nature

www.nature.com/nature

Vol 453 | Issue no. 7192 | 8 May 2008

Spring-cleaning in France

The French scientific research system is ripe for reform.

A year ago, on 6 May, France elected President Nicolas Sarkozy on a 'ticket of change' that included making research a priority and shaking up the science base. His government, to its credit, has made a fair start on both. Science and higher education, long relegated to puny junior ministries, have been given a full-ranking ministry, as have ecology and sustainable development — which also puts neglected issues such as climate change and energy high on the agenda. The state has upped science funding and has begun to modernize the way research is administered.

Sarkozy's government used his post-election momentum to swiftly push through a law shifting power away from the government research organizations, such as the CNRS and INSERM, to the universities. To an outsider, the law might seem quaint: it simply gives the state-run universities greater independence to manage their own budgets, to fix their own science strategies and to hire researchers on contract, with salaries negotiated on merit.

But this is France, where a naive interpretation of *égalité* has made taboo the competitive universities of Anglo-Saxon countries, and where most researchers are civil servants on identical pay scales. Attempting to change the status quo has been a recipe for street protests, and often political downfall. Students did protest against the new law, but the movement fizzled out after Valérie Pécresse, the country's science minister, faced it down while making only minor concessions. Forty years after the protests of May '68, the French have not lost their taste for revolt, but they seem to be slowly accepting that university reform is inevitable and overdue.

The universities will also be given a greater say in the running of the 1,300 laboratories they administer jointly with the research agencies. And thanks to the ramping up of the National Research Agency (ANR), created in 2005 to award grants on the basis of competitive proposals, university labs and young researchers are also less tied to funding by the agencies, which make rolling grants directly to labs.

Nonetheless, this change in status will take years to bear fruit, as most French universities are in a deplorable state. The government has funded the reforms with an extra €5 billion (US$8 billion), but the cost of redressing past neglect will stymie any immediate expansion of research and the recruitment of top talent.

This means that modernization of the research agencies remains imperative. A welcome restructuring of France's fragmented biomedical research has begun with a modest reform of INSERM, the national biomedical research agency, announced in March. The agency has been transformed into eight thematic institutes, which will regroup and concentrate research now spread across the CNRS and other agencies. A similar modernization of the CNRS, expected in June, is likely to transform it into a set of distinct institutes, each managing its own labs and long-term strategy.

"Forty years after the protests of May '68, the French seem to be accepting that university reform is inevitable."

It remains far from clear how the revamped research agencies will cohabit with the newly strengthened universities. Scientists must be vigilant to see that the right balance is struck among the agencies, the universities and the ANR. The research agencies have many merits. Rolling lab grants provide stability and opportunities for risk-taking that are rare elsewhere — even if too few researchers are taking advantage of them. And critics rightly complain that the young ANR is too skewed towards imposed top-down state 'strategic' themes to the detriment of investigator-driven basic research.

The broad inter-agency reforms also risk distracting from more pressing and targeted ones. The reform of INSERM does little to tackle the root problem of French biomedical research, which is that fixed pay scales and slow recruitment procedures make it almost impossible for France to compete in the highly competitive international job market. The softly-softly approach of introducing more flexible recruitment practices via the back door of the universities is not enough here.

So far, Pécresse has played a sensible opening game for the government, slowly moving her pawns forward to avoid confrontation while making reform a *fait accompli*. She may succeed where so many have failed. But sooner or later she needs to push forward a queen. ■

The gathering storm rages on

Two years on, a National Academies report on US competitiveness struggles to make an impact.

It is not often that a US senator talks about science, and even more rarely that she claims to bear good news. But that's the message Kay Bailey Hutchison, a Republican senator from Texas, tried to deliver last week to a restless crowd in a hotel ballroom in Washington DC.

The occasion was a look-back at a major 2005 report from the US National Academies on the state of science and engineering in America. Called *Rising Above the Gathering Storm*, it warned that the United States could soon lose its worldwide lead to other nations that have been investing aggressively in science and engineering education and research. Hutchison called the report a "wake-up call. Even in Congress, we got it".

But did they? Last August, Congress did respond by passing the America COMPETES Act, which vowed to double the physical sciences research budgets at such key agencies as the National Science Foundation, the Department of Energy's Office of Science and the National Institute of Standards and Technology. But in the United

©2008 Nature Publishing Group

States, legislation saying that budgets ought to be increased is separate from the legislation that actually increases them. And the latter promptly got lost in the government's budgetary meltdown, as Congress year after year fails to approve final numbers for each fiscal cycle until months later than expected. When the fiscal 2008 numbers were approved last December, the funding that Congress had pencilled in for the COMPETES Act — and that the agencies had been counting on — had disappeared. The resulting turmoil has forced research agencies to put major initiatives on hold, to put employees at national laboratories on unpaid leave, and to pinch pennies everywhere.

Many of the *Gathering Storm* authors in Washington last week were understandably furious. Broken promises are demoralizing, to say the least, and make it impossible for agencies to plan or manage coherently. Still, many of *Gathering Storm*'s best ideas could be implemented without waiting for Congress to collectively grow up and show financial responsibility. These ideas include bolstering programmes to train maths and science teachers; getting more students to enrol in advanced courses in high school; providing special funds to help young scientists start their own labs; and making it easier for foreign-born scientists to enter the country. Such measures would still require action from Congress, the president, or both. But they might very well be faster and easier to implement than the kind of major national commitment outlined in the America COMPETES Act.

In addition, it is important for supporters of the competitiveness initiative to remember that they, too, have a responsibility, which is to keep on communicating to legislators and to the American public at large why America COMPETES is more than just a 'Full Employment For Physical Scientists Act'. As David Ferraro of the Seattle-based Bill & Melinda Gates Foundation pointed out at the conference, the hotel ballroom was an "echo chamber": Americans elsewhere don't necessarily buy the notion that pouring money into research is the best way to spend their tax dollars. Indeed, some researchers argue that the statistics showing that the United States is falling behind have been misinterpreted (see H. Salzman & L. Lovell *Nature* **453,** 28–30; 2008).

So, while the *Gathering Storm* goals are worthy ones (see *Nature* **437,** 1208; 2005), supporters would be well advised to broaden their message beyond the usual suspects. Members of Congress are not going to stay on target for long when their constituents have other pressing issues, such as the economy or the war in Iraq, on their minds. ■

Bountiful noise

Whether in music or in nature, noise can be full of riches. The trick is to recognize the treasures.

Laughter and hisses — that's how a London promenade concert audience greeted the world premiere of a revolutionary musical composition in 1912. The response was hardly unusual, given that audiences of the day were regularly having their assumptions challenged by composers bent on redefining Western music. But unlike other dissonant masterpieces of that era, such as Igor Stravinsky's *The Rite of Spring*, the *Five Orchestral Pieces* of Arnold Schoenberg still come across to many as little more than noise. There are reasons for that, as a series of essays on science and music launched in this issue will make clear. But then, as other articles in today's issue illustrate, 'noise' has its treasures too.

Schoenberg's composition deliberately defied all the prevailing standards of music. It was, in his own words, "devoid of architecture or construction, just an uninterrupted changing of colours, rhythms, and moods". But it did have an expressive purpose, he insisted: "The music seeks to express all that swells in us subconsciously like a dream." Indeed, for today's sympathetic listener, the musical elements are distinctively recognizable and the emotional charge is tangible. Yet the language is still a challenge.

Of course, as Philip Ball explains in an Essay in this issue (see page 160), even more traditional music defies all attempts to explain its function in terms of mathematical or cognitive 'naturalness'. Subsequent essays in the series will highlight both the universalities in music — for example, how a mother's lullaby and rocking during early childhood are thought to lay a foundation for humans' aural and physical responsiveness — and music's diversity: the range of cultural conventions in such apparently fundamental elements as pitch scales and perceptions of rhythm. Essayists will also describe, for example, the challenges in acoustics of allowing audiences to hear music to its best advantage.

Drawing on musicology, statistics, cognitive and evolutionary biology and acoustics, the series will help us understand why most of Schoenberg's music is more challenging than that of his contemporary and champion, Gustav Mahler — let alone the music of Johann Sebastian Bach. But it will also remind us that none of these disciplines has yet been able to answer the fundamental question: why does music have such power over us? Nor can they explain how avant garde composers in the 1950s were able to take noise itself and make something new and true with it. Anyone who has performed Karlheinz Stockhausen's *Kontakte*, for example, which pioneered much subsequent electronic music by presenting manipulated electronic noise amid the sounds of percussion and piano, will tell you that the piece has an incomprehensible power. Anyone with an open musical ear who has listened to György Ligeti's *Atmospheres* for orchestra will say the same.

"The average listener isn't the least worried that musicologists and scientists cannot explain why we enjoy music."

The average listener isn't the least worried that musicologists and scientists cannot explain why we enjoy music. What matters is that its true bounties are recognized, and then explored and analysed. That applies not only to noise-like music, but also to nature. In that spirit, we can celebrate the fact that seismologists have begun to recognize and unpick the value of the ambient hum of the planet (see page 146). And we can enjoy the positive benefits that noise seems to have on living cells (see page 150).

Above all, what matters is that analysis strengthens rather than weakens humankind's sense of wonder — even as the natural terrain of exploration gets messier and as great composers make understanding music even more challenging. ■

©2008 Nature Publishing Group

RESEARCH HIGHLIGHTS

MATERIALS SCIENCE

Carbon on display

Nano Lett. doi:10.1021/nl080649i (2008)

Lumps of graphite are sooty; its constituent layers of graphene, however, are transparent, a property that researchers have used in a liquid-crystal device. Kostya Novoselov at the University of Manchester, UK, and his co-workers have shown that an electrode made from a thin graphene film can switch a device from opaque to transparent.

Most optical display devices have indium tin oxide as the transparent conductor, but this compound is apt to degrade because its ions are mobile. Graphene films, by contrast, are chemically stable and strong. The researchers find that the films can be cheaply spray-deposited onto glass from a suspension in an organic solvent.

ORGANIC CHEMISTRY

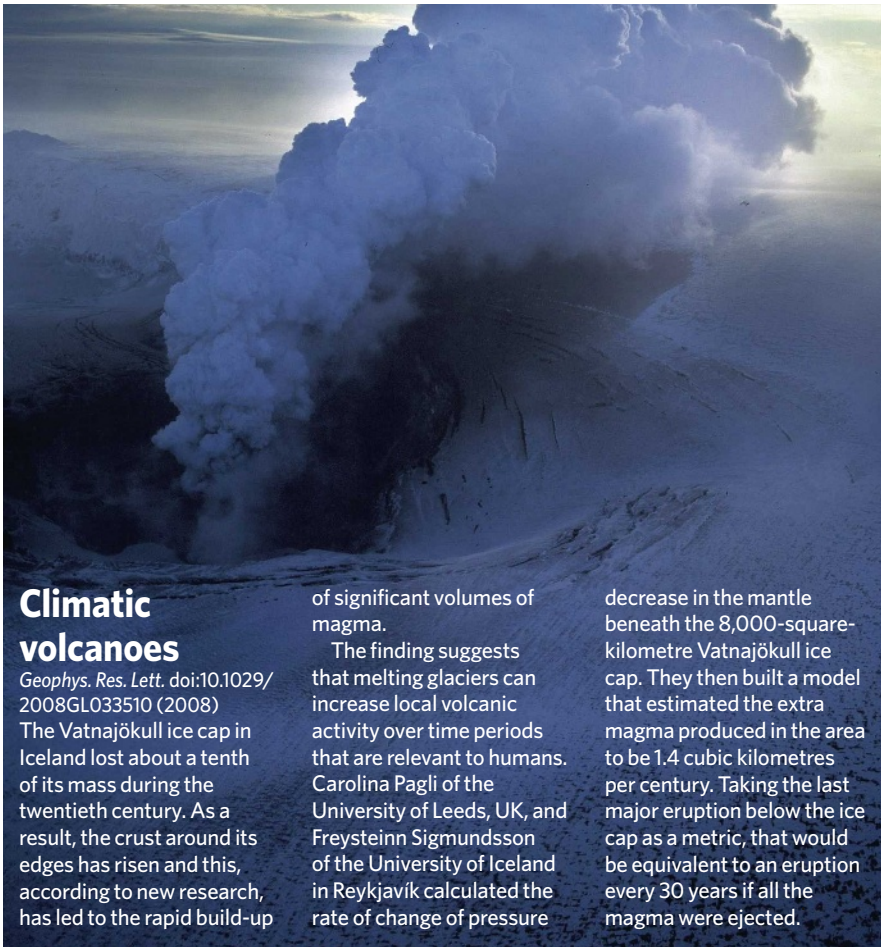
Flushing out HIV

Science **320,** 649–652 (2008)

The possibility of adding prostratin to antiretroviral therapies has been held back because it is scarce in nature, an impediment that the chemical's laboratory synthesis has just removed. Prostratin activates latent HIV viruses even in patients with undetectable viral loads, flushing out viruses that would otherwise be hidden. Including it in drug regimes might make patients who stop treatment after a few years less likely to experience 'viral rebound'.

Paul Wender and his colleagues at Stanford University in California found that they could make prostratin by first cleaving and then re-establishing the ring of three carbon atoms in phorbol, which is readily available. The process of putting the ring back together involves four steps that can be tweaked to produce various similar chemicals.

M. LANE/NHPA WRITEIMAGESPECS



P. STEFANSSON/REUTERS

Climatic volcanoes

Geophys. Res. Lett. doi:10.1029/2008GL033510 (2008)

The Vatnajökull ice cap in Iceland lost about a tenth of its mass during the twentieth century. As a result, the crust around its edges has risen and this, according to new research, has led to the rapid build-up of significant volumes of magma.

The finding suggests that melting glaciers can increase local volcanic activity over time periods that are relevant to humans. Carolina Pagli of the University of Leeds, UK, and Freysteinn Sigmundsson of the University of Iceland in Reykjavík calculated the rate of change of pressure decrease in the mantle beneath the 8,000-square-kilometre Vatnajökull ice cap. They then built a model that estimated the extra magma produced in the area to be 1.4 cubic kilometres per century. Taking the last major eruption below the ice cap as a metric, that would be equivalent to an eruption every 30 years if all the magma were ejected.

DEVELOPMENTAL BIOLOGY

Antler insight

PLoS One **3,** e2064 (2008)

Hans Rolf of the University of Göttingen in Germany and his colleagues report that they have found evidence of stem cells in the antler 'growth zones' of fallow deer (*Dama dama*; pictured left). They hunted for stem-cell 'markers' such as STRO-1 in several tissues located inside regenerating antlers and in regions called pedicles at the antlers' bases, pinpointing their quarry in several tissues.

The annual regrowth of antlers has long been considered the only case of complete appendage regeneration in mammals, and so this finding may prove useful in regenerative medicine.

NEUROSCIENCE

Bird brains

Science **320,** 630–634 (2008)

The babbling calls of baby songbirds involve separate brain circuitry from that used to generate more complex adult song, report neuroscientists at the Massachusetts Institute of Technology in Cambridge.

Michale Fee and his team gave adult zebra finches a drug that stopped the "high vocal centre" — a brain region that enables birds to sing melodious 'syllable' sequences — from working. The birds reverted to the stereotypical babblings of one-month-old chicks within 20 minutes, but regained their tuneful adult song when the drug wore off. Another brain region called the lateral magnocellular nucleus of the nidopallium proved necessary for juvenile babbles but not for adult song.

MICROBIOLOGY

A genetic monster

Proc. Natl Acad. Sci. USA **105,** 6730–6734 (2008)

A gargantuan bacterium carries tens of thousands of copies of its genome, researchers have found.

One species of the cigar-shaped bacterium *Epulopiscium* lives in the intestines of the unicornfish *Naso tonganus*, and can grow to more than half a millimetre in length. Esther

©2008 Nature Publishing Group

Angert of Cornell University in Ithaca, New York, and her colleagues have found that *Epulopiscium* cells also contain up to 250 picograms of DNA, compared with a human cell's 6 picograms, and have 50,000–120,000 copies of genes believed to occur only once in each genome.

Other bacteria contain multiple copies of their genomes, but so far none has been found to have nearly as many as *Epulopiscium*.

NANOTECHNOLOGY

Tiny carbon workers

Nature Nanotech. doi:10.1038/nnano.2008.98 (2008)

Micromechanical devices are most commonly made from silicon; now it is time to welcome carbon-based wafers to the party. Kenji Hata at the Nanotube Research Center in Tsukuba, Japan, and his co-workers say wafers such as those that they have built could provide an inexpensive approach to making these microstructures.

The researchers constructed the wafers by growing an array of widely spaced, vertically aligned carbon nanotubes. These were then 'squeezed' together by introducing an alcohol solution; the surface tension of the liquids and strong interactions between the nanotubes pulled the tubes into a densely packed arrangement.

These wafers can be processed using the same lithographic techniques that are used to work on silicon. The team has made mini-cantilevers and a range of other structures.

CHEMICAL BIOLOGY

Maths and malaria

Nature Chem. Biol. doi:10.1038/nchembio.87 (2008)

A 'guilt by association' test can correctly pinpoint the function of proteins, reports a team led by Elizabeth Winzeler of the Scripps Research Institute in La Jolla, California. An algorithm the researchers wrote found that the malaria-causing parasite *Plasmodium falciparum* expresses a gene encoding CDPK1 — a protein with an unknown role — at the same time as some other genes involved in cell invasion and movement. These all become active as the parasite prepares for its sexual stage.

The authors then screened a chemical library for compounds that block CDPK1. One of these inhibited the protein and also prevented *P. falciparum* from entering its sexual stage, indicating that the bioinformatics algorithm had provided accurate clues to CDPK1's role.

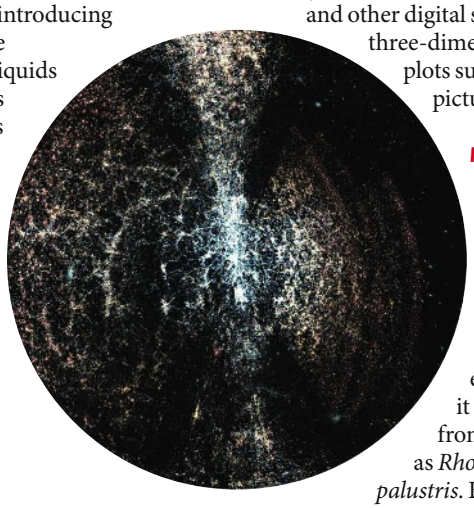
ASTRONOMY

Galactic mapping

Astrophys. J. **678,** 144–153 (2008)

Astronomers measure distance to galaxies in terms of 'redshift' — the far-off reddening of a galaxy's light as it zooms away from us. Current spectroscopic techniques measure at most a few hundred redshifts at once, but a new computer code can analyse larger batches of galaxies from digital images. It was developed by James Wray of Cornell University in Ithaca, New York, and Jim Gunn of Princeton University in New Jersey.

The algorithm works with general properties such as the colour and distribution of light across each galaxy. It gives good estimates of redshift when tested on 221,617 galaxies from the Sloan Digital Sky Survey, and could soon be used in the Sloan and other digital surveys to create three-dimensional galactic plots such as the one pictured left.



M. SUBBARAO, R. LANDSBERG & D. SURENDRAN

MICROBIOLOGY

Fuel cell

Environ. Sci. Technol. doi:10.1021/es800312v (2008)

Hydrogen is a useful and clean energy source, and it can be obtained from bacteria such as *Rhodopseudomonas palustris*. Bruce Logan of Pennsylvania State University, University Park, and his collaborators have found that this bacterium has another potentially useful skill: it can produce an electric current.

The researchers extracted *R. palustris* from a microbial fuel cell — a device in which bacteria deliver electrons derived from the oxidation of foodstuffs to an electrode, thus producing current. After culturing, this strain did the job more efficiently on its own than it had in the mixed bacterial population from which it came. The work might lead to the development of systems that capture electricity or hydrogen from the same cultures.

JOURNAL CLUB

François Balloux
Imperial College London

A population geneticist looks back in time in search of human origins.

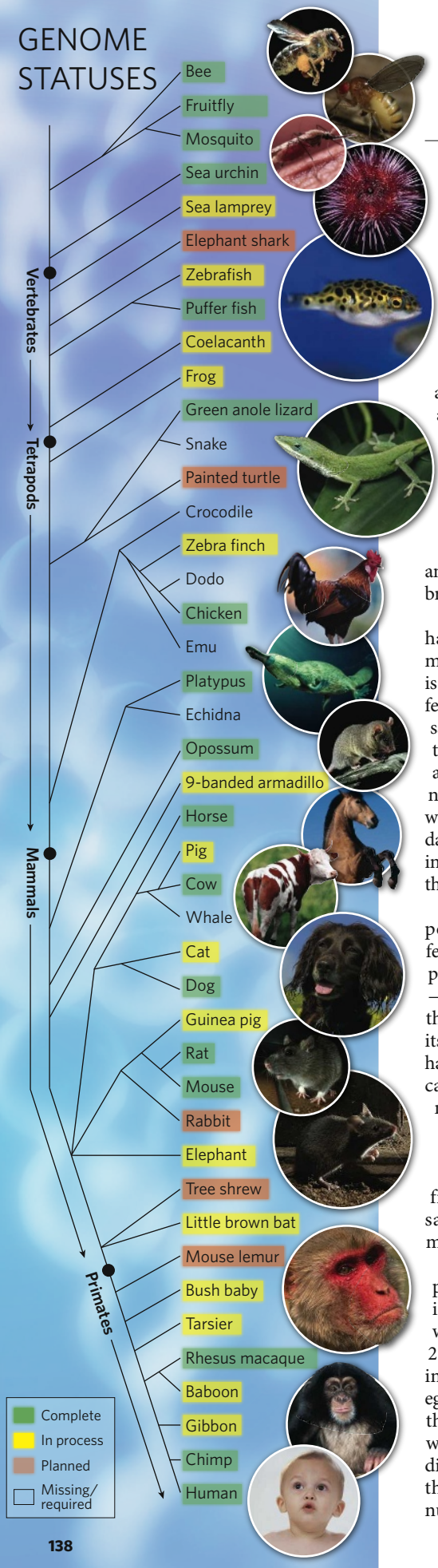
When and where anatomically modern humans evolved is arguably one of the most fundamental scientific questions. The issue also has philosophical and possibly even moral implications because it influences our definition of humanity. But I became involved in the subject for much more prosaic reasons. I was trying to make sense of the distributions among human populations of different versions of genes that imbue resistance to infectious diseases. It struck me that attempting to do this without a clear understanding of humans' past demography was bound to end in a muddle.

Despite decades of research, the origin of modern humans is still hotly debated. In a recent paper, Laurent Excoffier and his colleagues provide the first formal statistical evaluation of the likelihood for the various schemes that have been proposed (N. J. R. Fagundes *et al. Proc. Natl Acad. Sci. USA* **104,** 17614–17619; 2007). They conclude that a recent expansion from a single African origin is better supported by the current geographical spread of human genes than a multi-regional scenario. The multi-regional hypothesis proposes that modern humans hybridized with archaic humans, such as *Homo erectus*, as they spread.

This result may seem unsurprising because most genetic evidence points to an African origin some 60,000 years ago with no or negligible hybridization with archaic humans. However, there is a twist. By far the best supporting evidence for hybridization between modern and archaic humans has been the observation that, looking back, the amount of time it takes to reach the most recent common ancestor of some genes largely predates the age of our species. The extensive simulations in this paper debunk that argument by demonstrating that such cases can arise if modern humans had a recent and single African origin.

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

©2008 Nature Publishing Group



NEWS

Top billing for platypus at end of evolution tree

A draft sequence of the platypus genome reveals reptilian and mammalian elements and provides more evidence for its place in the ancestral line of animal evolution.

The platypus (*Ornithorhynchus anatinus*) is endemic to Australia and one of nature's oddest creatures, seemingly assembled from the spare parts of other animals. The semi-aquatic monotreme is a venomous, duck-billed mammal that lays eggs, nurses its young and occupies a lonely twig at the end of a sparse branch of the vertebrate evolutionary tree.

Now, the structure of its genome has revealed new clues to how mammals evolved. "The analysis is beginning to align these strange features with genetic innovation," says Wesley Warren of Washington University in St Louis, Missouri, the lead author of the genome analysis — a huge international project (see page 175). Comparisons with the genomes of other mammals will help to date the emergence of the platypus's distinguishing characteristics and reveal the genetic events that underlie them.

"There is nothing quite as enigmatic as a platypus."

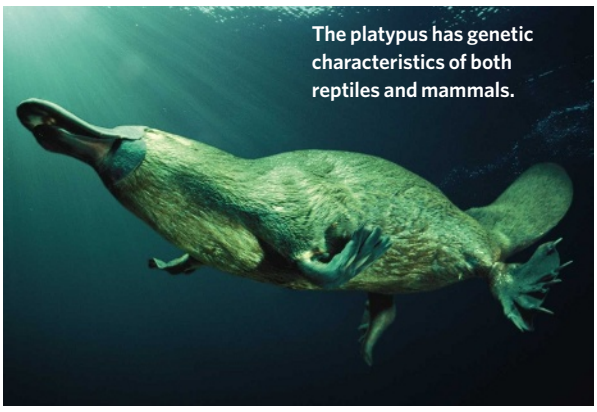
For example, mammals are defined by their possession of mammary glands, which in females can produce milk. Although the platypus doesn't have nipples, it produces true milk — full of fats, sugars and proteins — which the young suck through a glandular patch on its skin. The analysis shows that the platypus has genes for the family of milk proteins called caseins, which map together in a cluster that matches that of humans. This is a sign that one of the genetic innovations that led to the development of milk occurred more than 166 million years ago, and after mammals first split from the lizard-like sauropsids that gave rise to modern reptiles and birds.

The genes relating to the platypus's eggs offer further insight. The embryos develop within the maternal uterus for 21 days before they are expelled in a thumbnail-sized leathery egg. After 11 days of incubation, the young hatchlings emerge with their organs not yet fully differentiated. Like marsupials, they finish developing while nursing. The platypus shares with other mammals four genes associated with the zona pellucida, a gel-like coating that facilitates fertilization of the egg. But it also has two matches for *ZPAX* genes that had previously been found only in birds, amphibians and fish. And it shares with the chicken a gene for a type of egg-yolk protein called a vitellogenin. That suggests that vitellogenins, which are found in birds and fish, predate the split from the sauropsids, although the platypus retains only one vitellogenin gene, whereas the chicken has three.

Other characteristics that seem purely reptilian turn out to have evolved independently, the analysis suggests. Male platypuses have spurs on their hind legs that are loaded with a venom so potent it can kill a dog. Like the venom of reptiles, the poison is a cocktail of variations on at least three kinds of peptide. But the variations arose from duplications of different genes in platypuses than in modern reptiles. The similarity in venom is an example of convergent evolution between the two tetrapods.

"There is nothing quite as enigmatic as a platypus," says Richard Gibbs, who directs the Human Genome Sequencing Center at Baylor College of Medicine in Houston, Texas. "You have got these reptilian repeat patterns and these more recently evolved milk genes and independent evolution of the venom. It all points to how idiosyncratic evolution is."

The sex of the platypus is determined by a set of ten chromosomes, an oddity that sets it apart from all other mammals and from birds. These chromosomes link during meiosis to form a chain that ensures every sperm gets a set of all Xs or all Ys. Despite the similar designations,

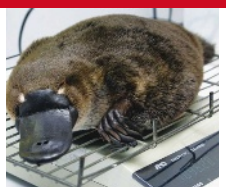


The platypus has genetic characteristics of both reptiles and mammals.

TOP TO BOTTOM: MEUL/ARCO/NATUREPL; N. A. CALLOW/NHPA; W. SLOSS/NATUREPL; J. ROTMAN/NATUREPL; J. JENSEN/IMAGEQUESTMARINE.COM; J. JENSEN/IMAGEQUESTMARINE.COM; G. ELLIS/FLPA; D. WATTS/NATUREPL; D. AUBREY/SPL; H. AUSLOOS/NHPA; T. J. RICH/NATUREPL; S. DALTON/NHPA; A. SANDS/NATUREPL; B. CASTELEIN/NATUREPL

D. WATTS/NATUREPL

©2008 Nature Publishing Group



D. PELED/AP PHOTO

PLATYPUS PLUS
Watch a video interview or listen to the podcast.
www.nature.com/news

none of the platypus X chromosomes resembles the human, dog or mouse X. "The sex chromosomes are absolutely, completely different from all other mammals. We had not expected that," says Jennifer Graves of the Australian National University in Canberra, who studies sex differentiation and is an author on the paper. Instead, the platypus Xs better match the avian Z sex chromosome. Another chromosome matches the mouse X, Graves and her colleagues report in *Genome Research* (F. Veyrunes *et al. Genome Res*. doi:10.1101/gr.7101908; 2008). This is evidence that placental mammalian sex chromosomes and the sex-determining gene *Sry* — found on the Y chromosome — evolved after the monotremes diverged from mammals, much later than previously thought. "Our sex chromosomes are a plain old ordinary autosome in the platypus," Graves says.

A team led by Gregory Hannon of Cold Spring Harbor Laboratory in New York sequenced microRNAs, which regulate gene expression, that were isolated from six platypus tissues. Again they found a mix of reptile and mammal examples (E. P. Murchison *et al. Genome Res.* doi:10.1101/gr.73056.107; 2008). "We have microRNAs that are shared with chickens and not mammals as well as ones that are shared with mammals, but not chickens," Hannon says.

"The reptilian characteristics [of miRNA] are not convergent features, and this is a feature of the genome as well," Hannon says. "Morphology didn't have to be reflected at the level of molecular biology, but in this case it was."

Adam Felsenfeld, who directs the Large-Scale Sequencing Program at the US National Human Genome Research Institute in Bethesda, Maryland, says: "I find it fascinating that genomic features of what are now two separate lineages can coexist in the genome of a single organism."

About half of the platypus genome contains non-coding DNA sequences. Many are 'interspersed repeats', copies of transposable elements that are characteristically abundant in other mammalian genomes. In contrast, repeats of very short sequences called microsatellite DNA are rarer in the platypus genome than in other mammals' and more closely resemble those of reptiles, with the balance of nucleic acids tipped toward A–T base pairs.

The sequence information has already generated useful genetic markers for studying the population structure of the elusive platypus in the wild. Differences in repeated elements, for example, separate the Tasmanian population from that on Australia's mainland, and could be used to improve understanding of the ecology of this enigmatic animal. There are as yet no plans to sequence the genome of its closest relative, the echidna. ■

Susan Brown

Chemists spin a web of data

A chemist running a computer server from his home is quietly solving one of his colleagues' biggest frustrations by providing the community with an open-access source of chemical information.

Although biologists have enormous public databases of genes and proteins, chemists usually have to pay for access to data on molecules. Chemist Antony Williams is hoping to change this in a move likely to ruffle the feathers of the American Chemical Society. Williams, a private consultant based in Wake Forest, North Carolina, has started a website called ChemSpider that has compiled data on nearly 20 million molecules in a year.

The modest project has made chemists interested in open access take notice — last week, the number of daily users of the site surpassed 5,000. "It's quite an exciting development," says David Wild, a chemical informatics researcher at Indiana University, Bloomington, who uses the service. "ChemSpider is working to integrate information in a unique way."

E. KOCH/ZEFA/CORBIS

Chemical data are becoming more freely available.

Chemical data have long been available, but at a hefty price. The largest supplier of such information is the American Chemical Society's Chemical Abstracts Service. The service, which is more than a century old, includes data on roughly 35 million molecules. But university and industry chemists must pay thousands of dollars to use the database. The society will not reveal numbers, but fees for using the database are thought to make up a substantial portion of its US$311-million annual income from 'electronic services'. Some have been highly critical of the society's grip on chemicals.

In recent years, several public sources for chemical information have appeared on the scene. The largest, PubChem, is run by the National Library of Medicine in Bethesda, Maryland, and contains data on some 19 million chemical structures. But PubChem's data focus on biological information, according to Williams. Other potential sources of information, such as Wikipedia, lack the algorithms needed to search chemicals according to their structure. "I noticed there was this gap," says Williams. "So I decided to try an experiment."

Rather than building up a database, the ChemSpider service scans open-access sources, including PubChem and Wikipedia, for chemical data. It compiles the publicly available information in a single location, and allows users to follow links to the original source material. The site is maintained with modest profits from advertising and the work of about 30 active volunteers who double-check the data pulled in from outside.

The site is not without its flaws. "There's an awful lot of chemical information, but there's an awful lot of rubbish as well," says Barrie Walker, a retired industrial chemist in Yorkshire, UK, who helps maintain the site. When working with such a large database, he says, "you're bound to end up with a quality issue". Williams adds that the site still has problems with certain searches. For example, it struggles to distinguish between isomers: molecules with the same chemical formula arranged in different structures.

But Williams nevertheless believes that the service may be able to compete with for-profit services. "What I'm doing is highly disruptive," he says. "I think it can be done and it needs to be done." The American Chemical Society declined to comment on ChemSpider. ■

Geoff Brumfiel

©2008 Nature Publishing Group

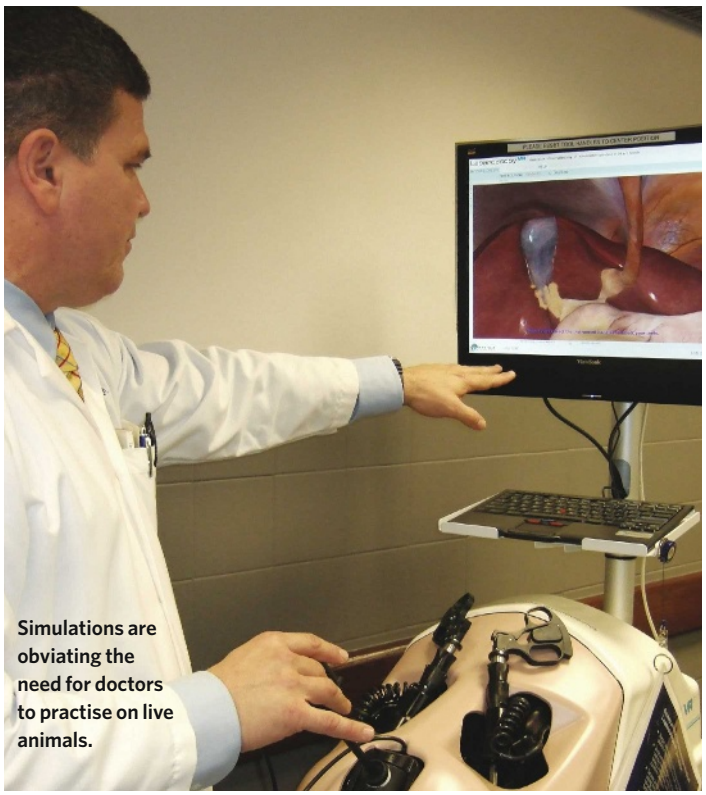
SPECIAL REPORT

Medical schools swap pigs for plastic

Doctors used to try out their surgical skills on animals before being allowed to work on patients. Now just a handful of US medical schools still have animal labs. **Meredith Wadman** asks if they've lost a vital tool.

This month sees the shutdown of the live-animal laboratory at Case Western Reserve University School of Medicine in Cleveland, Ohio. The lab is currently used to train medical students, allowing them to practise on anaesthetized pigs before attempting their first incision into humans. But the school, which has used live cats, dogs and ferrets in its surgery programme in the past, intends to stop using live animals at the end of this semester in favour of technologies such as virtual simulations.

It is the latest closure in a phase-out of animal labs across the United States: in 1994, live-animal experiments were on the curriculum in 77 of 125 medical schools; now it is thought that just eight use them. Several have stopped in the past year, including New York Medical College in Valhalla, which this year ended its practice of using live dogs to teach cardiovascular physiology to first-year students. And the trend is being played out across the globe (see 'All around the world').



Simulations are obviating the need for doctors to practise on live animals.

M. WADMAN

Cost is undoubtedly a factor — it is expensive to maintain animals and to employ veterinary staff. But schools such as Case Western and New York Medical College have said that the decision to eliminate live-animal experiments was based mainly on improvements in alternatives. The New York school now uses echocardiography on volunteer students and simulators that mimic cardiac arrest or a drug's action, for example.

Simulation has developed hugely over the past decade. "It is a lot more than a couple of mannequins," says Bruce Jarrell, vice dean of research at the University of Maryland School of Medicine in Baltimore, which a little over a year ago opened its surgery simulation and technology centre. Students practise using surgical instruments to lift coils of rope viewed over a monitor, much as intestines are lifted during bowel surgery. They use the controls during a simulated endoscopy while watching a realistic duodenum on a monitor. Nurses learn to intubate a mannequin that can be programmed to respond to administered 'drugs' with changes in heart rate and blood pressure. And minimally invasive surgery is tried by students using instruments that mimic those used in actual surgery to clip an 'artery' — complete with 'blood' — during a simulated gall-bladder removal, viewed on a computer screen.

The most advanced simulators have 'haptic feedback', which provides students with the sensation that their instruments are touching real tissue.

Advances such as these have made use of live animals for training in medical schools gratuitous, says John Pippin, a cardiologist based in Dallas, Texas. Pippin once used live dogs to study heart attacks but now works full-time as senior adviser for the media-savvy animal rights group Physicians Committee for Responsible Medicine in Washington DC. The group has become a thorn in the side of deans and administrators in the 6% of US institutes that continue to use live animals to train future doctors. Heading their list is Johns Hopkins School of Medicine in Baltimore, Maryland, which is consistently rated in the top few schools in the country and unapologetically uses live animals.

"Hopkins is the only top school that still uses animals in the medical school curriculum for any purpose," Pippin says. "It is on an island." He was one of seven doctors attending a demonstration outside Johns Hopkins hospital in March protesting against the use of live pigs.

Jonathan Lissauer, a doctor who recently trained at Johns Hopkins, concedes the argument for animal use in medical research and advanced surgical training. He says that sometimes they were used "as just a diversion for people who won't be using those skills at all. I think then you cross the territory from appropriate medical education to something worse than that," he says. "There was no medical utility in having pigs die so that people going into psychiatry could play around.

"From a purely academic perspective," he adds, "I thought there were substantial differences between human tissues and pig tissues — a lot of textural differences — and that the practising wasn't overly useful because of that."

But Johns Hopkins' director of surgery, Julie Freischlag, makes no apologies for her programme's use of roughly 50 pigs and US$75,000 a year. She argues that the two days students spend in the pig lab are important in helping them decide if they are drawn to a surgical specialty. It also trains doctors who won't become surgeons but still need to know how to start intravenous lines and work with sutures.

"The first time our graduates stitch you up in the emergency room as interns, they will have already done that on live tissue before," Freischlag says. "They will be safer and better. I think most of us would hope they have actually done that on someone or something else before us."

In the pig lab, students are taught how to take out the kidneys, part of the stomach, part of the liver, the gall bladder and the spleen. They learn how to operate on a lung and how to repair organs injured by trauma. They practise tying off arteries. They learn to control bleeding, handle tissues gently and finish the operation with the incision looking appropriate.

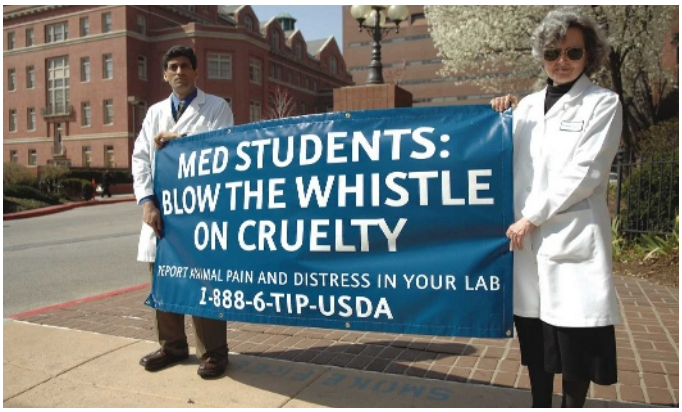
A veterinary assistant and veterinarian attend

©2008 Nature Publishing Group

HAVE YOUR SAY
Comment on any of our news stories, online.
www.nature.com/news

the lab; the latter anaesthetizes and euthanizes the animals, which are purchased from contractors.

Freischlag says that the lab "is a totally elective part of the surgery rotation" — it is not graded, and students can opt out. She says that no student has opted out of the lab in advance in the five years she has been in her job. "One person did decide they didn't like it when they got into it and then opted out." Lissauer, however, claims that when he participated in the pig lab two years ago, "students who didn't feel comfortable taking part in an animal lab felt pressured to do so".



K. HAIRSTON/BALTIMORE SUN

The activist group Physicians Committee for Responsible Medicine campaigned against the use of pigs at Johns Hopkins Medical School.

Freischlag says that no amount of book learning, lectures or computer simulation can substitute for the experience of working with living tissue, and the unpredictability and bloodiness of real surgery. Controlling bleeding is a priority in surgery, not least because excess blood obscures the surgeon's vision. "It is really a contact sport," Freischlag says.

Others agree that there is value in schools training doctors-to-be on live animals. "More and more institutions are opting out and I don't think that there has been an adequate assessment of the educational impact, because it is very difficult to do," says Alice Ra'anan, the director of science policy at the American Physiological Society in Bethesda, Maryland, which supports the use of animal labs. "How do you do a controlled experiment of what the impact is as medicine evolves?"

Larry Laughlin, the dean of medicine at the Uniformed Services University of the Health Sciences in Bethesda, Maryland, which uses roughly 75 pigs a year to teach medical students physiology and surgery, says: "I'm not troubled if 10 or 100 other medical schools do not use animals. My focus is on what's best for our students, based on the educational professionals who advise me." A committee is in the process of reviewing his institution's use of animals and is expected to report in September.

Laughlin points out that US law requires that animal labs are approved by ethics committees, which must be persuaded that there is a compelling justification. "Thousands of times more pigs are slaughtered and have worse lives and suffer worse demises in Iowa every day than we do in a year," says Laughlin, who grew up on a livestock farm in the midwest. "Therefore it is hard for me to rationalize the intense concern."

One third-year medical student at Laughlin's institute, whom school officials insisted remain anonymous, says: "We have our simulation centre down the road where we learn how human bodies are supposed to react. But in our pig lab we have the opportunity to see how life actually reacts." And having experienced both, he contends that the pig lab has made him and his colleagues better surgeons-to-be than the fourth-year students from other medical schools that he encounters during their visiting, surgical electives. "We are a lot better prepared for what goes on in an operating room. To me, it is a real shortfall of their education, he says."

"I always feel sorry, a little bit, for knowing that that an animal gave its life for me. But at the same time, if I can take from this animal and learn how to save the lives of patients, definitely to me the benefit outweighs that sadness." ■

All around the world

The decline in the use of live animals to train medical students can be seen worldwide.

In the United Kingdom, a law has banned the use of live animals to train medics since 1986, except in microvascular-surgery training. However, nine UK universities use rodents to teach biosciences to undergraduates, a process that involves lengthy applications for project licences, which must be rewritten every five years.

Last week, five of the country's major funders of animal research, including the Wellcome Trust and the Medical Research Council, published common guidelines for use of animals in research. Compliance with the guidelines, which include looking at alternatives to animals and reducing their use, will be a condition of funding for new grants involving animal work.

Elsewhere in Europe practices vary. In Germany, for example, the Animal Welfare Act allows animals to be used for training purposes if an institution can provide a compelling reason for why they are needed and why alternatives do not measure up.

The Federation of European Laboratory Animal Sciences Associations is launching a working group to review animal use in education in 19 member countries. Its recommendations will influence the 22-year-old European Union directive on the protection of animals, which is under review.

In Australia, where 15 years ago medical students were taught neurophysiology using feral cats, there has been a marked reduction in the use of live animals. "I am not aware of any live-animal use in practical classes in any medical school in Australia," says Geoff Dandie, the chief executive officer of the Australian and New Zealand Council for the Care of Animals in Research and Teaching.

The decline is the result of successive revisions to an Australian code of practice, which discourages the use of live animals in classes other than specific areas of professional development. Animals are now rarely found outside of refresher courses for seasoned emergency-department doctors. There, sheep are used in training for rare, but essential, life-saving interventions such as tracheotomies.

Only one of 18 medical schools in Canada — Memorial University in St Johns, Newfoundland — still uses live animals, and it is reviewing its curriculum with a view to eliminating the practice.

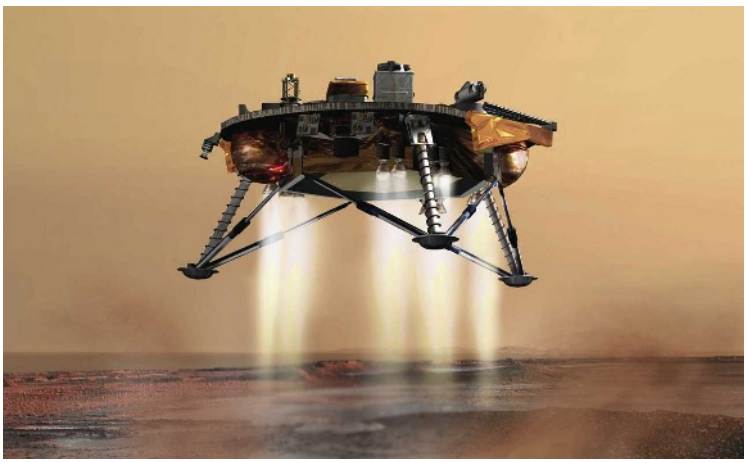
In India, animals are available for the teaching of medical students under restrictive conditions, says Maneka Gandhi, a member of parliament and animal-rights activist. But Harmeet Rehan, head of physiology at Lady Hardinge Medical School in New Delhi, where rabbits are used, says that he is not aware of any Indian institution that uses large animals to teach surgical skills to undergraduates. Rehan liaises with several medical schools on behalf of a national committee that must give the go-ahead for any large-animal experiments.

In Argentina, "it is not common", according to Pablo Pratesi, chief of critical care medicine at Austral University in Buenos Aires, where Pratesi's programme uses 20 pigs a year to train final-year medical students in surgery skills. **M.W.**

©2008 Nature Publishing Group

Phoenix descending

NASA's Mars strategy goes from "follow the water" to "arrive at the ice".



NASA

Since 2001, the slogan for NASA's Mars programme has been "follow the water".

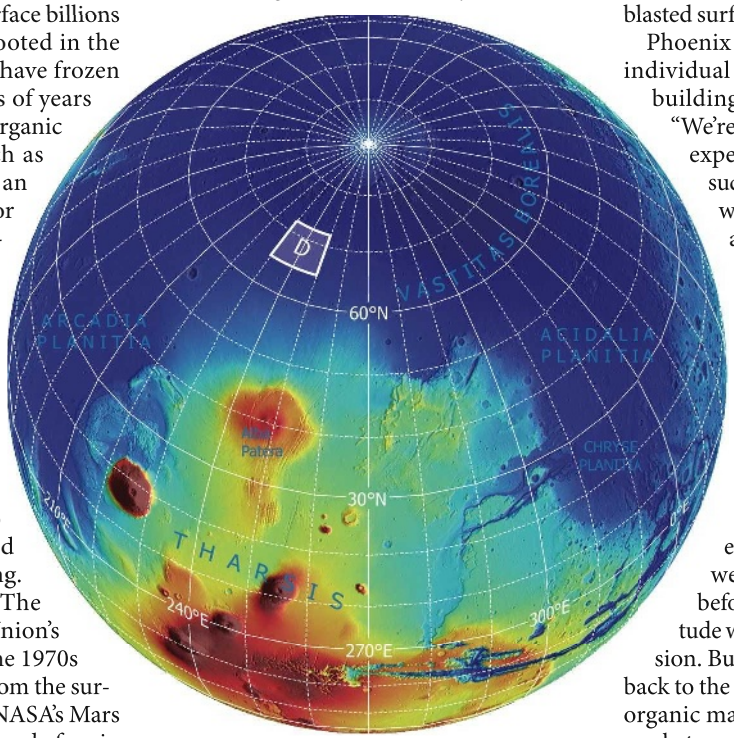
With Phoenix, a US$420-million mission to the edge of the planet's north ice cap, the agency hopes to finally touch its quarry, in the form of dirty water ice scraped from the subsurface and melted in the probe's on-board ovens.

If all goes as planned, Phoenix will reach the end of its 680-million-kilometre, 10-month-long journey on 25 May. Its landing site is in a region where NASA's orbiting Mars Odyssey spacecraft has detected gamma-ray and neutron signatures suggesting a significant amount of hydrogen — thought to be a constituent of frozen water — near the surface. Soon after unfurling twin solar arrays, the lander will extend a robotic arm to dig down as far as half a metre in search of that ice.

Older missions, including the rovers Spirit and Opportunity, have shown that there was water at and near the planet's surface billions of years ago. Phoenix will be rooted in the present, sampling ice that may have frozen in place mere tens of thousands of years ago and also sifting the soil for organic chemicals and substances, such as sulphur, that life might use as an energy source. "Phoenix will, for the first time, give us the opportunity to directly assess habitability," says Doug McCuistion, NASA's Mars-exploration programme director.

Before that, though, it has to get down safely. Five NASA spacecraft have landed successfully on Mars in the past: two 1976 Vikings landers, the 1997 Pathfinder and the two rovers that arrived in 2004 and are, remarkably, still going strong. But there have been misfires. The most successful of the Soviet Union's various attempts at landing in the 1970s sent back only seconds of data from the surface — but that was better than NASA's Mars Polar Lander, which was never heard of again after entering the atmosphere in 1999, and the British-run Beagle 2, lost in 2003.

Phoenix, a 350-kilogram lander, inherited hardware from the Mars Surveyor Lander, a mission cancelled in 2000 after the failure of the Mars Polar Lander, with which its design had much in common. As a result of its sibling's demise, Phoenix's principal investigator Peter Smith of the University of Arizona in Tucson and his team eliminated more than two dozen flaws that might have conceivably led to failure from hardware and software. "We may not succeed, but we deserve to succeed," he says.



Phoenix hopes for a picture-perfect landing (top) on Mars's northern plains (above).

Engineers at the Jet Propulsion Laboratory in Pasadena, California, which runs most of NASA's planetary missions, will have their last chance to tweak the trajectory on 24 May. The next day, Phoenix is expected to use the planet's atmosphere, and its own heat shielding, parachutes and retrorockets, to slow itself down from an orbital speed of 20,000 kilometres per hour to a soft, safe landing at just 2.4 metres per second.

Smith and his team hope that the ice and soil samples they will study with the spacecraft's mass-spectrometer and chemical analysis systems will contain hints of organic molecules, of which the Viking landers found no trace at lower latitudes. The polar environment might be more hospitable to persistent organics — either from incoming meteorites or primordial to Mars — than the corrosive and radiation-blasted surfaces which the Vikings scratched.

Phoenix can measure salts, pH levels and individual chemicals, but can't analyse the building blocks of life, such as proteins. "We're an ideal stepping stone for more expensive and sophisticated missions such as the Mars Science Laboratory, which can really look at molecules and say, 'This is DNA'," says Smith, referring to the Cadillac of NASA Mars missions — a long-duration rover weighing 850 kilograms and budgeted at nearly $2 billion — which is due for launch next year to one of a shortlist of six sites closer to the planet's equator.

By August, Phoenix will have used up its ovens and laboratory equipment. It will hang on as a polar weather station for a few months before the severe cold of the high-latitude winter in all likelihood ends its mission. But its results may lure other probes back to the poles. "If we find an abundance of organic matter in this ice," Smith says, "you can bet your bottom dollar there'll be another mission going there someday." ■

Eric Hand

©2008 Nature Publishing Group

Q&A

Research revolution?

J. HEKIMIAN/GETTY IMAGES

Valérie Pecresse has been a member of the French National Assembly (Yvelines department) since 2002. She rose to prominence as the combative spokeswoman for Nicolas Sarkozy's centre-right UMP party during the 2007 presidential race, after which she was appointed minister for higher education and research.

Reforming the country's research organizations and dilapidated universities is quite a challenge to take on.
It is something politicians — particularly those on the right — have paid little attention to, because you are up against such a politically hostile terrain. Every minister who has tried to reform French universities has fallen, so I felt I was taking a big political risk, but modernizing the research and education landscape is of major importance.

What is your vision for this?
With the law on university autonomy (introduced last summer) we will make the universities powerful and independent, in charge of their own budgets, staff and science strategy. We also intend to spend more to help the reforms succeed — that can't be done on existing budgets. We're the only OECD [Organisation for Economic Co-operation and Development] country that spends less on its university students than on those in high school. On 30 April we launched 'Operation Campus' which invests €5 billion [US$7.7 billion] in real-estate projects to create a dozen large world-class campuses allied to existing institutions. The logic is to break down the over-compartmentalization of French research.

Some worry that you will dismantle the CNRS, France's basic research agency, and turn it into a funding council.
We are not going to break what works. The CNRS is an incontrovertible player in French research, but we must empower the universities and make them independent. The role of the research organizations will be to steer research strategy at the national level, but I want to reform them to put an end to the dispersion of research and sometimes a lack of overall coherence. For example, biomedical research is spread across the CNRS, INSERM [the national medical research agency] and a series of dedicated agencies that in the past were created for every new priority disease. When Sarkozy announced that he was making Alzheimer's disease a priority, some people suggested that we create an Alzheimer's agency. That made me want to tear my hair out; we can't go on adding more structures like this. It Balkanizes research, stymies interdisciplinarity and leads to duplication.

How will this change?
A reform announced in April will bring all biomedical research under INSERM, with eight internal institutes representing the main research themes setting national strategies for each. It is INSERM that will organize and articulate the national biomedical strategy in France. The CNRS will propose similar reforms in June.

What will you do about the legendary French bureaucracy?
We absolutely must simplify the bureaucracy. At present, we have some 1,300 laboratories run jointly by the research organizations and the universities. One in four has more than four parent agencies: that's more than four different funding sources, four different accounting and IT systems and four different evaluation systems. We propose a single management agency, the host institution, which will usually be the university. Scientific strategy will continue to be jointly decided, but there will be a single administrative system, and a single consolidated stream for public funding.

Sarkozy told *Nature* last year that his dream was for more French scientists abroad to return home. Is this happening?
Not yet. But when autonomy comes into force in January 2009, universities will be free to recruit who they want, at the salary they want, on contract. The unions say we are killing the civil-servant status enjoyed by researchers. The truth is that only a small proportion of posts will be contracts at the start. Top researchers don't necessarily want civil-servant status. A young scientist, who has done postdocs in the United States, who has a family, won't return to France to earn €2,000 a month as an associate professor. ■

Interview by Declan Butler

See Editorial, page 133.

ON THE RECORD

"Love of God and compassion and empathy leads you to a very glorious place, and science leads you to killing people."

Ben Stein, star of anti-evolution movie *Expelled*, adds his sensible and rational voice to the science-versus-religion debate, during an evangelical webcast.

SCORECARD

▲ **Dino dung**
Two pieces of 130-million-year-old fossilized droppings fetched US$960 at a New York auction last week.

▼ **Ancient meteorite**
Meanwhile, at the same auction, the meteoric star of the show, expected to fetch as much as $2.75 million, was left unsold.

M. SEGAR/REUTERS

OVERHYPED

Virtual lego
The Danish company behind the world's favourite plastic bricks is preparing to launch Lego Universe, letting players build virtual constructions online. Mark Hansen, leader of the new universe, admits it will "never replace the physical experience" of Lego. Quite.

NUMBER CRUNCH

19 is the age of Alia Sabur, the world's youngest professor. She will teach physics at Konkuk University in Seoul.

2 is the number of years before Professor Sabur is able to legally purchase alcohol in her home city of New York.

1717 was the year in which the previous youngest-ever professor was appointed — Colin Maclaurin, who was awarded a professorship by the University of Aberdeen, UK, also at 19 years of age.

Sources: Trinity Broadcasting Network, Reuters, The Independent

SIDELINES

©2008 Nature Publishing Group

Institutes marshal locals to boost African physics

Two research centres will open their doors in Cape Town, South Africa, this month in a bid to bolster theoretical physics and mathematical sciences across the continent.

A pan-African centre of the existing African Institute for Mathematical Sciences will be launched on 12 May, followed the next day by the National Institute for Theoretical Physics, to be headquartered at the Stellenbosch Institute for Advanced Study.

The two institutes are funded mainly by the South African Department of Science and Technology. The pair will work closely with each other, using mainly local scientific talent to contribute to regional cosmology and astrophysics programmes such as SALT (the Southern Africa Large Telescope) in Sutherland and MeerKAT, a radio-telescope facility under construction in the Karoo desert.

Patent on Mexican yellow beans is reversed

The US Patent and Trademark office last week overturned a controversial patent on a breed of yellow beans. Opponents of the patent say the bean has been eaten in Latin America for more than a century, raising issues of biopiracy.

The patent was granted in 1999 to Larry Proctor of Delta, Colorado. According to the patent application, Proctor bought yellow beans in Mexico and bred them for two years to grow plants that gave a better harvest and produce beans with a distinctive yellow colour.

Proctor then began charging licensing fees on imports of yellow beans from Mexico, prompting the International Center for Tropical Agriculture, based in Cali, Colombia, to challenge the patent in 2001.

But the battle may continue. Proctor has the option of contesting the decision in federal court, and says he is consulting his lawyer. "Everybody may not be happy with what we're fixing to do now," he said, and declined to comment further.

CIAT 2008

Yellow beans have been freed from patent control.

Elephant-hunting season opens in South Africa

MCPHOTO/STILL PICTURES

South Africa's 13-year ban on elephant culling was lifted on 1 May to help manage the flourishing population.

The pachyderms (pictured) — once close to being wiped out in many parts of the continent — have more than doubled in number in the country since 1995, generating concerns about vegetation destruction, and threats to people and livelihoods within elephant ranges.

Culling — under strict provisos — has been legalized as a last resort, the government asserts. Yet some conservationists have condemned the action, warning that it may encourage ivory poachers and could threaten populations elsewhere.

Alternative ways of curbing elephant numbers include relocating the animals and hormone-based contraception.

Drug firm turns spotlight on basic systems biology

The pharmaceutical company Pfizer has launched a three-year, US$14-million systems-biology consortium to improve the understanding of diabetes and obesity.

Systems biology uses computer-intensive data analysis to derive models about specific biological phenomena, and can be used to help study the progression of some diseases. Drug companies have traditionally shied away from it, because it can take a long time to see any financial pay-off. But in March, systems-biology research, funded in part by Merck, identified genes involved in obesity.

In the new Pfizer programme, researchers at the University of California, Santa Barbara, the California Institute of Technology in Pasadena, the Massachusetts Institute of Technology in Cambridge, the University of Massachusetts, and biotechnology company Entelos based in Foster City, California, will examine the regulatory mechanisms involved in insulin signalling in fat cells.

NASA watchdog calls for Orion board suspensions

NASA's independent investigative arm has called for the suspension of 6 of the 19 members of a board charged with reviewing the agency's programme to replace the space shuttle.

The board members in question are employed by contractors already working on Orion, the shuttle's replacement. This creates a conflict of interest that violates US federal law, the Office of the Inspector General said in a report issued last week.

But NASA officials have said that the board should remain intact and that it is difficult to find completely independent aerospace experts. The inspector-general's office is calling that "nonresponsive". Further action may be taken after 28 May, when NASA's comments on the report are due.

Sacked whistle-blower demands reinstatement

A researcher who blew the whistle over animal-rights issues at the University of Nevada in Reno went to court last week to try to win his job back after being fired.

Nutrition researcher Hussein Hussein, a tenured professor, was sacked last month by university president Milton Glick, despite a recommendation by a university administrative hearing that he merely be reprimanded or demoted. Four years ago, Hussein reported deficiencies in the care of laboratory animals that led to US$11,400 in fines against the university. He claims the university sought to fire him in retaliation.

After the administrative hearing, he was cleared of charges of plagiarizing graduate student work, but found to have incorrectly reported on $377,000 in grants, thus denying the university overhead costs.

Hussein's lawyer asked a Nevada judge to reinstate him, arguing that he was improperly fired. A spokeswoman says the university acted appropriately, and will vigorously defend its actions.

Correction

The news story 'Rodent round-up' (*Nature* **453,** 9; 2008) incorrectly described the greater white-toothed shrew as a rodent, when it in fact belongs to the order Soricomorpha.

©2008 Nature Publishing Group

NATURE|Vol 453|8 May 2008

HARNESSING THE HUM

A new way to analyse seismic vibrations is bringing order out of noise to help predict volcanic eruptions or create detailed images of Earth's interior. **Rachel Courtland** reports.

S. GÉLABERT/LOOKATSCIENCES

Seismologists like noise — the louder, the better. Thumpers, airguns, explosives and earthquakes are their favourite tools. The vibrations made by these tools — called seismic waves — can be recorded on a detector, sometimes located thousands of kilometres away. With enough events, and enough detectors, seismologists can follow the waves' paths to illuminate the structure of the Earth below, just as X-rays illuminate the internal structure of a human body. For decades, waves from these large, single sources have told us what the planet looks like under the surface.

But Earth is also awash with much quieter vibrations that may turn out to be just as illuminating. This persistent hum doesn't have a single source, but instead comes from a combination of minor tremors, the long drawn-out echoes of major earthquakes and the crashing of ocean waves. These faint vibrations come from all directions and bounce around multiple times in Earth's interior. On a seismograph, they show up as a seemingly meaningless series of spikes and troughs. Yet over the past few years, seismologists have transformed the hum from a nuisance to a powerful tool to image Earth's crust and upper mantle.

The new field, known as ambient-noise tomography, has several advantages over traditional seismic imaging. Its improved resolution may help to answer some persistent geological questions. It also means that researchers do not have to wait for one-off events such as earthquakes, opening up the possibility of imaging Earth's crust over time. In addition, the technique works for seismically quiet areas that do not routinely see the earthquakes needed for traditional imaging studies.

Seismologist Andrew Curtis of the University of Edinburgh, UK, says that using the technique is like discovering an unknown continent. "An incredible new area of exploration has come upon us," he says.

Noisy beginnings

Random noise has long been used to tease out the intrinsic properties of materials. The theoretical underpinnings trace back to the 'fluctuation-dissipation theorem', formulated in 1928 to describe how noise in an electrical signal can reveal a material's response to current. The theorem suggests that, with proper analysis, random vibrations inside Earth could act like seismic waves, revealing structure and hinting at properties such as temperature, composition, orientation and stress.

> **"The idea is that in waves that are fully garbled, there is still some sort of residual coherence."**
> — Richard Weaver

All noise passing through Earth must pass through the same internal structures, and hence each signal — although inherently random — retains a sort of 'memory' of the material it has passed through. So by looking at correlations in noise between two seismometers, researchers can extract information on the material lying between the pair. "The idea is that in waves that are fully garbled, there is still some sort of residual coherence," says Richard Weaver, a physicist at the University of Illinois at Urbana-Champaign.

The research started to take off in 1999, when Weaver attended a workshop on random waves in Corsica and met Michel Campillo, a geophysicist at Joseph Fourier University in Grenoble, France. At the time, Campillo was looking at earthquake codas, the multiply reflected echoes of quakes, recorded by a seismometer network in southern Mexico. Campillo thought that codas could reveal how the crust and upper mantle scatter seismic waves. But he wasn't quite sure how to proceed — how, for instance, should he incorporate waves that bounce off the Earth–air boundary in his analysis? Weaver suggested a fix, and an additional direction: compare the noise in pairs of detectors a specified distance apart, look for evidence of correlations, then invert the timing of any correlated spikes to get the speed of the seismic waves. It was such a radical idea that Weaver says: "in hindsight, I'm not so sure why I was so confident."

After the workshop, the two went off to investigate how the technique could be used to probe the properties of materials. As a proof of principle, Weaver showed that ambient ultrasound vibrations in aluminium could be used to determine the material's elastic properties[1]. Then Campillo applied the same technique to the earthquake coda; the noise seemed to be correlated just as he hoped it would be[2]. So Campillo began to eye the ambient noise that dominates seismographs when the codas fade, in the hope of quickly and finely imaging the crust. It worked.

In 2005, two teams[3,4], one of which included

©2008 Nature Publishing Group

Campillo and his colleagues, used ambient noise to construct three-dimensional maps of seismic-wave speeds in southern California. The images had a horizontal resolution as small as 60 kilometres, more than four times the detail in traditional earthquake tomography.

The success of the technique was a surprise to Campillo. Seismic noise on the California coast is lopsided, as the eastward-moving waves from the Pacific Ocean tend to swamp noise from other directions. "If you want to use the noise, you have to assume the noise has the right properties, that there is an even distribution of sources," says Campillo. "And of course that's not true — it was not at all obvious that noise would work as well."

"It was not at all obvious that noise would work."
— Michel Campillo

Most of Earth's ambient noise reverberates through the shallower parts of the planet, making the technique most useful in the upper 60 kilometres of Earth. But since 2005, several teams have used the technique to look deeper into Earth's crust — and even into the upper mantle.

Collisions to climates

Paradoxically, some of the clearest pictures of those depths come from the highest plateau on Earth, the Qingzang Gaoyuan plateau (often referred to as the Tibetan plateau). At an average elevation of 4.5 kilometres above sea level, the plateau is being pushed ever-higher by the ongoing collision between the Indian and Eurasian continental plates. The thickest part is on the eastern side of the plateau, where the crust is roughly 80 kilometres thick — twice the average for the continental crust. Working out exactly how and when the plateau got so high could help to improve climate models, as some researchers posit that the arrival of Indian monsoons coincided with the rise of the eastern side of the plateau.

Geologists are also still debating how the crust on the eastern side grew so thick. One theory holds that it thickened as it was squeezed against the stronger crust of the neighbouring Sichuan basin. Others think that the crust from the Indian subcontinent is largely responsible, pushing and flowing through the plateau, inflating it like a tire. "It's sort of a conservation of mass problem. You're shoving one continent into the belly of another continent, and you want to know where all the mass goes," says Thorne Lay, an Earth scientist at the University of California, Santa Cruz.

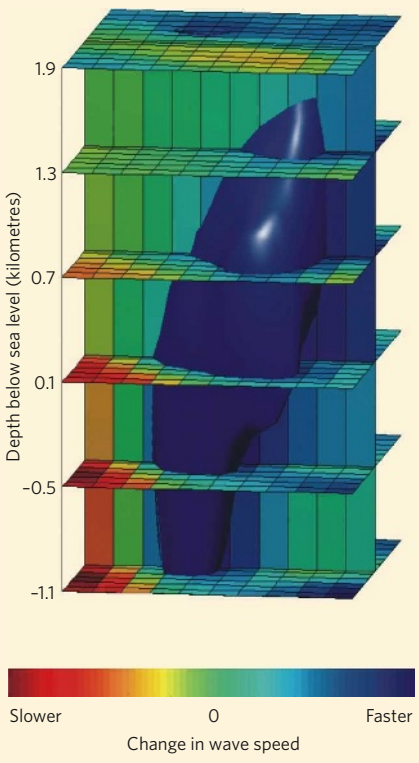
To investigate, Rob van der Hilst of the Massachusetts Institute of Technology in Cambridge and his colleagues set up 25 seismometer stations on the eastern edge of the plateau, then combined data from ambient noise and traditional tomography to produce a picture of the plateau down to a depth of 280 kilometres[5]. The clearest images, which came from ambient tomography done roughly 50 kilometres down, showed chopped-up areas where seismic waves travelled slower than usual. These low-velocity regions may be weak parts of the crust that have slowly flowed underneath and between other crustal layers. Although the images cover too small an area to be definitive, Lay says, "it's a step in the right direction".

Proponents of the flow model are encouraged. "I would say it's very new, exciting evidence. We didn't know there were low-velocity zones down there at all," says van der Hilst's colleague Leigh Royden, whose model of the plateau's uplift predicts crustal flow. The results also suggest that the model of crustal flow is more complicated than originally expected; the weak parts do not form a uniform, flowing layer, suggesting a complex network of conduits where the crust flows through the middle and lower mantle.

Later this year, van der Hilst plans to return to the area to collect data from a recently installed Chinese array of 300 seismometers. The additional data will help to improve resolution and might also help to differentiate between seismic speeds in different directions. This directional dependence, or anisotropy, may hint at material under strain, an indicator of flow.

IN THE HEART OF A VOLCANO

An ambient-noise image of the ground beneath the Piton de la Fournaise volcano on Réunion Island reveals a region where seismic waves sped up, delineating what may be a chimney of solid magma.



Early warning

The wealth of data offered by ambient noise tomography may also assist in more dynamic measurements, such as detecting precursors of volcanic eruptions. For the past several years, Campillo and his colleagues have set their sights on Piton de la Fournaise, an active volcano poised on the edge of the island of Réunion, near Madagascar. Piton de la Fournaise runs almost like clockwork: Averaged over the last 200 years, it has erupted every 10 months. Although most of the island's roughly 800,000 inhabitants live beyond the volcano's reach, past eruptions have overflowed the caldera, reaching the ocean as well as villages.

Over the years, seismologists have used Piton de la Fournaise as a test bed for several techniques that might help to predict future eruptions. Most rely on geodetics — sensors that measure strain, tilt or displacement on the surface of the volcano to detect slight changes, such as bulging from pent-up pressure. Seismometers are also set to look for any increase in shaking that might signal an impending explosion.

But in the absence of shaking, which sometimes starts just days or hours before an eruption, the volcano is a black box. "Seismic activity is less reliable for volcano eruption forecasting, because it can be linked to a lot of phenomena," says Florent Brenguier at the Institute of Earth Physics in Paris. "It is not a clear sign of an oncoming collapse."

Last year, Brenguier and his colleagues showed that 18 months of data from 21 seismometers scattered around the volcano could

REF. 6

©2008 Nature Publishing Group

be used to reconstruct its three-dimensional structure[6]. The images showed a unique structure that matched up with previous experiments: an anomalous area of high velocity, one kilometre east of the volcano's main vent, which seems to be a chimney full of solidified magma. Unexpectedly, comparing images separated by several months showed that seismic waves slowed down by a tiny amount — 0.1% — when travelling through this structure.

What's more, the waves started to slow down as early as 20 days before four eruptions that occurred between July 1999 and December 2000. Clear indicators of the change in speed showed up 5 days before the eruption. Why this happened isn't clear, but Brenguier says the speed changes probably match up with the opening of fissures within the edifice, through which the magma eventually flows to the surface.

But there are challenges in trying to use ambient noise to predict volcanic eruptions. For one thing, seismic waves in Piton de la Fournaise also show long-term changes in their speed, which may stem from mechanical changes due to tides or fluctuations in water level. "We don't know at the moment how to correct for these long-term variations or how they could be used for real-time monitoring," says Brenguier. The team's latest data, though, suggest that large changes in wave velocity correspond to larger eruptions, and down the line it might at least be possible to predict the size of an eruption[7].

Bernard Chouet, a volcanologist at the US Geological Survey in Menlo Park, California, questions whether warnings from ambient noise will ever be better than the few days' notice researchers typically get from the long-period seismic events generated when magma starts to gurgle in a volcano. But, he says, "all these things put together are going to put us closer to forecasting things more precisely". Extending the technique to volcanoes with more irregular eruptions will be a crucial test; such studies are already under way, including one at Mount Merapi in Indonesia.

"This is ultimately going to put the emphasis on processes rather than structures."
— Michael Ritzwoller

Graphic detail

In the years to come, the true bounty of ambient noise tomography may lie in high-resolution, three-dimensional maps of large swaths of Earth's crust, rather than having to wait for earthquakes. Some of the most detailed maps produced so far with this technique have come from the western United States.

In the past few years, researchers have studied ambient noise using USArray, which includes a rolling grid of 400 transportable seismometers that began mapping the west coast in 2004, but are moved eastwards each year. USArray was conceived before the advent of ambient noise tomography, but the dense array of detectors was perfectly suited for the technique. With proper analysis, the seismometers can combine to create 80,000 source-detector pairs that can be used to image the intervening Earth.

"One of the frustrations of seismology historically is that we have to wait for earthquakes as the source of energy, and they're most certainly not on a grid — they're in the Pacific Ring of Fire, in subduction zones or at plate boundaries," says seismologist Michael Ritzwoller of the University of Colorado in Boulder. The hum of ambient noise has revealed a number of features in fine detail, including subducting slabs, structures called mantle drips under the southern Sierra Nevada and evidence of the movement of Yellowstone's hotspot track across the Snake River Plain.

Like others, Ritzwoller is anticipating new results from anisotropy measurements, which may help illuminate stresses within the mantle and fractures and fault zones within the crust. "I think what this is ultimately going to do is put the emphasis on processes and not structures," he says. "It is actually a huge leap."

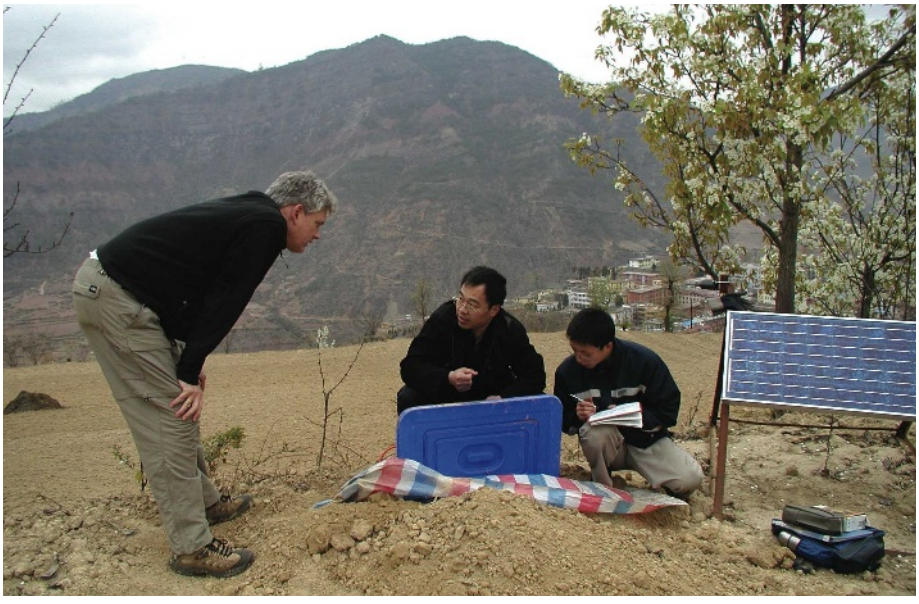
As rapidly as ambient noise tomography has been adopted, many researchers are still trying to perfect the analytical underpinnings of the technique. Challenges include figuring out how to account for the crashing of ocean waves that come from different directions, and for changes in topography at Earth's surface. Still, researchers are enlisting a growing number of seismometer arrays, including some in China, New Zealand, Korea, Japan, the United Kingdom, on glaciers and on the floor of the South Pacific Ocean. "I don't think the story's finished," says Campillo, "and it is not even really on the way."

Rachel Courtland is an intern reporter in *Nature*'s Washington DC office.

1. Weaver, R. L. & Lobkis, O. I. *Phys. Rev. Lett.* **87,** 134301 (2001).
2. Paul, A. & Campillo, M. *Eos* **82,** S21D-0610 (2001).
3. Shapiro, N. M. *et al. Science* **307,** 1615–1618 (2005).
4. Sabra, K. G., Gerstoft, P., Roux, P., Kuperman, W. A. & Fehler, M. C. *Geophys. Res. Lett.* **32,** L14311 (2005).
5. Yao, H., Beghein, C. & van der Hilst, R. D. *Geophys. J. Int.* **173,** 205–219 (2008).
6. Brenguier, F. *et al. Geophys. Res. Lett.* **34,** L02305 (2007).
7. Brenguier, F. *et al. Nature Geosci.* **1,** 126–130 (2008).

See Editorial, page 134.

C. Li



Rob van der Hilst and his colleagues use ambient noise and traditional tomography to image Earth's crust.

©2008 Nature Publishing Group

THE CELLULAR HULLABALOO

The inner life of a cell is noisy. **Helen Pearson** discovers how the resulting randomness makes life more challenging — and richer.

Sunney Xie's eight-year-old twins have a lot in common — starting with their genome. But they have different fingerprints, footprints and very different personalities. Any father might wonder how two apparently identical cells could turn into daughters fairly easily distinguished, at least by a parental eye. For a father who is also a biochemist at Harvard University in Boston, Massachusetts, that question is not just one for idle musing. "It's my biggest mystery," Xie says.

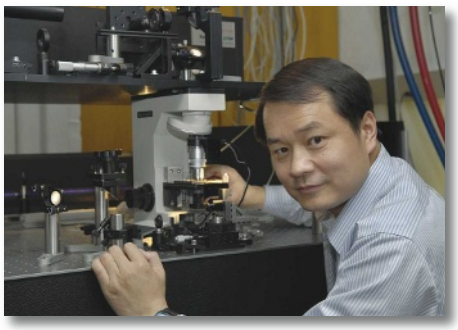
Two years ago, Xie's interest in 'identical' cells that act differently led him to produce the first comparisons of individual cells producing proteins one by one[1,2]. Although each of the cells studied under his cutting-edge fluorescent microscope produced the proteins he was looking for in bursts of up to 20 at a time, the sizes of the bursts and rhythm of their timing changed from cell to genetically identical cell. Xie wonders whether some such stochastic difference might have propelled the two identical cells that were to become his children down divergent biological paths.

In text books, molecular biology seems to leave little room for chance and random fluctuations. Cells are seen as exquisite machines, their component molecules fitting together to form mechanisms as neatly as any blind watchmaker could require. Genetic information is treated as software for this hardware to run, and the ways in which chance might make the software run differently in different cellular machines are seen as sources of error.

But in the past few years, some researchers have been exploring a far more haphazard picture of what goes on in the cell. Using techniques that allow one cell to be compared with another, they have found identical genes in identical cells doing wildly different things. Now researchers are trying to understand whether this variation or 'noise' actually matters. How hard do cells work to suppress it, what mechanisms do they have for tolerating it? A study from 2006 suggests that the workings of cells get noisier and noisier as they get older, and that this might contribute to the ageing body's decline[3]. "People are fascinated by how we do what we do despite this noise," says James Collins who studies stochastic processes at Boston University. "That's what everyone is ultimately trying to get at."

Noise in cells is inevitable. It traces back to the fact that each cell carries only one or two working copies of each gene. The DNA in which the genetic information sits is constantly shifting in both shape and structure, while jiggling proteins attach themselves to it and fall off again. At a given moment, the same piece of DNA may be bound by an activating protein in one cell, and not in another. In the first, a gene will be switched on, triggering the transcription of messenger RNA (mRNA) molecules and their subsequent translation into proteins. In the second, it will not, for the time being. There was no plan — it was all dependent on tiny changes in the configuration of the DNA, or the proximity of specific molecules. "You can never get away from the fact that it's all probability. Every event is up to chance," says Johan Paulsson, a systems biologist at Harvard Medical School. It is this that makes the noise in the cell so pervasive, and so interesting to those who wonder how cells control themselves. It is not just that the turning on and off of genes is noisy; the systems that regulate that turning on and off are themselves ineradicably noisy too. "Every chemical process is a constant battle between randomness and correction," says Paulsson.

An awareness of the part that chance might play has ebbed and flowed over the history of molecular biology, although many current researchers are "stupefyingly unaware" of the issue, according to Roger Brent, a molecular biologist at the Molecular Sciences Institute in Berkeley, California. The main difference between that earlier thought and today's interest, he says, is new tools that provide "all kinds of abilities to ask questions about single cells".



M. HALEVI

Sunney Xie has found that a random event can trigger one cell (yellow, below) to choose a different fate from that of its identical neighbours.

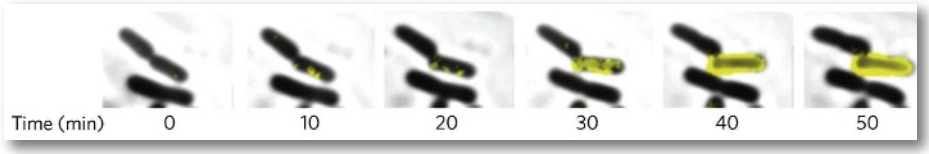
A million states

An early example of these tools is a 2002 study by Robert Singer at the Albert Einstein College of Medicine in New York. Singer's team used the relatively recently developed technology of fluorescent tagging to make four different coloured probes that bound themselves in different combinations to the mRNA transcripts of eleven different genes[4]. Each mRNA was thus given a different colour 'barcode'. The team showed that cells in a genetically homogeneous sample tended to each switch on a different combination of genes at different times. "No two cells seemed to be doing the same thing at the same time," Singer says. "Each cell could be in one of a million different expression states."

The amount of mRNA and protein produced by a population of cells typically forms a classic bell-shaped curve about the mean. The noisiness of a gene's expression is reflected in the width of the relevant bell curve. Some genes are quiet, with neatly peaked curves. Loud ones have their curves spread out. In the noisiest cases, the lowest performers may be making almost an order of magnitude less protein than the most productive cells.

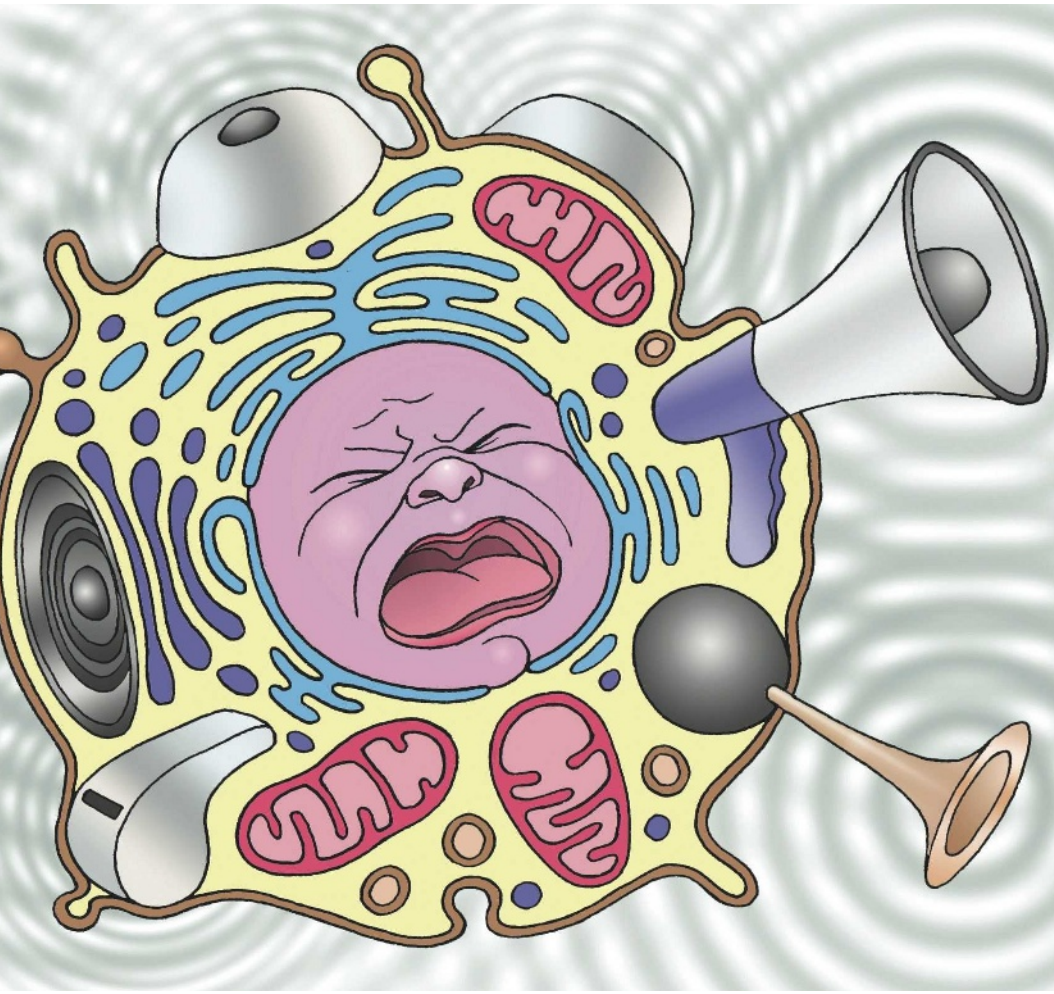
Simply doing away with all this noise is not an issue. In collaboration with Glenn Vinnicombe, an engineer at the University of Cambridge, UK, Paulsson has carried out theoretical work showing that noise abatement can take a huge amount of effort.

The pair modelled bacteria with varying numbers of DNA rings called plasmids. The plasmids control their own copy number by making repressor proteins that block plasmid replication. The more fluctuation there is in



P. CHOI

©2008 Nature Publishing Group



J. H. VAN DIERENDONCK

the number of repressors, the more fluctuation there will be in the number of plasmids: thus noise will tend to amplify itself. At the same time there is a very strong incentive for the plasmids to clamp down on the noise by producing more repressor proteins — because the more wildly the plasmid number swings, the greater its chance of that number dropping to zero. Selection will favour plasmids that avoid going extinct in this way.

Turning down the volume

According to Paulsson and Vinnicombe's unpublished model, a cell must produce enormous numbers of repressor proteins in order to control plasmid number noise — producing, for example, 160,000 repressors per cell division to make it possible to keep four plasmids to within a 5% standard deviation. Only by manufacturing so many repressors can the cell ensure that the inhibitor levels provide a faithful read-out of the underlying plasmid levels, so that there are enough around to control plasmid number. (Real versions of this type of plasmid use some of the strongest promoters found in nature to turn out many thousands of inhibitors per cell cycle.) Paulsson's theoretical work has convinced him that only the most critical cellular processes are worth the large investment it takes to keep noise under control. "If you want to do a good job you have to spend a huge amount of energy," he says. "If you want to do a half-decent job you pay much less, but it's still surprisingly expensive."

If removing the noise is not feasible, cells need to make a trade-off between the effort that is required to lower the noise levels and the effort that's required to live with them. The default noise level would thus be, almost by definition, the loudest that's bearable — any higher and the system breaks down. "I think in lots of cases the noise we see is simply the best the cell can do," says Paulsson. "People ask how come an organism works so well. Perhaps it doesn't work so well. Perhaps organisms without these fluctuations would outcompete us."

On that analysis, noise is simply something to be put up with. But another school of thought suggests that it might be more than that. It may be that for some purposes cells have learned to like it noisy. Rather than using energy to minimize all sources of noise or insulate all systems from its effects, the cells may be using the noise to their own ends.

"People are fascinated by how we do what we do despite this noise."
— James Collins

Collins is one of those who sees advantages to noise. His studies over the past 14 years have shown that mechanical and electrical noise applied to neurons can help them respond to weak incoming signals, because they are already partially excited. He also showed, on the same basis, that vibrating insoles can help elderly people balance, and has applied to patent the technology.

In 2006, Collins's team described engineering mutations into the control region of a gene that confers antibiotic resistance to create two strains of the yeast *Saccharomyces cerevisiae*, one with noisier expression of the gene, one with something more steady. Faced with a lethal antibiotic, the noisier strain survived better[5]. This result supports the idea that noise is a form of 'bet hedging' for cells: a population is more likely to survive in a changing environment if its members are noisy because some are likely to be making the quantity of a protein best suited to that situation. "A system that is covering more possibilities has a greater chance of survival in unpredictable settings," says Collins.

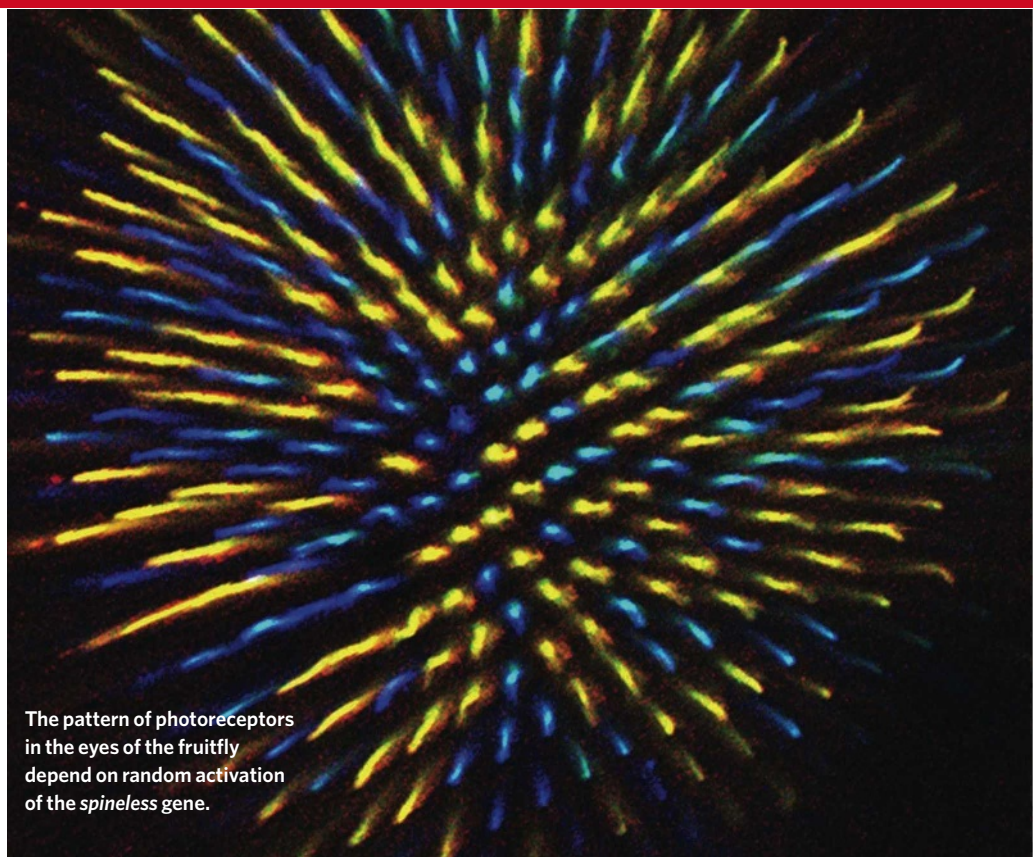
Two-colour eyes

Another situation in which cells may have used the stochasticity of gene expression to their advantage is to create a scattered pattern of photoreceptors in the developing retina of the fruitfly *Drosophila*. Here, some 30% of cell clusters called ommatidia include cells that are sensitive to blue light and 70% include those sensitive to green light. Claude Desplan at New York University and his colleagues showed that cells in the developing retina randomly switch on the *spineless* gene, making a transcription factor that controls other genes' activity[6].

Once on, *spineless* triggers a feedback loop that ensures the cell making it, and the one next door, together form a green-light sensitive ommatidium; those that do not turn on *spineless* make a blue-sensitive ommatidium (see picture, overleaf). Desplan says that evolution has taken advantage of small and random asymmetries between cells to ensure that the photoreceptors are spread across the retina. "Stochasticity makes life very difficult but it must have an evolutionary advantage," he says.

Other researchers are not yet convinced that noise has such biological value. "You can always make up a story of how cell-to-cell variation could be an advantage," says Brent. Noise could present a major obstacle when information is at stake — for example when a cell is measuring conditions outside its walls and needs to transmit that information faithfully to the nucleus. When yeast cells receive an external signal in the form of a pheromone, Brent has shown that their responses (such as recruiting signalling proteins to the cell membrane) are precisely calibrated to the strength of the signal, thus ensuring that they switch on pheromone-response genes in proportion to the pheromone's strength. But in a mutant strain, the signal and response are no longer perfectly aligned, and the cell's pheromone

©2008 Nature Publishing Group



The pattern of photoreceptors in the eyes of the fruitfly depend on random activation of the *spineless* gene.

D. VASILIAUSKAS

response becomes more unpredictable. "Noise is actively worked around," says Brent, which suggests that evolutionary pressure to ensure signals are transmitted in noise-resistant ways must be strong.

Cellular noise could also pose a problem in the development of an embryo, when cells have to act together to produce precisely timed and coordinated events. "How does a cell control that and make sure it doesn't take a wrong exit at the wrong time? That's a very interesting question," says Alexander van Oudenaarden, a physicist at the Massachusetts Institute for Technology in Cambridge. Working with *Caenorhabditis elegans* embryos, van Oudenaarden is examining the type of developmental mutations that are puzzling because they cause an obvious defect in some animals, but leave others intact. His idea is that fluctuations in noisy genes could explain this behaviour, nudging only some embryos down the mutant developmental pathway. A similar phenomenon might explain why some people with high-risk mutations that predispose them to breast cancer or another disease actually escape unscathed.

The chance of ageing

If noise-abatement is important in suppressing disease, its breakdown may have particular ramifications at the end of our lifespan. Jan Vijg of the Buck Institute for Age Research in Novato, California, measured the expression level of 15 genes in individual muscle cells plucked from the hearts of young, six-month-old mice and elderly 27-month olds[3]. The variation in gene expression was around 2.5 fold larger in older cells when compared with younger ones, showing that biological noise increased with ageing.

Vijg suggests that heart cells accumulate wear and tear to their DNA, and that this upsets the systems the cells normally use to keep noise reined in, a phenomenon that may contribute to the gradual decrease in heart function with age. "There is a lot of evidence that ageing has this stochastic component," he says. If it is true that ageing interferes with the control of almost all genes, then drugs or treatments that can reverse the ageing process may be unrealistic, Vijg says. "It will be very difficult to intervene in a basic process of life that you can't get rid of."

Vijg and other researchers now wonder whether there are genes and proteins the main function of which is in controlling or suppressing biological noise. Chaperones, proteins that help other proteins fold into the correct shape, are one candidate. They might help compensate for volatile protein concentrations by stalling the folding process until all of a system's components are present in the numbers needed to form a working piece of machinery.

Whereas some are working on the controls over noise, others are looking at its causes. In one of the first such experiments, van Oudenaarden studied genetically identical bacteria making a green fluorescent protein[7]. His group engineered mutated strains in which either transcription of DNA to RNA or translation of RNA to protein was altered — and showed that the noisiness in protein output could be changed systematically.

This and other studies have led to the idea that most noise stems from the rarer events during transcription — typically the infrequent activation of a gene or the slow, fitful production of mRNA molecules. If only five mRNAs for a given protein are present in each cell, then the creation or destruction of one or two extra can produce large variations relative to the mean. If, by contrast, a cell contains 500 copies of an mRNA molecule then the addition of a few extra makes little difference. Michael Elowitz, one of the leaders in the field at the California Institute of Technology in Pasadena, makes the comparison to raindrops falling on a roof. When rain is light, the numbers hitting each roof tile can vary dramatically; but when rain is heaver, those fluctuations are washed away. This means that you need a lot of Paulsson's repressors around to stop these variations in number from having a drastic effect.

> **"You can never get away from the fact that it's all probability. Every event is up to chance."**
> — Johan Paulsson

Experiments such as van Oudenaarden's have left some researchers wanting more. The total concentration of protein in one cell is itself the result of many separate stochastic moments, separate events in which mRNA and then protein were made one by one. Some scientists wanted to eavesdrop on this process. "We realized we needed to go one step higher in resolution," says Ido Golding, a physicist at the University of Illinois in Urbana-Champaign. This has been a technological challenge, because fluorescent cameras are usually only sensitive enough to detect the signal coming from 20 or 30 fluorescent proteins in one spot — and the signals can be blurred by diffusion of the protein.

Random bursts

In 2005, Golding described a way around this resolution problem, working at the time in the lab of Edward Cox at Princeton University, New Jersey. The researchers introduced a repetitive motif into one *Escherichia coli* mRNA molecule so that each molecule was bound by 50–100 copies of green fluorescent protein[8]. As each mRNA molecule was produced, enough glowing proteins latched onto it to see it in real time.

As they watched, the gene suddenly switched on and produced an intense burst of mRNAs rather than, as most had thought, constant, steady production. The group calculated that this gene would be off for an average of 37 minutes and then on for 6 minutes, during which time it would produce anywhere from one to eight RNA molecules. The bursts may be caused by random changes in the conformation of DNA that permit transcription, or the random binding and detaching of proteins

©2008 Nature Publishing Group

that activate or repress transcription. And it is these bursts, which vary both in length and in the numbers of molecules produced, that eventually produce much of the wide variation in protein levels between cells.

According to Singer, only certain genes may be prone to bursts of activity. Other genes may be permanently accessible for transcription, which proceeds at a more leisurely and continuous pace. For these 'housekeeping' genes, involved in running mitochondria, say, or cellular transport, it may be too risky for cells to leave production to random bursts. For others, such as those involved in responding to a rapidly and unexpectedly changing environment, it may be too risky not to.

Xie found echoes of burst activity in his studies of proteins[1,2], where each burst corresponds to the brief lifespan of a single mRNA before it decays. He was able to visualize single proteins because he studied some of those which are anchored in the cell membrane. Being secured in place, their fluorescent signals are not blurred by diffusion. And in the last two years, Xie has edged a little closer to understanding how his two genetically identical daughters became who they are. He has shown how the stochastic behaviour of a single molecule can drive two genetically identical cells to adopt different fates.

Xie's team studies bursts in the production of LacY, a membrane channel through which the sugar lactose enters into *E. coli*. When lactose is present at low levels, a repressor protein binds DNA and prevents production of the channel and other proteins involved in lactose metabolism. When lactose levels are high and enough has entered the cell, the repressor detaches and the cell switches into a state where it can metabolize and use the sugar. This classic system of gene regulation was first studied by Francois Jacob and Jacques Monod in the 1960s.

Xie's graduate student Paul Choi tracked the channel proteins and the repressor. He found that, now and then, the repressor protein drops off the DNA entirely and takes a minute or so to latch back on. This allows production of mRNA followed by a massive burst of 200–300 channel proteins. Lactose floods into the cell, preventing the repressor from rebinding and triggering lactose metabolism.

"How come an organism works so well? Perhaps it doesn't. Perhaps organisms without these fluctuations would outcompete us."
— Johan Paulsson

So a chance event — a single repressor protein dropping off DNA — thrusts the cell into an entirely different state, or phenotype. And by chance, only a proportion of cells in a population will trigger this switch. This may work to the bacteria's advantage when lactose concentrations are too low to support the entire population of cells, because only some will flip into a state where they can take up and make use of the sugar. "Stochasticity of a single molecule solely determines a life-changing decision of the cell," Xie says. "I don't think it's a stretch to think that these single molecule events will determine more complex phenotypes as well."

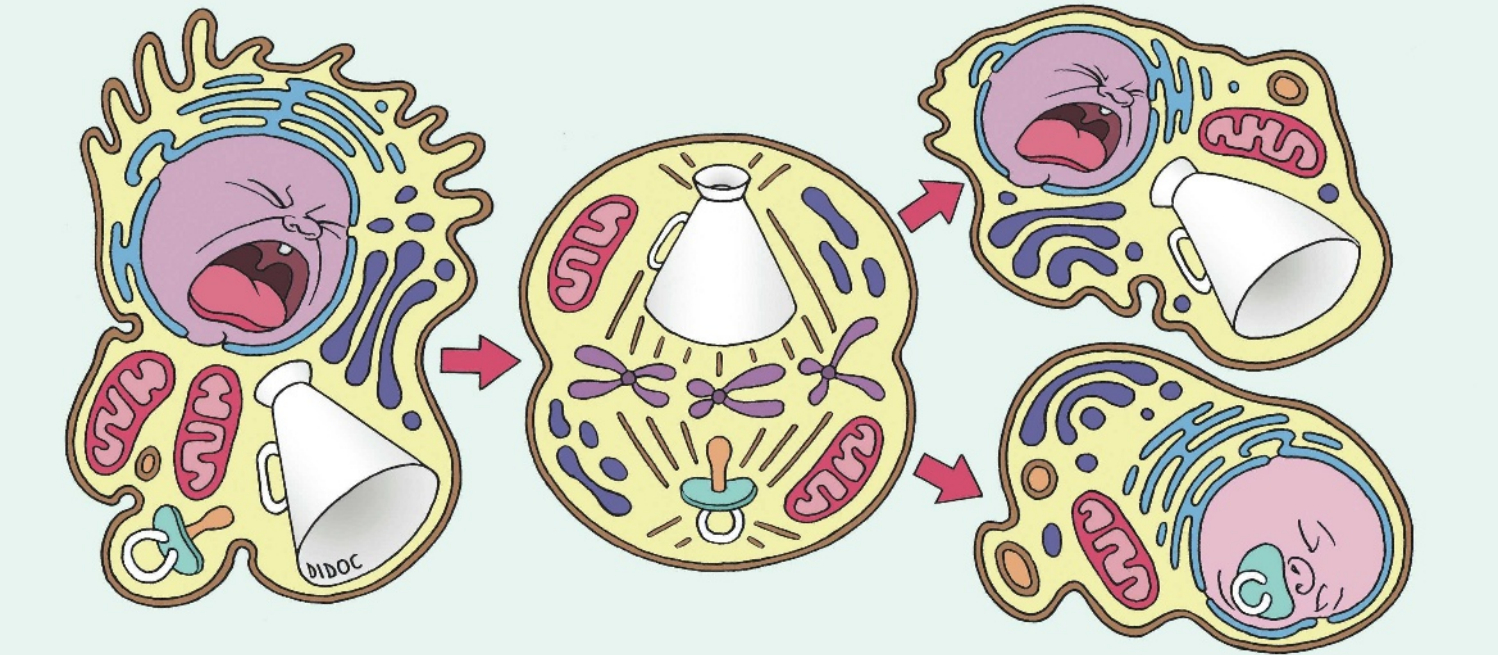
After Xie published his protein-bursting papers in *Science* and *Nature* in 2006, he says he received a surprise phone call from a programme manager at the Bill & Melinda Gates Foundation, in Seattle, Washington, asking him to think about whether his work could somehow help in their fight against tuberculosis. So Xie thought — and he is now embarking on a project to investigate whether stochastic events such as those he has observed in *E. coli* might explain why, from a population of genetically identical bacteria, only some will become extremely drug resistant.

Stochastic events can turn two genetically identical bacteria into ones that behave differently. Can the same really be said of Xie's genetically identical twins? They are, after all, far more complicated organisms than the unicellular types under his microscope. "In very simple cells we have shown for sure that a single stochastic event can lead to different phenotypes," Xie says. For his daughters, "I can't say for sure it's due to that". ■

Helen Pearson is *Nature*'s biology features editor.

1. Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X. S. *Science* **311,** 1600–1603 (2006).
2. Cai, L., Friedman, N. & Xie, X. S. *Nature* **440,** 358–362 (2006).
3. Bahar, R. *et al. Nature* **441,** 1011–1014 (2006).
4. Levsky, J. M., Shenoy, S. M., Pezo, R. C. & Singer, R. H. *Science* **297,** 836–840 (2002).
5. Blake, W. J. *et al. Mol. Cell* **24,** 853–865 (2006).
6. Wernet, M. F. *et al. Nature* **440,** 174–180 (2006).
7. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. *Nature Genet.* **31,** 69–73 (2002).
8. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. *Cell* **123,** 1025–1036 (2005).

See Editorial, page 134.



J. H. VAN DIERENDONCK

©2008 Nature Publishing Group

CORRESPONDENCE

These letters respond to the Commentary 'Dangerous assumptions' by R. Pielke Jr, T. Wigley and C. Green (*Nature* **452,** 531–532; 2008) on the scenarios of the Intergovernmental Panel on Climate Change (IPCC).

B. MELLOR

Long-range energy forecasts are no more than fairy tales

SIR — I largely agree with the overall conclusion of Pielke *et al.* that the IPCC assessment is overly optimistic, but I fear that the situation is even worse than the authors imply.

Debate about the energy challenge of climate change has ignored several key facts about global energy and its future.

All long-range energy forecasts fail in various ways. The scenarios for 2100 in the 2000 IPCC *Special Report on Emissions Scenarios* are risible: even if one of them were to approximate actual demand, we will remain ignorant of what its regional and sectoral composition would be a century from now — and it is this that will determine emissions of greenhouse gases. Basing policies on computerized fairy tales is inadvisable.

Why argue about plausible rates of future energy-efficiency improvements? We have known for nearly 150 years that, in the long run, efficiency gains translate into higher energy use and hence (unless there is a massive shift to non-carbon energies) into higher CO_2 emissions.

The speed of transition from a predominantly fossil-fuelled world to conversions of renewable flows is being grossly overestimated: all energy transitions are multigenerational affairs with their complex infrastructural and learning needs. Their progress cannot substantially be accelerated either by wishful thinking or by government ministers' fiats.

Carbon sequestration is irresponsibly portrayed as an imminently useful large-scale option for solving the challenge. But to sequester just 25% of CO_2 emitted in 2005 by large stationary sources of the gas (9.6 Gm^3 at the supercritical density of 0.468 $g\,cm^{-3}$), we would have to create a system whose annual throughput (by volume) would be slightly more than twice that of the world's crude-oil industry, an undertaking that would take many decades to accomplish.

China, the world's largest emitter of CO_2, has no intention of reducing its energy use: from 2000 to 2006 its coal consumption rose by nearly 1.1 billion tonnes and its oil consumption increased by 55%.

Consequently, the rise of atmospheric CO_2 above 450 parts per million can be prevented only by an unprecedented (in both severity and duration) depression of the global economy, or by voluntarily adopted and strictly observed limits on absolute energy use. The first is highly probable; the second would be a sapient action, but apparently not for this species.

Vaclav Smil University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada

Energy assumptions were reasonable at the time, but not now

SIR — Pielke *et al.* correctly point out that almost all of the IPCC emissions scenarios assume continuing improvements in the energy intensity of the global economy and in the carbon intensity of the global energy system, even in the absence of explicit climate policy. This assumption is not hidden, as Pielke *et al.* imply, and it is not unreasonable.

Over several decades before 2000, that is exactly what happened. Improved manufacturing and buildings, increased efficiency in power production, and increasing economic activity in services and other low-energy activities, all contributed. Especially with historically high energy prices, there is every reason to expect future improvements in the efficiency of energy generation and use, improvements that don't depend on climate policies. Consistent with its mandate, the IPCC developed scenarios that build on historical trends to characterize future possibilities.

Since 2000, the historical patterns have not held. The carbon intensity of the energy system and the energy intensity of the economy have both increased, driven partly by a surge in new electricity from coal, and partly by rapidly accelerating construction and manufacturing, especially in China and India. This trend, if it continues, will greatly increase the challenge of meeting an atmospheric concentration goal of 500 parts per million, or any other level. Pielke *et al.* are right to emphasize the stark contrast between the pattern before 2000 and that in the past few years.

The trends towards increased carbon and energy intensity may or may not continue. In either case, we need new technologies and strategies for both

©2008 Nature Publishing Group

"We need a better understanding of how the alchemy of music depends on texture." see page 160

endogenous and policy-driven intensity improvements. Given recent trends, it is hard to see how, without a massive increase in investment, the requisite number of relevant technologies will be mature and available when we need them.

Christopher B. Field Carnegie Institution for Science, Stanford, California 94305, USA

Future scenarios for emissions need continual adjustment

SIR — Pielke *et al.* show that the 2000 *Special Report on Emissions Scenarios* (*SRES*) reflects unrealistic progress on both the supply and demand sides of the energy sector. These unduly optimistic baselines cause serious underestimation of the costs of policy-induced mitigation required to achieve a given stabilization level.

This is well known among experts but perhaps not to the public, which may explain why some politicians overstate the impact of their (plans for) climate policy, and why others argue incorrectly that 'available' off-the-shelf technologies can reduce emissions at very little or no cost.

The numbers presented by Pielke *et al.* are revealing, but they divert attention from a more serious problem underlying the *SRES* approach to calculating mitigation costs: a failure to incorporate the dynamic nature of the decision problem into climate-policy analysis. Until we can keep adjusting the analysis by continually incorporating uncertainty, correction and learning, we shall continue to offer policy-makers an incomplete guide to decision-making.

The focus of policy analysis should not be on what to do over the next 100 years, but on what to do today in the face of many important long-term uncertainties. The minute details of any particular scenario for 2100 are then not that important. This can be achieved through an iterative risk-management approach in which uncertain long-term goals are used to develop short-term emission targets. As new information arises, emission scenarios, long-term goals and short-term targets are adjusted as necessary. Analyses would be conducted periodically (every 5–10 years), making it easier to distinguish autonomous trends from policy-induced developments — a major concern of Pielke and colleagues. If actual emissions are carefully monitored and analysed, the true efficacy and costs of past policies would be revealed and estimates of the impact of future policy interventions would be less uncertain.

Such an approach would incorporate recent actions by developed and developing countries. In an 'act then learn' framework, climate policy is altered in response to how businesses change their behaviour in reaction to existing climate policies and in anticipation of future ones. This differs from *SRES*-like analyses, which ignore the dynamic nature of the decision process and opportunities for mid-course corrections as they compare scenarios without policy with global, century-long plans.

Richard G. Richels Electric Power Research Institute, 2000 L Street NW, Suite 805, Washington DC 20036, USA
Richard S. J. Tol Economic and Social Research Institute, Whitaker Square, Sir John Rogerson's Quay, Dublin 2, Ireland
Gary W. Yohe Department of Economics, Wesleyan University, 238 Church Street, Middletown, Connecticut 06459, USA

Climate policies will stimulate technology development

SIR — Roger Pielke and his colleagues argue that the IPCC "seriously" underestimates the scale of the technological changes required to stabilize greenhouse-gas concentrations, and hence conveyed an inappropriate message to policy-makers on policies required for mitigation. We believe that this argument is based on a flawed analysis.

The authors repeat a 'thought experiment' done by the IPCC, to find out what would happen over the next century if technology were frozen at present levels. In that case, the emission reductions required to reach any of the assessed greenhouse-gas stabilization levels would far exceed those computed by any of the mitigation scenarios reviewed in the IPCC fourth assessment report (AR4).

It is most unlikely, however, that technology will be frozen. Over the past 30 years, the decrease in energy intensity has been 1.1% a year — well above the 0.6% a year assumed in 75% of the energy scenarios assessed by the IPCC.

Developments in China since 2000 do raise concerns that the rate of decrease in energy and carbon intensity could slow down, or even be reversed. However, similar short-term slow-downs in technical progress have occurred in the past, only for periods of more rapid development to compensate for them. India, for example, does not show the decreasing trend in energy efficiency seen in China.

The increase in fossil-resource prices triggered by high economic growth will lead to an increase in energy efficiency. Admittedly, a possible increase in carbon intensity caused by a renaissance of coal is a worst-case scenario for any climate policy. But the impact of increasing fossil-fuel prices on technological change and on mitigation costs, or policies, cannot be analysed in any meaningful way for policy-makers by assuming a 'frozen technology' scenario.

The IPCC's main policy conclusions stand: present technologies can stop the rise in global emissions. But they will depend on governments' policies to ensure that the technologies reach the market in time. A carbon price on emissions will promote investment in lower-carbon technologies, and climate policies will stimulate technological development to bring emissions and mitigation costs down further.

Ottmar Edenhofer, Bill Hare, Brigitte Knopf, Gunnar Luderer Potsdam Institute for Climate Impact Research, Germany

IPCC's climate-policy assumptions were justified

SIR — Pielke *et al.* suggest that the IPCC underestimates the challenge of global warming. I find their analysis misleading.

They criticize the IPCC for implicitly assuming that the challenge of reducing future emissions will mostly be met without climate policies. But the IPCC's *Special Report on Emissions Scenarios* makes clear that, although the scenarios don't technically have climate policies, they can and do have energy-efficiency and decarbonization policies, which amount to the same thing (see IPCC reference emission scenario B1, which includes aggressive policies to help limit total global warming to about 2 °C). So advances towards reducing emissions are indeed policy-driven.

The authors also caution the IPCC against assuming that spontaneous advances in technological innovation will be instrumental in cutting future emissions. They claim that the IPCC is actually diverting attention away from policies that could stimulate technological innovation, pointing out that enormous advances in energy technology will be needed to stabilize atmospheric carbon dioxide to acceptable concentrations. This claim is unjustifiable: in fact, the IPCC report makes clear that we have the necessary technologies, or soon will, and focuses on creating the conditions for rapid technological deployment.

Joseph Romm Center for American Progress, 1333 H Street NW, Washington DC 20005, USA

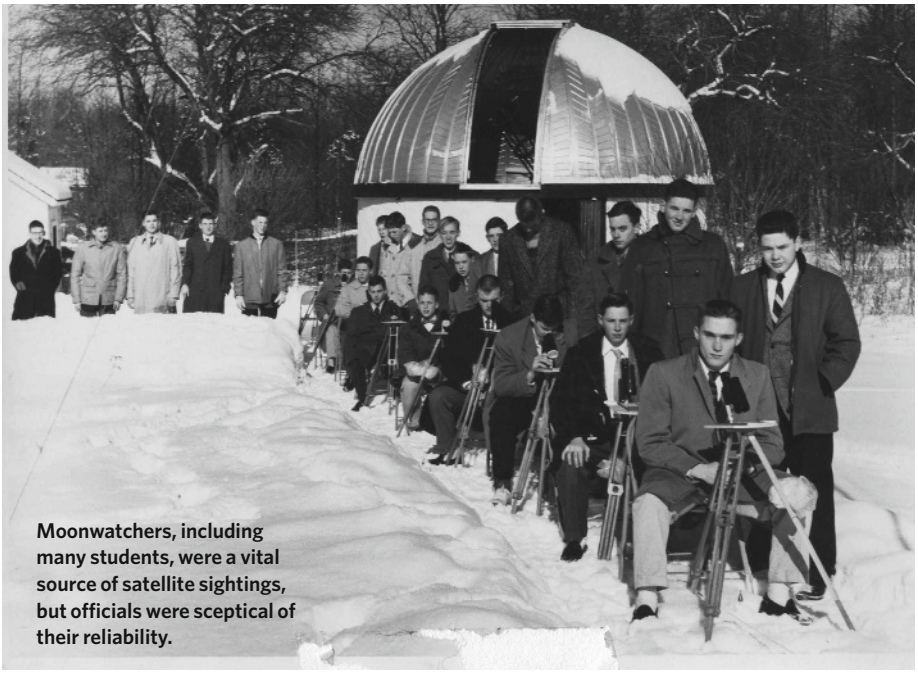
©2008 Nature Publishing Group

BOOKS & ARTS

Not so amateur

Volunteer star-gazers tracking satellites at the start of the space age often surpassed the professionals.

Keep Watching the Skies! The Story of Operation Moonwatch and the Dawn of the Space Age
by W. Patrick McCray
Princeton University Press: 2008. 324 pp. $29.95, £17.95



Moonwatchers, including many students, were a vital source of satellite sightings, but officials were sceptical of their reliability.

Technological advances come in such small increments that we rarely think about their accumulated effect. Today, it is hard to imagine a world without photocopiers, colour television, mobile phones or e-mail. Or satellites, imperceptibly circling overhead, ignoring national boundaries, spying on hurricanes, solar activity and terrorist encampments, positioning us globally.

In *Keep Watching the Skies!*, Patrick McCray reconstructs an era when the world was taking its baby steps into the space age. He views it through the eyes of amateur star-gazers who experienced the excitement of those Sputnik days by joining Moonwatch, a worldwide effort to track the satellites that were launched for the International Geophysical Year of 1957–58.

Fred Whipple, the energetic new director of the venerable and nearly moribund Smithsonian Astrophysical Observatory (SAO), oversaw the observatory's move north from Washington DC to join Harvard College Observatory in Cambridge, Massachusetts, and propelled the SAO into the space age. Overriding the doubts of many of his peers, it was Whipple's vision to engage amateurs across the United States, and then around the world, to watch for the passage of satellites. At each location, he proposed the creation of an 'optical fence' across the sky, using a row of modest wide-field telescopes to spot any satellite that might cross the celestial meridian.

Whipple's stubbornness in defending the inexpensive Moonwatch programme paid off when the unexpected launch of the Russian satellite Sputnik 1, and shortly thereafter Sputnik 2, occurred before more elaborate satellite-tracking programmes were ready. The Moonwatchers' observations gave the SAO just enough data to calculate the orbits of the satellites and to follow their decay. Three weeks after its launch, the batteries of Sputnik 1 died, rendering it silent to the multimillion-dollar radio tracking stations and amateur radio operators. The observations of the amateur Moonwatch teams suddenly became the main source of information.

Despite the crucial role of the Moonwatch network, Whipple faced opposition from administrators who were sceptical of the reliability of amateurs, an encounter vividly described by McCray. But Whipple's confidence was vindicated when the amateur teams spotted a lost, radio-silent Vanguard satellite launch vehicle in the skies in May 1958. A missing satellite was found by another Moonwatch team in Adelaide, Australia, on New Year's Day 1959; when the satellite got lost again, Moonwatchers searched for six months before finding it once more.

Writing the history of a widespread and amorphous group of amateurs was not simple. Many file boxes of communications from the Moonwatch teams survive in the Smithsonian Institution Archives. But it was a challenge to locate and integrate the more personal aspects of recruiting team members and sustaining their enthusiasm in the cold, pre-dawn hours when there was often nothing to be seen. McCray went beyond the official documents, ferreting out records from several of the most effective team leaders, and spotlights these throughout his well-illustrated presentation. At times the story seems rambling or padded, but at its best, *Keep Watching the Skies!* becomes a genuine page-turner.

The high points of McCray's narrative are the mini-biography of Whipple, leading up to his visionary defence of the Moonwatch proposal against its detractors, and the account of those fraught days after Sputnik's launch in 1957, when it took a heroic scramble to make the programme work in real time. The tale is always related from the perspective of the Moonwatch teams, with almost nothing about the space technicians and computer novices at the SAO who were caught equally unprepared. McCray's account is an important contribution towards preserving the history of a fascinating episode at the dawn of the space age, but other such stories are yet to be told.

McCray ends his book wistfully, wondering if there can ever again be such a public outreach on this scale by scientists, and whether amateurs can be marshalled to play a vital part in science today. "Will curious citizens have the opportunity to learn about the science that explicates global climate change and contribute to research in a meaningful way?" he asks. I would guess that it won't be citizen star-gazers who continue their skyward observations, but more likely amateur naturalists — looking not for satellites, but for birds. An ongoing census, which includes all animals and plants, might allow tracking of gradual or swift changes in our terrestrial environment. There are still observations to be made. ■

Owen Gingerich is professor emeritus of astronomy and history of science at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts.

©2008 Nature Publishing Group

Learning from climates past

Fixing Climate: What Past Climate Changes Reveal About the Current Threat — and How to Counter It
by Wallace S. Broecker and Robert Kunzig
Hill and Wang: 2008. 272 pp. $25

As we are alerted to the perils of climate change, we need a long-term perspective to understand the capabilities of Earth's climate. Looking to the past can help us evaluate the risks and judge the best way to manage them.

Fixing Climate gives a personal account of scientific endeavours to understand past, present and future climate change. Geoscientist Wallace Broecker at the Lamont–Doherty Earth Observatory of Columbia University, New York, has teamed up with journalist Robert Kunzig to argue that the current crisis is urgent but not insurmountable, requiring technological development and political leadership. The authors use the analogy that dumping greenhouse gases into the atmosphere is akin to the sewage problem faced in the nineteenth century, an issue that also invited vehement debate at the time.

Few scientists have been as active in their work and outreach as Broecker. Starting with his upbringing and early career, Broecker explains what inspired him and outlines his tremendous contributions to climate science. Describing in an accessible way how Earth's climate system worked in the past, Broecker and Kunzig look at future climate predictions and the technological developments that may help to scrub the air clean.

Along the way, we meet some extraordinary characters. In 1896, the Swedish scientist Svante Arrhenius proposed that a doubling of carbon dioxide levels in the atmosphere would drive global temperatures up by nearly 6 °C — close to the predictions made by today's climate models. Arrhenius did not foresee how quickly we would pump greenhouse gases into the air and took a benign (if naive) view of climate-change effects: "Our descendants, albeit after many generations, might live under a milder sky and in less barren natural surroundings than is our lot at present."

The broad topical coverage of *Fixing Climate* is admirable, yet there is a strong US emphasis. Climate-change adaptation and mitigation are not easily covered in a single popular text, as the authors admit. The book describes promising developments for stripping the air of excess carbon dioxide, but other avenues are less thoroughly explored. Little is made of the underlying problem that we are hooked on a high-energy lifestyle. Sustainable energy sources are underplayed despite the recent advances made, for example, with biofuels produced from algae and thin-film solar-power cells that are cutting the cost of these technologies.

There are some intriguing omissions from the book, notably the idea that humans may have influenced the climate for longer than we thought. In 2003, William Ruddiman of the University of Virginia, Charlottesville, proposed that more than 8,000 years of agriculture have raised atmospheric greenhouse-gas levels, warming the atmosphere by some 0.8 °C. This work is not discussed, even though it supports the underlying thesis of *Fixing Climate* that the past is the key to the future.

The book is a call to arms for us to take responsibility for our fossil-fuel dependency. The climate has changed; now let's fix this open sewer of the twenty-first century. ■

Chris Turney is professor in physical geography at the University of Exeter, Exeter EX4 4QJ, UK, and author of *Ice, Mud and Blood: Lessons from Climates Past*.

Mountains into molehills

Burning the Future: Coal in America
Screens on the Sundance Channel and is released on DVD on 13 May.
Mountain Top Removal
At the Fine Arts Theater, Asheville, North Carolina, on 15 May; on DVD in December.

Coal provides half of the United States' electricity. Increasing demand and eased regulation are fuelling coal-mining operations such as mountain-top removal, in which miners access coal seams, without digging below the surface, by dynamiting the summits of mountains and dumping the blasted rock in adjoining valleys.

Two new documentaries — *Burning the Future: Coal in America* by director David Novack and *Mountain Top Removal* from producer and director Michael O'Connell — investigate how mines affect health and the environment. They focus on the Appalachian mountains just inland of the eastern US seaboard, an area known for its traditional music, rural poverty and religious fervour.

The sight of demolished mountain tops is dramatic and well suited to the screen. Both films feature aerial shots of the mines, which look as if someone has skinned the top halves of mountains down to the rock, then snapped off the peaks. Residents such as Ed Wiley of Coal River, West Virginia, are understandably upset: "Once these mountains are gone, there is no more Appalachia; there is no more West Virginia. It don't grow back."

Mining companies move mountains to get to coal.

A. CASKEY, WWW.ANTRIM-NEW.COM

In 2005, the Environmental Protection Agency estimated that 6.8% of forested land in Appalachia will be affected by mountain-top mining by 2012. According to *Mountain Top Removal*, 207,000 hectares of mountains have been destroyed and 1,900 kilometres of mountain streams have been buried. Mining companies are required to do some remediation of mined-out surface sites, although this is often waived, and in any case does not return the area to its former state.

Both films describe the mines as causing increased flooding in the valleys below, and polluting local water with toxins, including arsenic and lead. Chilling but anecdotal evidence charts the effects on the health of local people and on the school children of Marsh Fork elementary school in West Virginia. This school and many local individuals appear in both films.

Choosing to focus on the wound-like alteration to the landscape, the film-makers barely mention climate change, probably the most dire and far-reaching effect of the US coal habit. Nor do they discuss arguably worse mining methods elsewhere, like hard-rock mining for minerals such as gold and copper, which is undertaken on a much larger scale.

Both films stir up a sense of loss and outrage. But neither offers solutions — political or technical — beyond suggesting that coal companies resume underground mining, which they are unlikely to do voluntarily as it is less efficient. Coal companies would rather keep opening up a mountain like taking the top off an egg. ■

Emma Marris is a correspondent for *Nature*.

©2008 Nature Publishing Group

Saving art *in situ*

A conservation scientist explains how borrowing gadgets from Mars rovers helps preserve culture on Earth.

Art conservation protects our cultural heritage by preventing loss. During the past 50 years, conservators have respected the material integrity of objects, adopting an approach of minimum intervention. Removing even small samples from works of art for scientific analysis is acceptable only when no other technique can provide the necessary information. Until recently, few analytical choices were available that did not necessitate sampling, but the development of non-invasive, portable instruments has greatly changed the art conservation landscape.

Conservation scientists apply analytical procedures to important works of art and architecture to establish what an object is made of and how it was made (technical art history), as well as when and where it was created (dating and provenance). Research also reveals how materials change over time, and suggests ways to slow their deterioration. This information helps art historians, archaeologists and conservators to better understand the objects they are studying and preserving.

Much is to be gained from bringing the analytical instrument from the laboratory to the work of art, rather than the reverse. It is inherently risky to move art objects: insurance companies charge high fees to cover damage during transportation, often prohibiting the movement of an object for study, and some items are simply not transportable. Portable instruments permit on-the-spot brainstorming among conservators, museum curators, art historians, site managers and others, allowing good solutions to be reached and such partnerships to be strengthened.

Many items of analytical equipment are now produced in miniaturized portable versions. One example is an X-ray fluorescence spectrometer, used to determine the elemental composition of a material. X-ray fluorescence has limitations because it cannot tell you the absolute quantities of each element, does not distinguish between various layers beneath the surface or show how the elements are combined to form molecules and crystals. But in its non-invasive and portable form, the technique is revolutionizing the conservation field.

Conservation science is too small a market to attract the analytical-instrument industry. Conservation scientists often rely on tools that have been developed for other fields, such as the medical, oil and automobile industries or homeland security, adapting the technologies to the needs of the art world. Recently, they have even looked to the instruments of space programmes and planetary exploration.

G. CHIARI

Multi-spectral imaging reveals inscriptions on a museum mummy shroud.

For example, NASA has developed an innovative X-ray diffraction and fluorescence instrument that will be deployed on Mars in 2010 as part of the Mars Science Laboratory rover. This technology was redeveloped for applications on Earth by the California-based company inXitu, and the firm's chief technology officer Philippe Sarrazin has collaborated with the Getty Conservation Institute, a programme of the J. Paul Getty Trust in Los Angeles, to build an adapted portable instrument for art analysis. The instrument was redesigned to reflect X-rays off a surface — thus avoiding the need for sample collection — rather than using transmission mode, which would require the X-ray detector to be placed beneath the object surface.

On location, the instrument is adjusted in front of the artwork using precise laser alignment guides and a high-resolution camera. A fine X-ray beam is directed at the surface and the radiation scattered back is collected by an energy-dispersive detector that allows the diffraction and fluorescence signatures to be measured simultaneously in minutes. The combined diffraction and fluorescence data can differentiate not only between some common salts but also between hydration states, which may pinpoint the causes of deterioration and lead to the design of a suitable conservation strategy for an art object.

The new instrument will be extensively used to analyse the collection of antiquities, outdoor sculptures and manuscripts at the Getty Museum in Los Angeles. Later this year in the Roman ruins of Herculaneum, Italy, it will help examine excavated murals for pigments, salts and deterioration products. Other non-invasive portable instruments (such as Raman spectrometers and multi-spectral analysers) have been used to identify pigments and reveal inscriptions on the shroud of the Fayum mummy Herakleides, held at the Getty Museum. The additional shroud inscription 'son of Thermoutarion' was revealed only when light of a specific wavelength was used.

Of course, all portable instruments come with limitations. Miniaturization and portability may compromise accuracy and sensitivity. The user-friendly nature of portable instruments may convince non-experts to try to operate them, perhaps leading to incorrect analyses. Certain studies (such as stratigraphic analysis) necessitate cross-sections and therefore still require samples to be taken. Nevertheless, the use of such instruments will solve many key problems, such as analysing precious manuscripts without taking samples, or heavy statues, marble works and outdoor murals that cannot be moved.

To implement this technology skilfully, I foresee an increasing need for a new role in art conservation — the 'scientist–conservator'. Combining knowledge of artists' methods, the science of conservation and the technical instruments, these hybrid professionals will be able to ask the right questions to conserve our future. ■

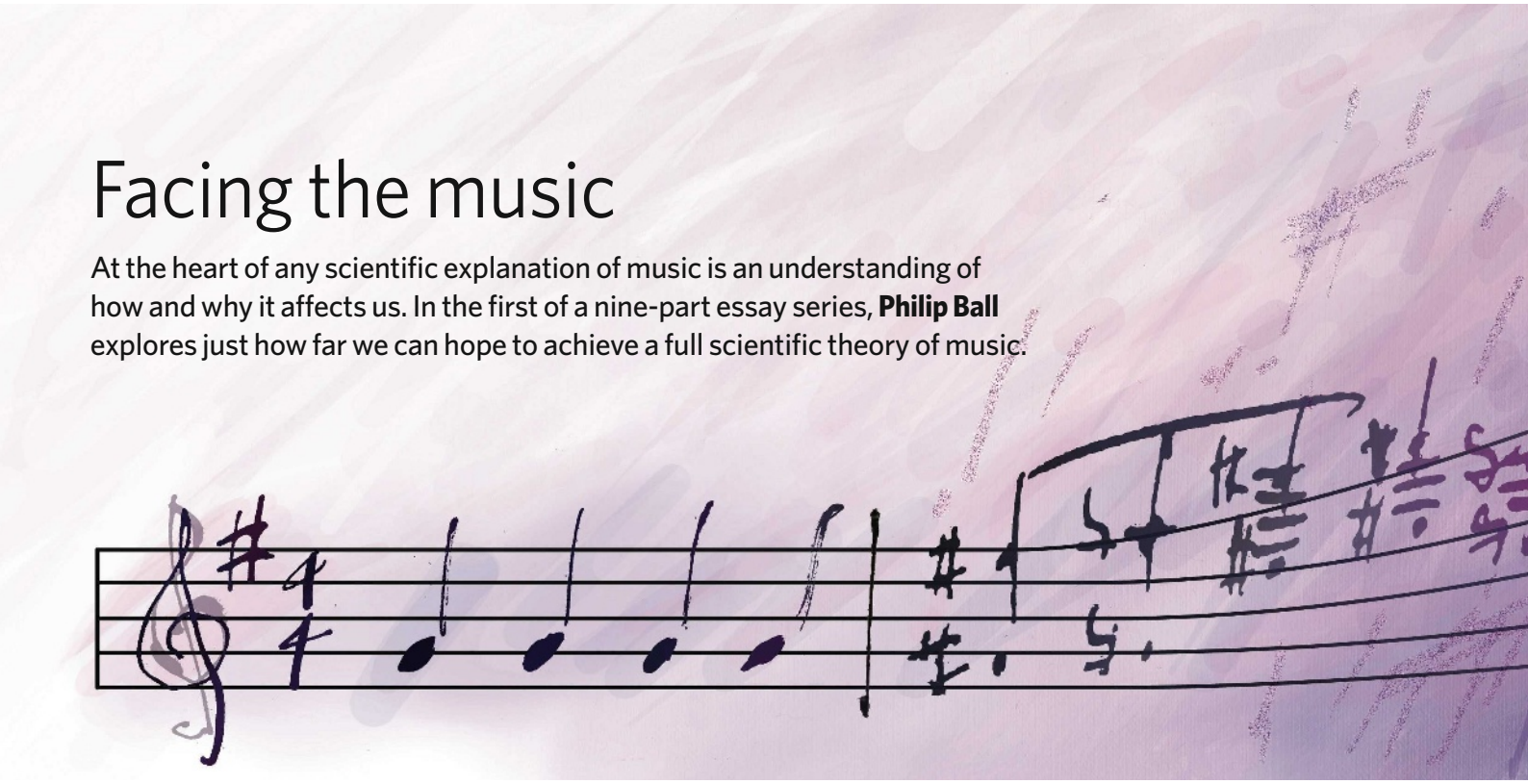
Giacomo Chiari is chief scientist at the Getty Conservation Institute, 1200 Getty Center Drive, Los Angeles, California 90049-1684, USA.

©2008 Nature Publishing Group

ESSAY

Facing the music

At the heart of any scientific explanation of music is an understanding of how and why it affects us. In the first of a nine-part essay series, **Philip Ball** explores just how far we can hope to achieve a full scientific theory of music.



How strange that music is deemed a phenomenon in need of scientific explanation. We don't, in general, construct objective theories of how great paintings 'work', or great literature, dance or sculpture. We are interested in what is happening at a perceptual level when we experience these arts, but there is always a space in which we leave them to speak for themselves, beyond the reach of cold facts. Yet with music, scientific studies seem to be on the trail of an absolute, all-encompassing explanation that connects neurology with creativity, auditory physiology with acoustic physics. There seems to be a conviction that the composer Arnold Schoenberg was right when he cautioned: "One day the children's children of our psychologists will have deciphered the language of music."

This 'scientification' of music is part of a very old tradition. In antiquity and the Middle Ages music was not an art in the modern sense; it was one of the four sciences of the syllabus called the liberal arts, alongside geometry, arithmetic and astronomy. Scholars studied music to learn about the natural harmony of the world, and performed music was often dismissed as frippery. The early sixth-century Roman philosopher Boethius ranked it as the least of his three classes of 'music', and agreed with Pythagoras that music should ideally be studied while "setting aside the judgement of the ears".

The practice of music does have something of the mathematical about it. Some of the experiments in compositional symmetry, such as the palindromes and mirror reflections of Wolfgang Amadeus Mozart and Joseph Hadyn, are little more than the parlour tricks of an age that delighted in such amusements. But many other musical forms and theories have deeper, more formal organization, from the interwoven fugues of Johann Sebastian Bach to the quasi-mathematical laws of composition developed by Paul Hindemith.

In the final throes of Schoenberg's twelve-note serialism in the 1960s, composers such as Pierre Boulez insisted on a mathematical rigidity that almost sucks their music dry of expression and makes onerous demands of the listener's ability to perceive ordered forms. And in some types of non-Western music, pattern and structure rather than emotion or tone-painting provide the foundations of composition. This is the case in polyrhythmic African drumming, for instance, and the shimmering soundscapes of Javanese gamelan.

Even musicians are uncertain of what kind of art it is they are engaged in, and what, if anything, can be said about it. 'Is there meaning in music?' asked US composer Aaron Copland. He felt there was, but admitted to being unable to articulate what that meaning is.

Almost the only thing we can say about music as a cultural phenomenon is that it seems to be universal. Music serves very diverse ends, sometimes with more apparent emphasis on the ritualistic than the hedonistic. Even when it is taken very seriously — in some Native American cultures a ceremony has to be started again if a single note is out of place — anthropologists have often struggled to understand how or to what extent cultures apply intellectual and aesthetic judgements. Sometimes music is a commodity for sale and exchange; elsewhere it is inseparable from dance.

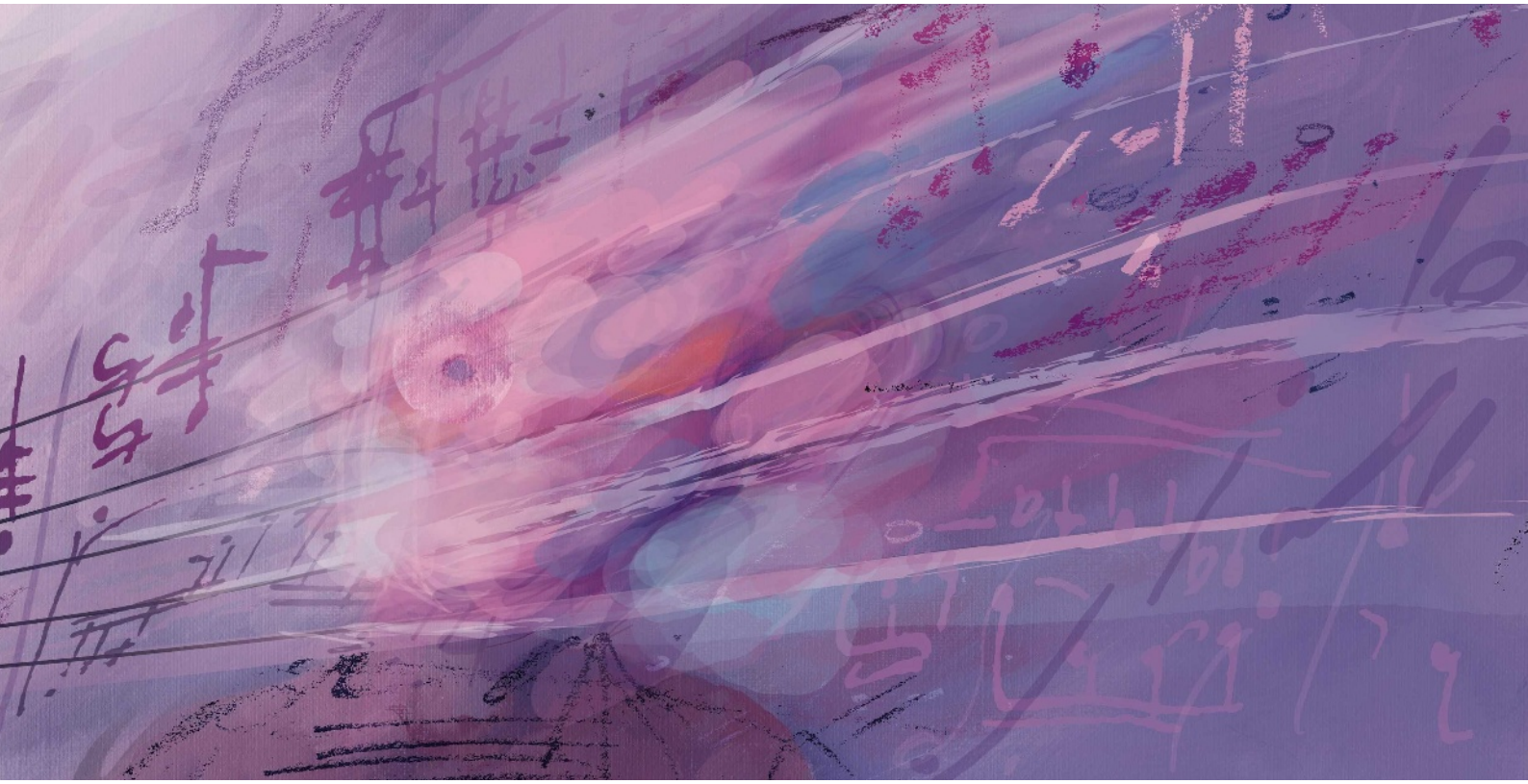
Given this range of what music is and what functions it serves, how can we make sense of it as an acoustic, cognitive, cultural and aesthetic phenomenon? That need not be deemed an entirely hopeless task, but it is not one that science will accomplish alone.

What is to 'understand'?

Trying to understand music is a little like trying to understand biology. The problem is so hard that you have to be reductionist, breaking it down into the building blocks and how they function. Then you find that the original problem has evaporated: in this atomistic view, 'life' or 'music' ceases to be visible at all.

Nonetheless, it makes sense to start with the nucleotides of music: single notes, idealized perhaps to pure tones with a single acoustic frequency. Here, like Pythagoras with his

©2008 Nature Publishing Group



D. PARKINS

vibrating strings, we ponder the apparent primacy of simple ratios in proportion and frequency, getting an octave from 2:1, a perfect fifth interval from 3:2 and a perfect fourth from 4:3.

Already there are complications. For one thing, musicians do not generally create pure sine tones. Instruments, like any resonating natural object, produce blends of a fundamental tone and harmonics with frequencies that are integer multiples of the fundamental. To us, these complex tones sound like a single note and not a chord.

"The practice of music does have something of the mathematical about it."

The brain seems to have evolved simple empirical rules of interpretation that furnish a good guess about the nature of the sounds we are hearing. As nearly all natural sounds are harmonically complex, the brain attributes related tones to a single sound source, combining them into a single auditory object. This is one of the 'Gestalt principles' the brain uses to decode sounds. It seems to be a basic aspect of sound processing — part of the natural auditory conditioning through which music is filtered.

These non-musical acoustic processing principles influence how music is composed. The rules of counterpoint developed in the early Baroque period, for example, use Gestalt grouping mechanisms to prevent separate melody lines from merging[1]. Bach and others also exploited the capacity to generate auditory illusions in the technique called virtual polyphony: splitting a single melodic line into two by means of alternating large pitch jumps.

Some have suggested that, through the series of overtones, conventional Western scales are inherent in a single, harmonically complex note. This idea was formalized by the eighteenth-century French composer Jean-Philippe Rameau, who showed that the major triad (tonic–third–fifth) can be derived from the harmonic series. Other music theorists have claimed that all the notes in the major scale (the white piano notes starting on C, say) originate this way, as higher harmonics of the fundamental tone.

Similar proposals have been made to explain why certain combinations of notes are consonant, or comfortable to the ear: simple arithmetical relationships between their frequencies are said to confer a favoured status in auditory processing. Such pythagorean reasoning fails to unravel convention and conditioning from physiology. For one thing, several 'consonant' intervals in the conventional Western 'equal-tempered' major scale (the major third in particular) have frequency ratios far from their 'ideal' values. Equal temperament, which divides the octave into twelve equally spaced semitones, is needed because transposing between keys does not preserve the ideal ratios. An E in the key of C does not have the same pitch as an E in the key of A, say. So a scale defined by pythagorean proportions in C will be increasingly out of tune the farther the key moves away from C. Our ears don't seem to object to the adjustments that equal temperament demands.

There is no apparent reason why a scale based on the harmonic series should sound more 'natural' than any other. This is borne out when we look at non-Western music. Some Javanese scales, for instance, pay no heed to intervals with small-integer ratios. Yet scales are not arbitrary. Most have between four and seven notes arranged asymmetrically within the octave, with some pitch steps bigger than others. They also tend to avoid steps much smaller than the Western semitone. The idea that some systems, such as those of Indian classical music, use finer visions, or 'microtonal' scales, has given way to a recognition that these microtones are basically embellishments of a simpler basic scale.

So how do different cultures decide on their musical scales? Cognitive studies on infants and primates offer some evidence that the brain recognizes the octave, and possibly the

©2008 Nature Publishing Group

fifth as 'special'. Indeed, these intervals feature in nearly all musical cultures that use scales. The other notes in a scale seem to be constrained in other ways, too. If there are too many notes per octave, it is hard to tell them apart, and instruments are difficult to tune. There is probably a good reason why most scales have unequal steps, as in the way the Western diatonic scales switch between whole notes and semitones. This asymmetry offers clues about a melody's tonal centre, letting a listener quickly figure out 'where the tune is' in relation to the tonic note.

It is also not obvious how much of the relative consonance and dissonance of different intervals, if any, is a 'natural' phenomenon. Certainly, notions of consonance in Western music have been fluid, defined largely by convention. But there does seem to be a genuine sensory dissonance in some combinations of tones, caused by the unpleasant sensation of beating between two tones that differ only slightly in frequency. Hermann von Helmholtz first did the maths in the nineteenth century and showed that sensory dissonance dips at the intervals corresponding to the Western scale, suggesting that physics does play a part in determining this scale. Yet there is considerable flexibility in the range of tunings that our ears will tolerate. It may even be that acclimatization to a convention can completely override these acoustic facts.

"How interval, melody and harmony act on the emotions is central to our understanding of music."

Why does music move us?

How interval, melody and harmony act on the emotions is central to our understanding of music. But we still have only hints of the mechanisms — or even of the basic phenomena. Take the tonic–minor third interval, for example: this ubiquitous musical element does not feature in the harmonic series at all. Some theorists have tried to argue that 'pulling down' the major third by a semitone to create the minor third creates an intrinsically sad effect. A passing acquaintance with Irish, Spanish or Hungarian folk music demolishes any idea that this association is anything more than a convention. Medieval church music largely ignored major keys, but not because it was trying to be 'sad'; the secular music of medieval troubadours used the major third, but not to be 'happy'. And Aristotle insisted that the Phrygian mode, a kind of scale that includes a minor third, "inspires enthusiasm".

Unlike literature, music cannot convey complex semantic ideas in any universal way. Musical phrases did have particular, conventional 'meanings' in the Western classical era of Mozart and Haydn, and some have tried to argue that instrumental music of other times has specific meanings that can be objectively decoded. Generally such efforts end up as dogmatic assertion, evoking French composer Albert Roussel's poignant image of the musician "alone in the world with his more or less unintelligible language".

Unlike visual art, there is no tangible referent in music that we can point to and contemplate at our leisure. Music works its invisible magic for a moment, then it is gone. It refers to nothing else in the world except by occasional, intentional mimicry. It is hard enough to understand why we can make the slightest sense of these eliding acoustic signals, let alone why they move us to tears and laughter, make us dance or rage. Acculturation alone does not seem sufficient to account for it: we can enjoy the music of other cultures intensely, yet miss all its important allusions and structures.

Someone coming to the psychology of music to understand how those passions are stirred will probably be disappointed. Current studies are often woefully inadequate, even simple-minded. When neurologists examine how and when people classify musical extracts as 'happy' or 'sad', the music lover may reasonably say it traduces the emotive qualities of music, as though we sit through Igor Stravinsky's *The Rite of Spring* either beaming or moping.

But we have to start somewhere. And even these reductive attitudes to musical affect can reveal useful things about how the brain processes musical input. Take the case of a patient with brain damage studied by Isabelle Peretz and her colleagues at the University of Montreal in Canada, for example. This patient had lost all ability to recognize either melody or rhythm but somehow retained the ability to make a 'correct' distinction between happy and sad. She claimed to still enjoy music but lacked what seemed to be the basic neural mechanisms for comprehending it[2]. This finding suggests that affective and cognitive processing of music might involve different neural pathways.

Crudely speaking, there are two camps in the analysis of emotion in music. One says that the emotional content is inherent in musical cues: the choice of mode (major/minor), tempo, timbre, melodic contour and so forth. The other says that it is all about how the music unfolds in time: how a combination of innate and learned responses set up expectations about what the music will do, and emotional tension and release flow from the way these are manipulated, violated and postponed[3]. Much of this is achieved during performance. An expressive performer uses subtle changes of timing, loudness, phrasing and improvized ornamentation to bring out emotional qualities that become barren under the hands of an inexpressive, mechanical player.

What's missing?

All this still falls pitifully short of telling us 'how music works'. It can provide endless narratives about musical events that tend to be somewhat arbitrary and untestable. Do several repeated notes create an expectation of further repetition, or of imminent change? When is a violation of expectation pleasing, and when does it jar, confuse or irritate? In complex polyphonic music, violations can take too many potential directions for us to develop meaningful expectations about them. Beyond a rough sense of shifting tonal centres, I feel no real expectations at all when confronted with the dense, multifaceted slabs of sonic matter that make up, say, Arthur Honegger's first symphony.

We need a better understanding of how the alchemy of music depends on texture. This may be less easily atomized than melody, rhythm and harmony, but it is a more 'musical' characteristic. Here, perhaps, music is working like visual art, just as Mark Rothko's paintings are not mere rectangles of maroon but complex, textured paint surfaces that massage and enliven the brain. Violations of expectation can't account for how that happens.

A lot of music is, after all, less about sequences of notes or beats than about sound sculptures, rich in timbre and composed of interlocking and overlapping layers that function as composite entities. This type of music, whether it is by Messiaen or Ministry, is far ahead of music psychologists still looking for tension peaks in a Mozart sonata.

Far from discouraging scientific studies as futile, these current lacunae should be a stimulus to it. We might start by accepting that it is fruitless to try to define 'music'. We will either leave something out, or include a lot of noise. We might accept too that we should not expect anything like a fully scientific theory of something so fluid. Perhaps there will always be some fundamental limitation in connecting how our brains work with what we do with them. ■

Philip Ball is a consultant editor for *Nature*. His new book *Universe of Stone: Chartres Cathedral and the Triumph of the Medieval Mind* is published this month by Bodley Head.

1. Huron, D. *Music Perception* **19,** 1–64 (2001).
2. Peretz, I., Gagnon, L. & Bouchard, B. *Cognition* **68,** 111–141 (1998).
3. Meyer, L. B. *Emotion and Meaning in Music* (Univ. Chicago Press, Chicago, 1956).

See Editorial, page 134, and http://tinyurl.com/559f2c for further reading.

©2008 Nature Publishing Group

NEWS & VIEWS

PLANETARY SCIENCE

Music of the stratospheres

Timothy E. Dowling

Fifteen-year oscillations in Saturn's equatorial stratosphere bear a striking resemblance to the shorter-term oscillations seen on Earth and Jupiter — akin to notes played on a cello, a violin and a viola.

Planetary lower atmospheres — the tropospheres — are clamorous. If, in addition to sound waves, one could hear buoyancy waves, which have periods measured in minutes, and vorticity waves, which have periods measured in days, the effect would be a deafening cacophony of enormous range. Time series of atmospheric data reveal a surprising response to this tropospheric din: the overlying stratospheres are answering back, and playing the deepest notes of all.

Studies by Orton *et al.*[1] and Fouchet *et al.*[2] (pages 196 and 200 of this issue) reveal that Saturn's equatorial stratosphere exhibits a 15-year oscillation in wind direction and temperature that is remarkably similar in structure to 2-year and 4-year oscillations in the stratospheres of Earth and Jupiter, respectively. Considering that a stratosphere is like a complicated sound box attached to the elaborate musical instrument that is the troposphere, this discovery is like listening to notes played on three random orchestral instruments and finding that they all are from the string section — and perhaps all from the same instrument.

Stratospheres are so named because they resist convective forces and so are comparatively stable. They form because the transmission of thermal infrared radiation to space by participating molecules (greenhouse gases) becomes efficient for pressures of 200–100 hectopascals and lower (1 hPa = 1 millibar; pressure, of course, decreases with height). Each of the seven stratospheres in the Solar System — those of Venus, Earth, Jupiter, Saturn, Titan, Uranus and Neptune — has its bottom, the boundary with the troposphere known as the tropopause, at approximately 100 hPa. (The other four atmospheres, those of Mars, Io, Triton and Pluto, have surface pressures significantly less than 100 hPa.) In terms of waves, the tropopause plays much the same part as the bridge on a violin or cello (Fig. 1), and its consistent placement near 100 hPa marks a pivot point for comparative planetology.

Orton *et al.*[1] analysed more than two decades' worth of observations from NASA's ground-based Infrared Telescope Facility (IRTF), focusing on wavelengths that are sensitive to methane and ethane emissions, which are diagnostic of stratospheric temperatures. A 14.8 ±1.2-year oscillation stands out clearly in their time series. Fouchet *et al.*[2] used the Cassini spacecraft's Composite Infrared Spectrometer (CIRS) to obtain thermal spectra from Saturn's limb — the edge of the planet where the line of sight passes through the most atmospheric mass, yielding high vertical resolution for the upper stratosphere. Their observations reveal the vertical structure of this oscillation, showing alternate bands of eastward and westward winds from an altitude of 0.2 hPa down to 5 hPa.

There are two terrestrial candidates as analogues to Saturn's 15-year signal, the quasi-biennial oscillation[3] (QBO) and the semiannual oscillation[4]. The QBO has a mean period of 28 months, during which time the direction of the jet stream at a given altitude in Earth's equatorial stratosphere switches from eastward to westward and back again, alternating with time, just as is now seen on Saturn.

The mechanism driving the QBO is more complicated than the ratcheting friction between bow and violin string, but it is not more complicated than the trials and tribulations of everyday life. Imagine a father trying to retrieve two (or more) rambunctious toddlers from a tall climbing frame. He clambers up the west side to coax down tot no. 1, blocking all the other kids' access on that side in the

B. MCDERMID/REUTERS

Figure 1 | An atmosphere as an upside-down violin. The strings are the troposphere, vibrating with many eastward- and westward-propagating waves. The bridge is the tropopause, transmitting the waves up into the soundbox, the stratosphere. The primary difference is that, unlike a typical string instrument, a stratosphere responds by dropping the input frequencies down by two or more orders of magnitude — from periods of minutes and days down to years.

©2008 Nature Publishing Group

process. Meanwhile, tot no. 2 scrambles up the east side. Dad plants tot no. 1 on the ground, and heads up the east side to retrieve tot no. 2, while tot no. 1 breezes back up the west side. When dad and tot no. 2 return to the bottom, one full cycle is complete. Of course, no self-respecting parent would put up with this nonsense, but what jet streams lack in intelligence, they make up in perseverance.

In fact, the troposphere is a playground for a large array of westward- and eastward-propagating waves that are constantly clambering up into the stratosphere. Forty years ago, this was recognized[5] as being key to the QBO mechanism. Each type of wave has a different personality. Buoyancy waves (internal gravity waves) travel in all directions; they are just like the waves on the ocean, but internal to the atmosphere. Near the equator, the vorticity waves (Rossby waves) travel slowly westward. In addition, there is a class of 'half waves' (equatorial Kelvin waves) that lean against each other across the equator; they travel rapidly eastward.

When any of these waves drifts upwards and encounters a stratospheric jet going in the same direction, it deposits its momentum just shy of the jet maximum, which has the effect of coaxing the jet downwards in a slow but continuous manner; this led to the first complete description[6] of the QBO mechanism. Meanwhile, waves that travel in the opposite direction are not blocked but can scramble all the way to the top of the climbing frame, thereby starting a new jet in their direction, which then slowly descends. The upshot is that the roughly 2-year period of the QBO on Earth is governed more by the strength of the wave flux, and the size and shape of the stratosphere, than by the rate of rotation or revolution of the planet.

Significantly, Jupiter also exhibits a QBO-like oscillation, with a period of 4.5 (Earth) years, appropriately called the QQO (quasi-quadrennial oscillation)[7]. As described by Fouchet *et al.*[2], similarities between Saturn's 15-year oscillations and the QBO/QQO include strong equatorial confinement of temperature extremes; asymmetry between the eastward and westward wind shears, with stronger eastward shears; and out-of-phase temperature changes between the equator and 15–20° latitude.

My guess is that the lengthening of the period from Earth to Jupiter to Saturn relates to their decreasing proximity to the Sun, which reduces the total energy budget available to their waves. But watch out: in addition to the QBO, Earth's stratosphere exhibits a strong signal with a 0.5-year period; this is the semiannual oscillation (SAO) mentioned earlier (Mars also seems to show an SAO[8]). The Saturn year is 29.5 Earth years, meaning that the Saturn wave is, at least descriptively, an SAO (for comparison, the corresponding time ratio for Jupiter's QQO is 4.5/11.9 = 0.38, which is not semiannual). Whereas Earth's SAO is driven by the different response to surface heating between the ice of Antarctica and the surrounding ocean, it is not obvious what the analogy is on Saturn (although its large ring shadow probably has a role).

The influence of the QBO and SAO on Earth's weather cannot be overstated. They modulate seasonal activity, the behaviour of the Hadley cell (the overturning circulation that predominates in the tropics), the strength of the polar vortex, the mixing of atmospheric trace species, and even the predictability of regional patterns such as the Indian monsoon in August and September[9]. Because the portfolio of eastward waves is distinct from that of westward waves, the eastward and westward phases of the QBO are different. The big news is that this asymmetrical, long-period response has now been observed in the stratospheres of three planets. The question for modellers is whether these stratospheres are like three different string instruments, or are more like three of the same instruments being played differently. ■

Timothy E. Dowling is at the Comparative Planetology Laboratory, University of Louisville, Louisville, Kentucky 40292, USA.
e-mail: dowling@louisville.edu

1. Orton, G. S. *et al. Nature* **453,** 196–199 (2008).
2. Fouchet, T. *et al. Nature* **453,** 200–202 (2008).
3. Baldwin, M. P. *et al. Rev. Geophys.* **39,** 179–229 (2001).
4. Andrews, D. G., Holton, J. R. & Leovy, C. B. *Middle Atmosphere Dynamics* (Academic, San Diego, 1987).
5. Lindzen, R. S. & Holton, J. R. *J. Atmos. Sci.* **25,** 1095–1107 (1968).
6. Plumb, R. A. *J. Atmos. Sci.* **34,** 1847–1858 (1977).
7. Leovy, C. B., Friedson, A. J. & Orton, G. S. *Nature* **354,** 380–382 (1991).
8. Kuroda, T., Medvedev, A. S., Hartogh, P. & Takahashi, M. *EPSC Abstr.* **2,** EPSC2007-A-00204 (2007).
9. Claud, C. & Terray, P. *J. Clim.* **20,** 773–787 (2007).

COMPUTATIONAL BIOCHEMISTRY

Old enzymes, new tricks

Giovanna Ghirlanda

Although enzymes are superb catalysts, their range of reactions is limited to those that support life. Their repertoire could be expanded by a method that allows artificial enzymes to be made from scratch.

Enzymes are astoundingly good catalysts: they allow reactions to occur billions of times faster than would be possible without them, at temperatures much lower than those required by typical synthetic catalysts. But enzymes have evolved to accelerate only biological reactions, under the narrow set of conditions that are compatible with life. Two papers from the same group, one in this issue (Röthlisberger *et al.*[1], page 190) and another in *Science* (Jiang *et al.*[2]), show how these limitations can be overcome. They describe a method for designing enzymes that catalyse unnatural reactions, and demonstrate its use for two different chemical transformations.

Enzymes work by lowering the activation energy of reactions, specifically by confining substrates in binding sites that stabilize the highest-energy arrangement of atoms in the reaction pathway (known as the transition state). They also shield the reactants, thus preventing possible side reactions. The idea behind the latest work[1,2] is simple — model the transition state for a reaction, stabilize it by surrounding it with carefully placed chemical groups, graft the resulting active site into an existing protein and then alter the amino-acid sequence of the protein to accommodate the changes. In practice, this is a complicated procedure. For starters, building an accurate model of a transition state requires a detailed understanding of the reaction's mechanism, which isn't always available. Furthermore, transition states are modelled using quantum-mechanical calculations, but currently available methods can handle only a limited number of atoms, and are often inadequate for modelling enzyme reactions.

Designing a protein that folds into a given structure is equally challenging. For a protein made of 100 amino acids, there are about 10^{130} possible sequences, each of which can adopt many different conformations. The thermodynamic stability of every sequence and conformation must therefore be calculated to find the lowest-energy structure (that is, the one most likely to form). Some simplifications can be made using advanced computational methods to quickly eliminate unfavourable combinations. This has resulted in several notable accomplishments, such as the complete redesign of a protein consisting of 28 amino acids[3], the design of an amino-acid sequence that forms a structure not found in nature[4], and the engineering of naturally occurring proteins into biosensors for trinitrotoluene (TNT) and other small molecules[5].

With these precedents, you might think that designing catalytic proteins should be straightforward, but success has been limited. Catalytically inactive proteins have been converted into modestly catalytic ones for two different reactions, but the observed enhancements of rate[6,7] were only about a millionth of those produced by naturally occurring enzymes. It is also sometimes difficult to prove that designer enzymes are truly catalytic on the basis of biochemical observations, and some exciting claims have been found to be flawed.

©2008 Nature Publishing Group

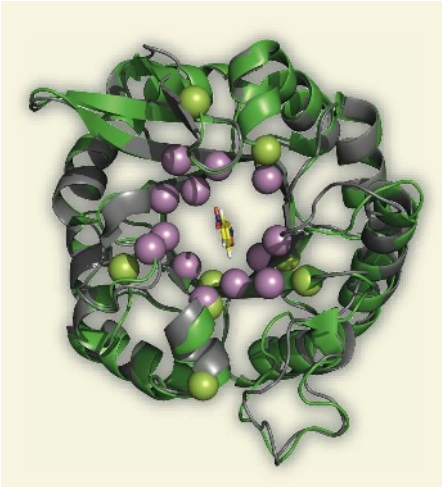


Figure 1 | Enzymes by design. Röthlisberger *et al.*[1] have computationally designed and prepared the first enzyme capable of catalysing a non-biological reaction. Here, the computational model (grey) is overlaid with the crystal structure of the actual protein (green); the two overlap almost perfectly. The substrate is shown at the centre of the structure. The design process involved modifying the amino-acid sequence of a naturally occurring protein. Residues selected computationally to form the active site are shown as purple spheres. Additional mutations that were introduced *in vitro* to optimize the enzyme's performance are shown as green spheres.

But some reports of catalysis by designed enzymes have fared rather better — especially those that are based on sound crystallographic evidence[6–8]. An essential step in demonstrating the success of a designer enzyme, therefore, is the determination of a high-resolution crystal structure for the protein, to verify that the designed catalytic features are present. The results of Röthlisberger *et al.*[1] and Jiang *et al.*[2] are remarkable in the spectacular agreement between their computationally predicted enzyme models and the experimentally determined structures (Fig. 1).

Röthlisberger *et al.*[1] made an enzyme that catalyses the Kemp elimination reaction (see Fig. 1a on page 190 for a reaction scheme). The Kemp elimination is initiated by the removal of a hydrogen ion from a carbon–hydrogen bond in the substrate; the minimum requirement for catalysis of the reaction is the presence of a base to perform this step. The authors therefore identified two amino acids — aspartic acid and histidine — that have side chains that can act as bases under physiological conditions, and used these as the starting points of their putative active sites. They decorated models of the proposed active sites with other chemical groups found in proteins, choosing those that could interact favourably with groups in the substrate. They then used state-of-the-art quantum-mechanical methods to precisely place all the groups in the models to maximize stabilization of the transition state of the substrate. The authors thus obtained a large ensemble of designs for catalytic sites in enzymes.

Next, Röthlisberger *et al.* selected about 100 proteins that could be used as scaffolds for their proposed active sites. The criteria for selection were the availability of high-resolution crystal structures and the presence of pre-organized cavities, with a preference for proteins that behave well in experiments (that is, those that have good solubility, are expressed easily in cells, and so on). The authors then used computational methods to search each of the proteins for specific regions that could accommodate the sites, narrowing down the vast number of possibilities to about 100,000 promising leads. These were whittled down further using an automated modelling technique to find the optimal amino-acid sequence in defined shells around the active site, selecting sequences that maintained protein stability and integrity.

This computational screening method picked out 59 candidate enzymes, which the authors expressed in cells and evaluated for their ability to catalyse the target reaction. Only eight of the proteins had measurable catalytic activity. The team then used *in vitro* evolution to further optimize one of their successful leads (designated KE07), mutating the amino-acid sequence in both random and directed locations. After several rounds of mutation and screening, Röthlisberger *et al.* obtained improved enzymes that were up to 200 times more active than KE07. The best two of these mutants accelerate the rate of the Kemp elimination reaction to about a million times that of the uncatalysed version.

The strategy used by Röthlisberger *et al.*[1] promises to be general, as the same group[2] has successfully applied the procedure to another chemical transformation known as the retro-aldol reaction, which is very different from the Kemp elimination. The complexity of the design procedure is underlined by the number of interdisciplinary groups involved in the work, and by the huge amount of computational power required to solve the problem — donated from hundreds of thousands of idling computers around the world as part of a project known as Rosetta@home[9].

Those in the know might say that the performance of the designed enzymes is far from impressive — the reaction-rate enhancements for typical, naturally occurring enzymes are anywhere between 10,000 and 1 billion times higher than those of the artificial enzymes described in these papers[1,2]. Furthermore, the chosen reactions are relatively easy targets. The Kemp elimination is accelerated by several catalysts, including various synthetic compounds, catalytic antibodies and even serum albumin. Similarly, the retro-aldol reaction is catalysed by antibodies[10] and by various peptides[11,12]. Indeed, the rate enhancements reported by Röthlisberger *et al.*[1] are equivalent to those of only the most sophisticated catalytic antibodies[13,14]; the enhancements obtained by Jiang *et al.*[2] for the retro-aldol reaction are even more modest.

Another limitation of the design process is that, although naturally occurring enzymes

50 YEARS AGO

A Hundred Years of Evolution. By Dr. G. S. Carter — It is fundamental to the neo-Darwinian theory that Weismann's concept of the inviolability of germ plasm by soma is correct, and that mutational changes in the gene complex arise solely at random; Dr. Carter (p. 87) accepts Weismann's doctrine as "undeniable when once pointed out". It is arguable, however, that the "separateness of the gonad from the rest of the soma" is a philosophical concept of the same order as that of the soul and the body. As such it may have been valid in the state of biological knowledge in Weismann's time, but it has to-day become undermined to the point of collapse ...That mutation is random is purely theoretical, depending in the first place on the validity of the divorce between germ plasm and soma, and in the second upon the absence of evidence to the contrary.

From *Nature* 10 May 1958.

100 YEARS AGO

To be told that life exists on Mars tells us but little of its nature ... Perhaps on Mars there is only one living being, a gigantic vegetable the branches or pseudopodia of which embrace the planet like the arms of an octopus, suck water from the melting polar snows, carry it to other parts of the planet, and are visible to us as the Martian canals. Lowell adduces the straightness of the canals as a proof that they are artificial products of intelligent beings. But they are certainly no straighter than the somewhat similarly interlaced pseudopodia seen in certain Heliozoa, Foraminifera, and Radiolaria ... My position is that one may admit that Prof. Lowell's brilliant researches prove the existence of life on Mars, and still ask from him further evidence before we are convinced that that life is intelligent.

From *Nature* 7 May 1908.

50 & 100 YEARS AGO

©2008 Nature Publishing Group

have evolved to optimize steps other than just catalysis (such as the binding of substrates and the release of products), the model used by the authors[1,2] to design their enzymes doesn't attempt to address these factors. This is understandable, because many of the finer features that provide enzymes with their unique properties are not yet understood. For example, the mutations introduced by the authors into their enzymes by directed evolution did not modify the active site itself, but occurred at neighbouring positions (Fig. 1). The effect of some of these mutations can be easily understood with hindsight, but others are much less obvious. It was therefore wise of the authors to let nature lend a helping hand in their designs.

Nevertheless, these results[1,2] are a milestone in biochemistry. For the first time, artificial enzymes have been designed for non-biological reactions, providing rate accelerations that are about 1,000 times faster than previous examples of computationally designed enzymes. Biochemists have long wanted to build artificial enzymes to identify and validate the minimal requirements for enzyme-like catalysis. These reports provide an accurate framework for this enterprise to which further features can be added. As Röthlisberger *et al.*[1] note, the ability to design enzymes will truly test our understanding of enzyme catalysis. ■

Giovanna Ghirlanda is in the Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287-1604, USA.
e-mail: giovanna.ghirlanda@asu.edu

1. Röthlisberger, D. *et al. Nature* **453,** 190–195 (2008).
2. Jiang, L. *et al. Science* **319,** 1387–1391 (2008).
3. Dahiyat, B. I. & Mayo, S. L. *Science* **278,** 82–87 (1997).
4. Kuhlman, B. *et al. Science* **302,** 1364–1368 (2003).
5. Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. *Nature* **423,** 185–190 (2003).
6. Bolon, D. N. & Mayo, S. L. *Proc. Natl Acad. Sci. USA* **98,** 14274–14279 (2001).
7. Kaplan, J. & DeGrado, W. F. *Proc. Natl Acad. Sci. USA* **101,** 11566–11570 (2004).
8. Geremia, S. *et al. J. Am. Chem. Soc.* **127,** 17266–17276 (2005).
9. http://boinc.bakerlab.org/rosetta/
10. Sinha, S. C., Barbas, C. F. & Lerner, R. A. *Proc. Natl Acad. Sci. USA* **95,** 14603–14608 (1998).
11. Tanaka, F. & Barbas, C. F. *Chem. Commun.* 769–770 (2001).
12. Tanaka, F., Fuller, R. & Barbas, C. F. *Biochemistry* **44,** 7583–7592 (2005).
13. Hu, Y., Houk, K. N., Kikuchi, K., Hotta, K. & Hilvert, D. *J. Am. Chem. Soc.* **126,** 8197–8205 (2004).
14. Thorn, S. N., Daniels, R. G., Auditor, M.-T. M. & Hilvert, D. *Nature* **373,** 228–230 (1995).

DEVICE PHYSICS

Chance match

Robert M. Westervelt

A clever device uses the quantum statistics of electron tunnelling to match image patterns. The circuit is low-power, works at room temperature — and could point to a way forward for silicon electronics.

Over the past three decades, as the components that make up integrated circuits have been made smaller and smaller, the power of computer chips has grown exponentially, even as their cost has fallen drastically. But sooner rather than later — by around 2020, according to one estimate[1] — the scaling-down process will become difficult to maintain[2,3]. The energy required to represent a bit of information will become larger than the heat that can be carried away from a tiny circuit element; what's more, as devices approach the size of atoms, quantum-physical phenomena will become important, changing even the ground rules of how bits are processed. Writing in *Applied Physics Letters*[4], Nishiguchi *et al.* detail what might be one way to circumvent, and even exploit, these issues. They describe a circuit that allows them to perform the computing operation of pattern matching by harnessing the stochastic, quantum-mechanical tunnelling of single electrons into a transistor[5].

Pattern recognition is a natural enough task for people, but is often difficult for computers. We would like our computer processors to be like us and recognize an object (a cat or a dog, say) in a photographic image, understand the meaning of spoken words, or translate efficiently from one language to another. But pattern recognition also has more abstract, fundamental uses: extracting a simple conclusion from a great body of input data, for instance.

Nishiguchi *et al.*[4] build on previous work[6–8] to construct a simple pattern-matching circuit using a basic building-block of two transistors (more precisely, metal–oxide–semiconductor field-effect transistors, or MOSFETs) patterned on a silicon-on-insulator wafer. They trap and store single electrons on the first of these nanoscale transistors, the 'T-FET'. They are able to reduce the rate at which electrons tunnel quantum-mechanically into a storage node on the T-FET to very low levels of around one per second. The authors show that the trapped electrons obey Poisson statistics, and represent a statistically random source of events that can be used for stochastic signal processing[9].

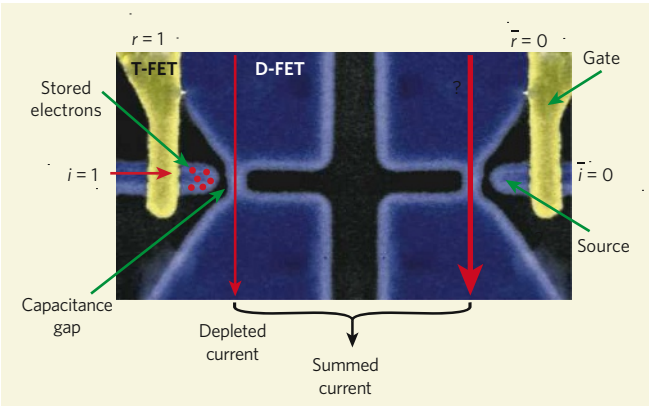


Figure 1 | Dual processor. Nishiguchi and colleagues' pattern-recognition processor[4] uses two basic components that each consist of two capacitatively coupled transistors: a transfer transistor (T-FET) and a detector transistor (D-FET). The probability that an electron will tunnel from the source of the T-FET, under the gate and into the storage node is determined by the source voltage, which is set by the value of a bit i in the input image, and by the gate voltage, which is set by a bit r in the reference image. The more electrons accumulate in the T-FET storage node, the lower the current that flows through the capacitatively coupled D-FET. In the instance depicted, both the input and reference bits are turned on, $i = r = 1$, and electrons accumulate in the storage node, reducing the detector current. The second unit (right) is fed with the inverse inputs of the first, $\bar{i}$ and $\bar{r}$. If the original inputs were matched at 0, the inputs here would be 1, and this half of the processor would record the depleted current characteristic of matched bits. (Figure adapted from ref. 1.)

The job of the second transistor that makes up the authors' processor, the 'D-FET', is to detect the number of electrons stored in the T-FET. It does this through a capacitative coupling: as the number of electrons stored in the T-FET increases, the current passing through the D-FET is progressively reduced. The coupling is sensitive enough that the tunnelling of a single electron into the T-FET is registered as a discrete drop in current in the D-FET.

To perform pattern matching, the individual bits of an input image must be compared with those of a reference image. Nishiguchi and colleagues set an input bit, i, to 0 or 1 by stepping the input 'source' voltage of the T-FET. Similarly, they set a reference bit, r, by changing the T-FET's 'gate' voltage, which controls the passage of current from the source into the storage node (Fig. 1). When $i = 0$, the tunnelling rate into the T-FET is negligible. When $i = 1$, tunnelling occurs, and the number of electrons stored in the T-FET slowly builds up. The precise rate of this tunnelling is controlled by the gate voltage, and thus the reference bit: it is large when $r = 1$, and small when $r = 0$.

Essentially, this set-up creates a detector that flags up when the input and reference bits are both on, $i = r = 1$: in this case, electrons build up in the T-FET particularly

©2008 Nature Publishing Group

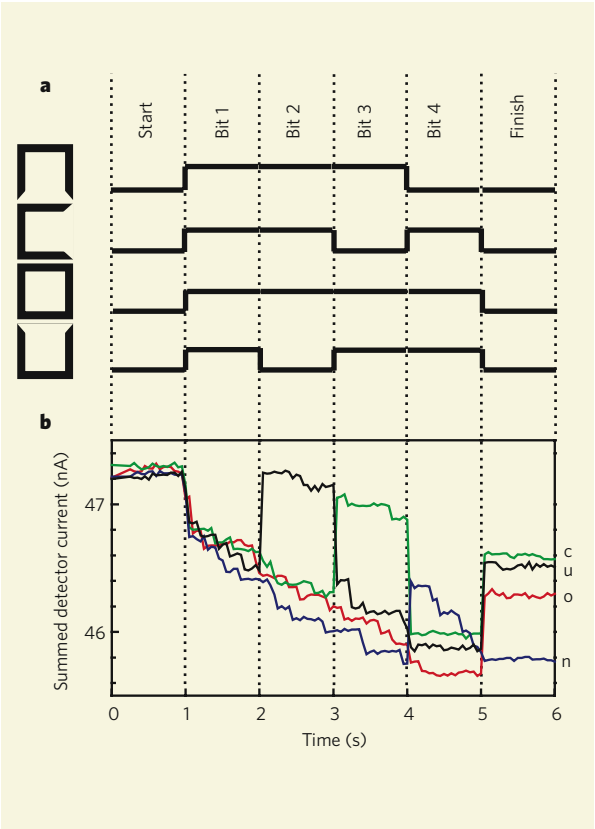


Figure 2 | Comparing letters. **a**, Images of the letters n, c, o and u can be translated into sequences of four bits (working clockwise from the left of each letter). **b**, Nishiguchi *et al.*[4] test their pattern-matching processor by sequentially feeding in the source image bits encoding n, and comparing them in turn with reference bit strings for n, c, o and u. The first bit in each reference sequence is 1, so the technique cannot distinguish between them — the detected current falls similarly for each. But as subsequent image and reference bits differ, the currents registered for those reference images increase — until after all four bits, the 'winning' reference image, n, is clearly marked by its low detector-current signal. The individual steps in the current record correspond to the tunnelling of individual electrons into the processor; the sharp steps in output current at the transitions between bits occur because the D-FET picks up changes in T-FET source and gate voltage. (Figure adapted from ref. 1.)

fast, and the current passing through the D-FET decreases rapidly. To recognize a pattern of bits such as that in an image — a series of 0s and 1s — a second unit is required that responds similarly when $i = r = 0$. This can be done simply by setting the T-FET source and gate voltages on the second unit to represent the logical inverses of whatever the input and reference bits are on the first unit. The output currents from both D-FETs are then added together. The result is the equivalent of an exclusive NOR logic gate: it flags up through a sharp drop in current whenever the input and reference bits match up (whatever the values of those bits are), but does not respond when they differ.

The authors demonstrate the principle of pattern matching using the circuit by feeding it sequentially with four bits representing an image of the letter n, and comparing it with four four-bit reference sequences encoding the letters n, c, o and u (Fig. 2a). The total error was found in each case by summing the number of electrons collected in the two T-FETs, as deduced from the drop in summed D-FET current. When the reference coding for n was used, the current drop was by far the greatest (Fig. 2b).

Nishiguchi and colleagues' advance is opportune for two reasons. First, it comes at a time when computing is moving away from single processors towards many processors operating in parallel. 'Cloud' computing, which uses very many parallel processors, is what allows search engines such as Google to provide rapid answers to our web enquiries, and now even many laptop computers contain chips that have two or more processors, or 'cores'. In this multi-processor environment, it is increasingly likely that we would wish to add special circuits dedicated to a single purpose, such as the fundamental task of pattern recognition.

Second, we are now recognizing that entirely new chip architectures and new approaches to computation are needed to continue the rapid growth in computing power to which we have become accustomed[1–3]. To that end, the semiconductor industry has developed a set of goals[3] for research aimed at producing truly nanoscale switches with low power requirements that work at room temperature. Nishiguchi and colleagues' circuit is a meaningful step in this direction: it uses silicon-on-insulator circuits that are compatible with conventional silicon technology, it works at room temperature, and it can respond to the stimulus of just a single electron.

In its present form, this work[4] represents just a proof of principle. But extended to a system that can handle many bits, such single-electron circuits that exploit stochastic quantum-mechanical effects to produce low-power devices could be an important part of a brave new electronic future. ■

Robert M. Westervelt is in the School of Engineering and Applied Sciences, and the Department of Physics, Harvard University, 29 Oxford Street, Cambridge, Massachusetts 02138, USA.
e-mail: westervelt@seas.harvard.edu

1. Bourianoff, G. *IEEE Computer* **36** (8), 44–53 (2003).
2. Cavin, R. K. *et al. J. Nanopart. Res.* **7,** 573–586 (2005).
3. Bourianoff, G. I., Gargini, P. A. & Nikonov, D. E. *Solid-State Electron.* **51,** 1426–1431 (2007).
4. Nishiguchi, K., Ono, Y., Fujiwara, A., Inokawa, H. & Takahashi, Y. *Appl. Phys. Lett.* **92,** 062105 (2008).
5. Likharev, K. K. *Proc. IEEE* **87,** 606–632 (1999).
6. Kobayashi, D., Shibata, T., Fujimori, Y., Nakamura, T. & Takasu, H. *IEEE Trans. Electron Dev.* **52,** 2188–2197 (2005).
7. Nishiguchi, K., Inokawa, H., Ono, Y., Fujiwara, A. & Takahashi, Y. *Appl. Phys. Lett.* **85,** 1277–1279 (2004).
8. Nishiguchi, K., Fujiwara, A., Ono, Y., Inokawa, H. & Takahashi, Y. *Appl. Phys. Lett.* **88,** 183101 (2006).
9. Uchida, K., Tanamoto, T., Ohba, R., Yasuda, S. & Fujita, S. *Tech. Digest Int. Electron Devices Meet.* 177–180 (2002).

QUANTUM INFORMATION

Stopping the rot

Philip C. E. Stamp

Uncontrollable outside influences undermine the whole enterprise of quantum computing. Nailing down the sources of this 'decoherence' in a solid-state system is a step towards solving the problem.

In the quest for a quantum computer, no obstacle is more formidable than decoherence — the 'collapse' of an information-encoding quantum wavefunction when it couples to its surroundings. We pressingly need to understand what causes it, how it works and how to get rid of it. Bertaina *et al.* (page 203 of this issue)[1] have passed a milestone on that road. They report the first observation of Rabi oscillations, a signature of coherent spin dynamics, in a magnetic molecule of a kind envisaged as the basic physical carrier of a 'qubit' of quantum information in a quantum computer. Perhaps more importantly, they have also succeeded in pinpointing the sources of decoherence in their system, and so taken the first step towards eliminating them.

Magnetic molecules come in all shapes and sizes, and have spins with values ranging from 1/2 , the smallest that quantum theory allows, to more than 30. Their great advantage for making qubits is that all molecules of a species are the same, and have a structure governed purely by quantum mechanics. The authors focus on the vanadium V_{15}^{IV} molecule, which, at just over a nanometre in diameter, is small-to-middling in size. It has an interesting spin structure, in which 15 vanadium ions, each with a net electronic spin of 1/2, couple strongly into three groups of five.

©2008 Nature Publishing Group

Because of the way spins add as vector quantities (direction, as well as magnitude, counts), the whole molecule can have an overall spin of 1/2 or 3/2, depending on how the individual electron spins line up. These low-energy spin states are very widely separated from the many higher-energy states. The spin-1/2 state in particular, which has two energy levels corresponding to molecular spin 'up' and molecular spin 'down', is a natural candidate for a two-state qubit.

But it's here that interactions with the environment — decoherence — become a problem. Decoherence was long thought to be a relatively simple process. A popular view was to model the environment as a 'bath' of oscillators that are not localized, but extend throughout space[2]. Decoherence was the result of transitions in the bath caused by its interactions with the central quantum system of interest. The results of experiments on simple quantum-optical systems[3] and on superconductors[4] agreed with this picture.

But there were also good reasons to suppose that the oscillator-bath picture should not work in describing low-temperature decoherence in most solid-state systems, in which decoherence is mostly caused by the influence of entities in the local, rather than the extended, environment[5]. These might be nuclear spins, found almost everywhere, or else one of the many defects (some charged), dislocations and spin impurities found in anything but a perfect crystal. All these objects hop or flip quantum mechanically between a few different states, so that they act as a reservoir of quantum states known as a spin bath. A spin bath often causes little dissipation of energy, but can cause quite devastating decoherence by its interactions with the central quantum system.

Bertaina *et al.*[1] were able to spot Rabi oscillations between the low-energy qubit states of their vanadium molecule — the first time the phenomenon had been seen in a molecular magnet, and clinching proof that a degree of coherence is present in the system. But the authors were also able to work out what was causing decoherence, as manifested in the decay of the Rabi oscillation. They found that the prime source was the 15 vanadium nuclear spins in each molecule, with a rather smaller contribution from hydrogen nuclei (protons) also present in the structure. The experimental decoherence rate differed by only a few per cent from that expected theoretically for spin-bath decoherence in this system[5,6].

This result indicates that the decoherence mechanism is as follows. Each time the spin state of the qubit flips from up to down, it also flips the field on the vanadium nuclear spins. But because both other internal and externally applied fields are present, this nuclear flip is not through fully 180°. The nuclear spins attempt to realign with the field, but because the field is constantly jumping, they end up precessing in a complicated way that depends on the motion of the qubit. Quantum mechanically, this means that the dynamics of the nuclear spin bath are entangled with the qubit dynamics — decoherence has occurred, even though no energy has dissipated from the qubit into the nuclear spin bath[2] (Fig. 1). This is a remarkable finding, because the magnetic moments resulting from the nuclear spin are thousands of times smaller than those associated with the electronic spin of the qubit; and yet, like David overcoming Goliath, they prove the stronger party.

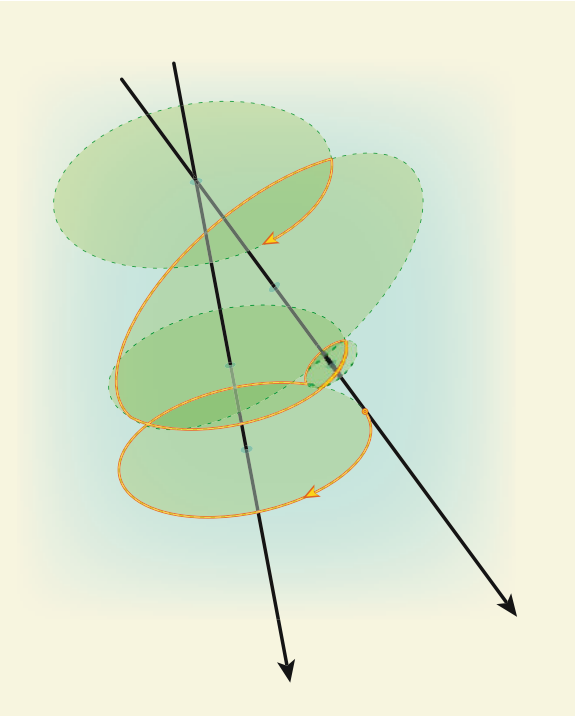


Figure 1 | Descent into decoherence. In Bertaina and colleagues' experiment[1], a spin qubit flipping between its up and down quantum states (a Rabi oscillation) also flips the field acting on nearby nuclear spins between two orientations. The nuclear spins try to precess in this qubit field, but each time it suddenly changes they must begin anew. Thus, the path they follow is conditional on the specific trajectory of the qubit — the two are quantum-mechanically entangled, which leads to the decoherence of the qubit.

As far as other possible sources of decoherence are concerned, Bertaina *et al.* calculated the contribution of lattice vibrations (phonons) and found it to be more than a hundred times weaker than the nuclear-spin contribution. This is because the phonon frequency is much higher than that of the qubit's Rabi oscillation, so that the phonons smoothly follow the qubit dynamics, rather than destroying it. Dipolar interactions between separate vanadium molecules are potentially more dangerous than any other decoherence source, because they are effective over long ranges[7,8]; but the authors were able to suppress these effects simply by spacing the vanadium molecules far apart in a solvent.

What are the implications of these results for future work? Certainly, the prospects for using magnetic molecules as qubits are good. If one can get rid of nuclear spins — perhaps using systems with only zero-spin nuclei, prepared by isotopic purification — then the intrinsic decoherence time is about 100 microseconds for a two-level system with an energy separation of around 10 gigahertz. That should be enough to permit a quantum computer to work, given sufficiently weak dipolar interactions[7].

The nature of spin-bath decoherence has now been addressed experimentally in both molecular magnets[1] and rare-earth metals[9,10], and Rabi oscillations have been seen in both[1,9]. Such systems would thus seem to have a clear edge over a rival system posited as a viable basis for a qubit — electron transitions in the semiconductor structures known as quantum dots. In quantum dots, roughly a million nuclear spins can couple to each qubit (although ingenious methods have been proposed to deal with these[11]). Similarly, the magnetic-molecule qubits are superior to superconducting qubits, which are so large that they inevitably harbour many defects.

But before we get carried away by these latest achievements, two urgent 'architectural' problems must be solved. The first is that, in a real quantum computer, one might not have the option of keeping the qubits very far apart — so a way must be found to arrange the qubits and their interactions to suppress errors arising from dipolar interactions. The second is that the small size of the qubits means that reading out the quantum state of a large number of them, as well as controlling individual qubits externally, has so far defeated our experimental guile. But there is no fundamental reason why these problems cannot be solved. With advances such as that of Bertaina and colleagues[1], there would seem to be good grounds for optimism for the future of spin-based quantum computation. ■

Philip C. E. Stamp is at the Pacific Institute for Theoretical Physics, and the Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada.
e-mail: stamp@phas.ubc.ca

1. Bertaina, S. *et al. Nature* **453,** 203–206 (2008).
2. Caldeira, A. O. & Leggett, A. J. *Ann. Phys. (NY)* **149,** 374–456 (1983).
3. Myatt, C. J. *et al. Nature* **403,** 269–273 (2000).
4. Clarke, J. *et al. Science* **239,** 992–997 (1988).
5. Prokof'ev, N. V. & Stamp, P. C. E. *Rep. Prog. Phys.* **63,** 669–726 (2000).
6. Stamp, P. C. E. & Tupitsyn, I. S. *Phys. Rev. B* **69,** 014401 (2004).
7. Aharonov, D., Kitaev, A. & Preskill, J. *Phys. Rev. Lett.* **96,** 050504 (2006).
8. Morello, A., Stamp, P. C. E. & Tupitsyn, I. S. *Phys. Rev. Lett.* **97,** 207206 (2006).
9. Bertaina, S. *et al. Nature Nanotech.* **2,** 39–42 (2007).
10. Rønnow, H. M. *et al. Science* **308,** 389–392 (2005).
11. Yao, W., Lao, R.-B. & Sham, L. J. *Phys. Rev. Lett.* **98,** 077602 (2007).

©2008 Nature Publishing Group

OBESITY

What's your fat-cell allowance?

In the American sitcom *Friends*, Monica bakes mouth-watering cookies for her housemates, but — having been overweight as a teenager (pictured) — she is reluctant to eat them herself. Kirsty Spalding and colleagues now provide scientific evidence for why Monica has every reason to be cautious (K. L. Spalding *et al. Nature* doi:10.1038/nature06902; 2008).

Two factors contribute to an increase in fat mass: the number of fat cells and how much fat each of them stores (their volume). The authors studied the dynamics of fat-cell number in some 700 adults, both lean and obese, and combined their data with previous observations in children and adolescents.

A clear pattern emerged: irrespective of weight, the number of fat cells seems to rise steadily from birth to the early twenties, but remains constant thereafter. Moreover, in patients observed before and up to two years after surgical treatments that facilitate weight loss by reducing stomach size, no decrease in fat-cell numbers was detected — although their volume did drop.

So, are fat cells that are generated in early life doomed to remain with us till death us do part? In animal studies, this question can be addressed by labelling DNA nucleotides with radioactive isotopes such as ^{14}C. Differentiated fat cells do not divide, and so radioisotopes, incorporated in their DNA in the last round of division before differentiation, remain there throughout the cells' life. The time of radiolabel incorporation, which is worked out from its half-life, is therefore the 'birth date' of these cells. But the potential toxicity of radioisotopes means that such studies cannot be performed in humans.

Spalding *et al.* cleverly thought of the next-best option. Atmospheric levels of ^{14}C have remained relatively constant for centuries, with the only major increase occurring between 1955 and 1963, when nuclear bombs were being tested above ground. A chain of reactions ensures that, at any given time, the radioisotope content of human DNA matches that of the atmosphere. The authors could thus follow fat-cell dynamics in individuals born around 1955–63.

As Spalding and colleagues' results show, fat cells have a high turnover: new cells are continually being born to replace their dead predecessors. The average age of a fat cell seems to be about 10 years in both lean and obese individuals, and the number of fat cells as a proportion of all cells remains constant in each weight group. But the total number of new fat cells was higher in obese subjects, suggesting that they are replenishing an existing larger pool.

So do the lean among us need to worry about our diet if we have fewer fat cells? Yes, we do: our fewer fat cells can still store large amounts of fat. Also, can obese people do anything about their weight? After all, they've already accumulated a large pool of fat cells in childhood and adolescence? Again, the answer is yes. As Monica seems to have guessed, they can still reduce the volume, if not the number, of their fat cells. A further corollary of the paper is that researchers seeking drugs to cure obesity should consider targeting the mechanisms underlying fat-cell turnover.

Sadaf Shadan

NBCUPHOTOBANK/REX FEATURES

MOLECULAR BIOLOGY

An HIV secret uncovered

Eddy Arnold and Stefan G. Sarafianos

With two catalytic activities and many substrates, how does HIV's reverse transcriptase enzyme know what to do to which substrate? Zooming in on the enzyme's molecular interactions provides tantalizing clues.

To replicate within their host cell, retroviruses such as HIV must make a double-stranded DNA copy of their single-stranded RNA genome. This elegant process, by which a viral enzyme — reverse transcriptase — synthesizes some 20,000 nucleotides, with the flow of genetic information moving in the opposite direction to normal (DNA to RNA), is nothing short of amazing. It also provides numerous opportunities for therapeutic intervention, as reverse transcriptase is the target of nearly half of the drugs approved for treating AIDS. But reverse transcription is not simple, for the enzyme comes into contact with diverse nucleic-acid substrates (DNA and/or RNA). On page 184 of this issue, Abbondanzieri *et al.*[1] describe how this enzyme discriminates between substrates at intermediate steps of reverse transcription.

Reverse transcriptase has two distinct enzymatic activities: it is a DNA polymerase capable of copying either an RNA or a DNA template into a complementary DNA sequence; and it is an RNase H, capable of degrading the RNA strand of an RNA–DNA duplex into small pieces once it has been used as a template for the first DNA strand (called the 'minus' strand; Fig. 1). But the enzyme overlooks some RNA segments, called polypurine tracts (PPTs), sparing them from degradation; these segments have an unusual sequence (and so structure), which prevents the enzyme's RNase H domain from cutting them[2].

Like other DNA polymerase enzymes, reverse transcriptases initiate DNA synthesis using short nucleic-acid segments called primers. The primer used by HIV reverse transcriptase to generate the minus-strand DNA is a host-cell transfer RNA sequence. For the formation of the second (plus) DNA strand, which is complementary to the minus strand, the enzyme uses PPT segments that it has

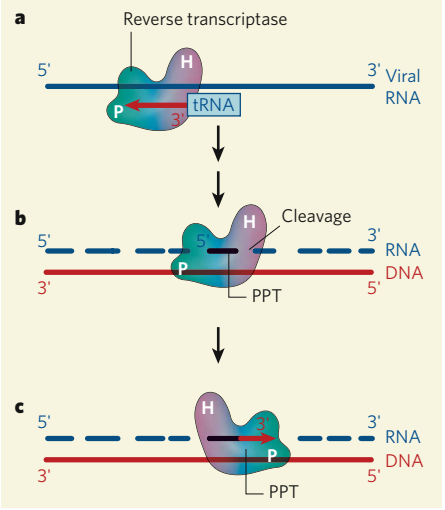


Figure 1 | Reverse transcription. Like other reverse transcriptase enzymes, that of HIV mediates transcription of the virus's RNA genome into double-stranded DNA. **a**, To form the first DNA strand (the minus strand) the enzyme uses a transfer RNA as a primer and interacts with the tRNA 3′ end in a polymerase (P) binding mode. **b**, As the complementary minus-DNA sequence is being synthesized, the enzyme cleaves the RNA template (but leaves the PPT sequences of the RNA intact) by binding to it in an RNase H (H) mode. **c**, Finally, to initiate the synthesis of the second (plus) DNA strand, the reverse transcriptase uses the PPT sequence as a primer, once again binding in the polymerase mode.

©2008 Nature Publishing Group

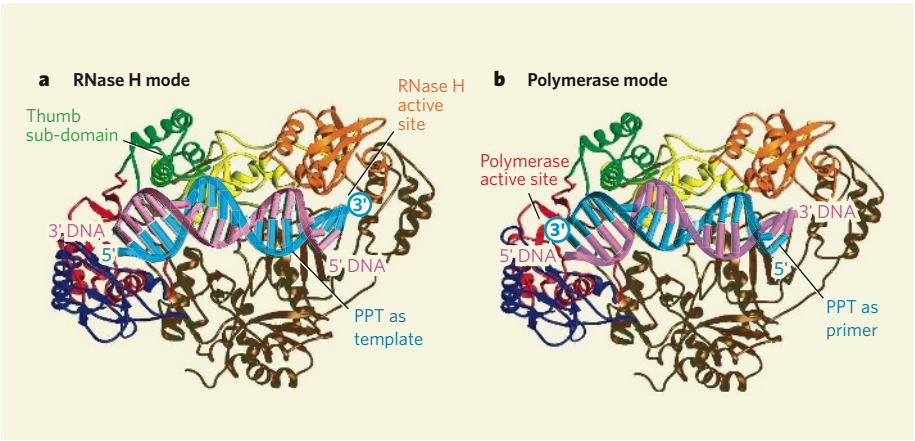


Figure 2 | The dual binding mode of HIV reverse transcriptase. As part of the virus's RNA genome, the PPT sequence acts as a template for the synthesis of the minus-strand DNA. **a**, Abbondanzieri *et al.*[1] show that, before cleavage of the PPT RNA, HIV reverse transcriptase binds to PPT (blue sequence) in an RNase H mode — that is, the RNase H domain of the enzyme is close to the 3′ end of PPT (circled), thus blocking the polymerase domain's access to it and preventing premature extension. **b**, When the RNA nucleotides next to the PPT sequence are cleaved, allowing PPT to function as a primer, the enzyme can bind in a polymerase-competent mode (with its polymerase domain close to PPT's 3′ end) to initiate DNA synthesis. The HIV reverse transcriptase seems to rapidly switch orientation between these two binding modes.

generated by cleaving the template RNA on both sides of these fragments (Fig. 1b).

The molecular details of the interactions between HIV reverse transcriptase and its various substrates have been revealed through structural analyses of the enzyme when in complex with nucleic acids (DNA–DNA or RNA–DNA strands)[2,3]. They have also been studied by analyses of nucleic acids (DNA–DNA) in the presence of dNTPs (the incoming free nucleosides that join and extend a nucleic acid)[4]. Together with complementary biochemical studies, these structures showed that reverse transcriptase interacts with both nucleic acids and the incoming dNTPs when in the polymerization mode, with its polymerase active site aligned at one end (the 3′ end) of a primer. When in the RNA-cleavage mode, it interacts with RNA–DNA, with the active site of the RNase H domain aligned with the RNA strand[2] (Fig. 1). A remaining puzzle has been how reverse transcriptase can distinguish PPT RNA, as an efficient primer for the initiation of second-strand-DNA synthesis, from the rest of the RNA sequence, which, having already been used as a template, is cleaved by RNase H.

Abbondanzieri *et al.*[1] present fascinating clues about the interactions between HIV reverse transcriptase and various nucleic-acid substrates. To determine the orientation in which individual enzyme molecules bind to an RNA–DNA hybrid, the authors used the technique of fluorescence resonance energy transfer (FRET). Typical biochemical experiments involve analysing populations of molecules or complexes. By contrast, single-molecule spectroscopy techniques such as FRET allow direct, time-dependent observation of individual molecular events. For FRET analysis, two molecules of interest — in this case, reverse transcriptase and the nucleic acid — are labelled with fluorescent chemical groups, and the transfer of light energy from one molecule to the other, which depends on the distance between them, is used as a measure of their interactions.

The authors find that, when reverse transcriptase binds to a nucleic acid in which a relatively short RNA segment is hybridized to a longer DNA strand (mimicking the binding of a random primer to the minus-strand DNA), this enzyme binds almost exclusively in an RNase H cleavage mode, and the active site of the polymerase domain does not bind to the 3′ end of the RNA to extend this sequence. But if the RNA is the PPT primer, a portion of the reverse transcriptase binds to it in a polymerase-competent orientation, with the 3′ end of PPT available for extension by the polymerase active site (Fig. 2).

Abbondanzieri and colleagues also show that, in the presence of dNTPs, the number of reverse transcriptase molecules that bind to the PPT primer in the polymerase-competent orientation increases. Moreover, if a short DNA primer is bound to the longer DNA template, most reverse transcriptase enzymes bind in the polymerase-competent orientation. These results indicate that the PPT sequence 'design' is such that not only does it avoid degradation, but it is also an efficient primer for the initiation of second-strand DNA synthesis.

Remarkably, HIV reverse transcriptase can 'flip' its orientation on a nucleic-acid substrate without dissociating from it[1]. Such a transition between two binding orientations is unexpected, given the extensive contacts between this enzyme and its nucleic-acid substrate. The authors also find that an incoming dNTP substrate can bind to a reverse transcriptase in complex with a template–primer duplex (which lacks a hydroxyl group in the 3′ end of the primer to prevent dNTP incorporation), thereby stabilizing the polymerase-competent orientation of reverse transcriptase. Formation of such a 'dead-end' complex has previously been reported[5] to prevent a reaction that unblocks chain-terminated primers and to increase the efficacy of some chain-terminating drugs that target HIV reverse transcriptase.

The authors have made another intriguing observation: the drug nevirapine — an inhibitor of HIV reverse transcriptase that binds in a pocket at the base of the flexible 'thumb' subdomain of the enzyme (Fig. 2a) — allows the enzyme to flip more rapidly. Nevirapine binding forces the thumb subdomain to undergo a conformational change that Abbondanzieri *et al.* propose relaxes the enzyme's grip on the nucleic-acid substrate, thus allowing it to switch orientations more rapidly. This observation implies that certain steps in the process of reverse transcription may be particularly susceptible to antiviral treatments — an idea that has also been proposed previously[6,7].

Abbondanzieri and colleagues' work vividly illustrates the importance of understanding how the orientational dynamics of reverse transcriptase relate to the enzyme's activities. Their findings also raise the question of whether similar surprises are in store with regard to the enzyme's interactions with the host tRNA, during the initiation of minus-strand-DNA synthesis. Another question is whether other multifunctional enzymes indulge in similar molecular acrobatics. Examples of such enzymes are DNA polymerases, which, like reverse transcriptases, have separate non-polymerase activities, or enzymes that interact with nucleic acids, such as helicases, ligases, topoisomerases, integrases and restriction endonucleases. Whatever the answers, this work suggests that reverse transcriptase is even more agile in its handling of nucleic acids than we thought. ■

Eddy Arnold is at the Center for Advanced Biotechnology and Medicine, and the Department of Chemistry and Chemical Biology, Rutgers University, 679 Hoes Lane, Piscataway, New Jersey 08854, USA.
Stefan G. Sarafianos is at the Christopher Bond Life Sciences Center, and in the Department of Molecular Microbiology and Immunology, School of Medicine, University of Missouri, 1201 Rollins Street, Columbia, Missouri 65211, USA.
e-mails: arnold@cabm.rutgers.edu; sarafianoss@missouri.edu

1. Abbondanzieri, E. A. *et al. Nature* **453,** 184–189 (2008).
2. Sarafianos, S. G. *et al. EMBO J.* **20,** 1449–1461 (2001).
3. Jacobo-Molina, A. *et al. Proc. Natl Acad. Sci. USA* **90,** 6320–6324 (1993).
4. Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. *Science* **282,** 1669–1675 (1998).
5. Meyer, P. R., Matsuura, S. E., Mian, A. M., So, A. G. & Scott, W. A. *Mol. Cell* **4,** 35–43 (1999).
6. Palaniappan, C., Kim, J. K., Wisniewski, M., Fay, P. J. & Bambara, R. A. *J. Biol. Chem.* **278,** 3808–3816 (1998).
7. Grobler, J. A. *et al. J. Biol. Chem.* **282,** 8005–8010 (2007).

©2008 Nature Publishing Group

Q&A

MATERIALS SCIENCE

Supramolecular polymers

Tom F. A. de Greef and E. W. Meijer

Most polymers consist of long molecular chains made up of many units connected by covalent bonds — but supramolecular polymers are different. The strikingly dynamic properties of these materials arise from the reversible bonds that hold their chains together, and open up the prospect of many new applications.

Do we need another class of polymer?

In short, yes. Conventional polymers have excellent properties as materials, but when they melt they become highly viscous — the result of entanglement of their macromolecules. High temperatures and pressures are typically required to provide a melt of sufficiently low viscosity for processing, and this limits their applications. But supramolecular polymers combine good material properties with low-viscosity melts that are easy to handle. Some supramolecular polymers also have remarkable characteristics unique to their class, such as the ability to self-heal fractures in their structure.

What exactly are supramolecular polymers?

They are polymeric arrays of monomer units, held together by reversible and highly directional secondary interactions — that is, non-covalent bonds, such as hydrogen bonds. The resulting materials therefore maintain their polymeric properties in solution. The directions and strengths of the interactions are precisely tuned so that the array of molecules behaves as a polymer (that is, it behaves in a way that can be described by the theories of polymer physics). The high reversibility of the non-covalent bonds ensures that supramolecular polymers are always formed under conditions of thermodynamic equilibrium, and hence the lengths of the chains are directly related to the strength of the non-covalent bond, the concentration of the monomer and the temperature.

How do they differ from conventional polymers?

Most, if not all, of the material properties of conventional polymers stem from the covalent nature of their spaghetti-like molecules, although secondary interactions between the molecules are also often involved. For example, hydrogen bonding in nylon affects the properties of the polymer's crystals. Supramolecular polymers, on the other hand, make use of reversible interactions in the main chain. Hence, all the mechanical properties of supramolecular polymers are a direct result of secondary interactions, in particular the strength, reversibility and directionality of these interactions.

What are the origins of supramolecular polymers?

Supramolecular interactions between certain molecules have long been recognized. In most cases, only short chains of up to about 10 units are created, whereas useful supramolecular polymers require at least 100 to 1,000 monomers to be connected at any one time. The key to creating longer chains is to strengthen the interactions between the monomers; put more accurately, the average lifetime of the bonds must be increased. But the window of opportunity is small — if the secondary interactions are too strong, the polymers lose their dynamic flexibility, and their bulk properties become more like those of covalent polymers.

What kick-started the field?

Perhaps the most crucial step was the development in 1997 of a unit that could form four hydrogen bonds, known as 2-ureido-4[1*H*]-pyrimidinone (UPy). A monomer containing two of these self-complementary units self-assembles into polymers (Fig. 1). The relatively long lifetime of the bonds (0.1–1 seconds) was essential for making networks that

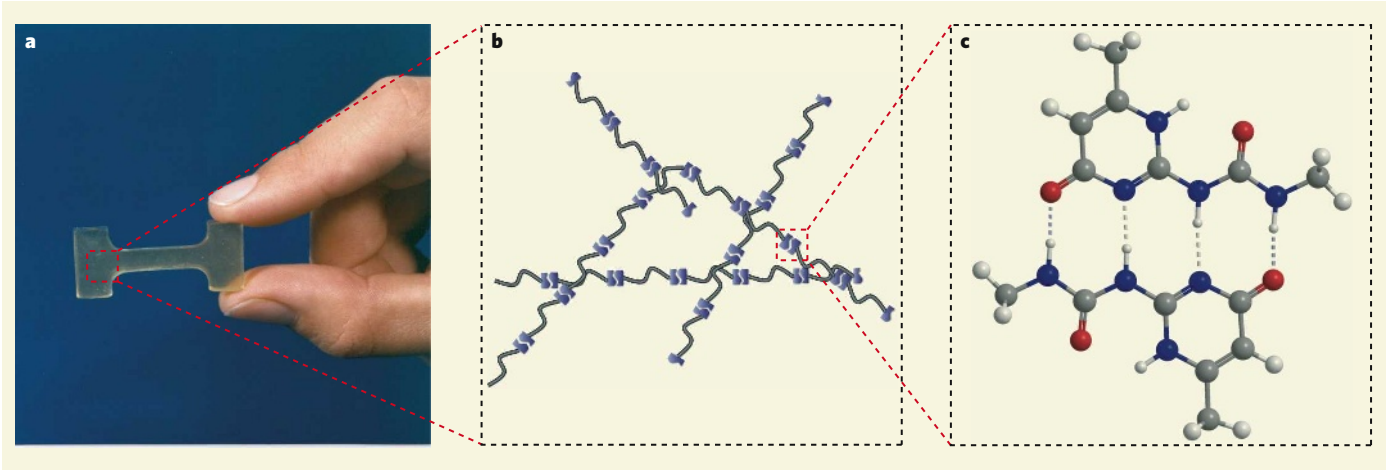


Figure 1 | Zooming in on supramolecular polymers. a, This plastic component is made from a supramolecular polymer. **b**, The molecular structure is a network of monomers connected by hydrogen bonds. Each monomer consists of a hydrocarbon chain with a hydrogen-bonding unit (2-ureido-4[1*H*]-pyrimidinone, UPy, in blue) at each end. **c**, The UPy units form hydrogen bonds to each other (dotted lines). The monomers therefore spontaneously self-assemble to form the polymeric network. Carbon atoms are shown in grey; nitrogen in blue; oxygen in red; and hydrogen in white.

©2008 Nature Publishing Group

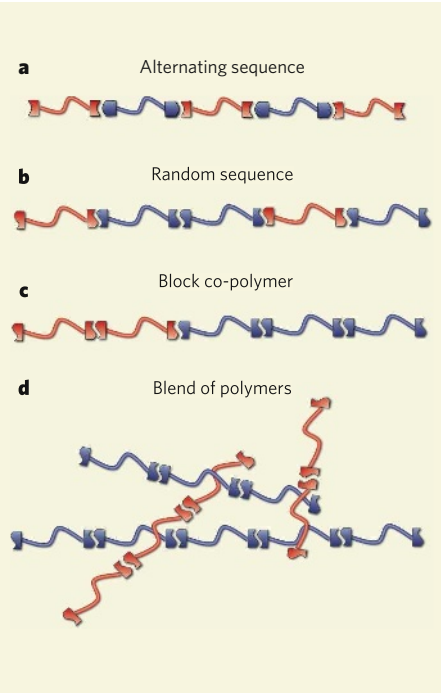


Figure 2 | Monomer ordering in supramolecular co-polymers. These co-polymers form from mixtures of different kinds of self-assembling monomer. The ordering of the monomers depends on the binding modes of their hydrogen-bonding groups. **a**, If the different kinds of group on each monomer form complementary bonds (only to each other), an alternating sequence of monomer types forms. **b**–**d**, If the groups are self-complementary (they bind to their own kind) but are also able to form hydrogen bonds to other kinds of group, there are three different possibilities. **b**, Normally, a random chain of monomers forms. **c**, Alternatively, a chain assembles containing distinct blocks of one kind of monomer. This can happen when, for example, there are two self-complementary arrays AA and BB that can form a self-complementary AB interaction of much lower strength. **d**, If the monomers repel each other and don't mix, the polymer forms as a blend of chains that each contain one kind of monomer.

displayed true polymeric properties. Before the discovery of UPy, supramolecular polymers were based on pairs of different monomers that formed complementary bonds with each other — but even a slight excess of one of the monomers greatly reduced the degree of polymerization.

Are hydrogen bonds the best way to hold supramolecular polymers together?

Hydrogen bonds are ideal, as they combine high strength with excellent reversibility. Furthermore, because they form bonds only in one direction (that of the polymer chain), unwanted interactions in other directions are minimized. The strength and reversibility of hydrogen bonding between monomers can be precisely tuned by installing an array of hydrogen-bonding groups in the monomers, so allowing several bonds to form. The interactions between such strongly bound units are maintained even when the polymer is dissolved in water, where the solvent molecules compete to form hydrogen bonds with the monomers. Nevertheless, in water, the hydrogen bonding between monomers is much weaker than in the solid material. In order to make supramolecular polymers of sufficient length in water, the monomers must form hydrophobic pockets in which hydrogen bonds are protected from the detrimental influence of the solvent.

Are supramolecular polymers easy to make?

Because the polymers self-assemble from monomers (Box 1), the real question is whether the monomers can be easily made. The answer is that it depends on the monomer — some require several steps of sophisticated organic synthesis, but others are easy to make, even on a large scale. Luckily, most of the interesting materials fall into the second category. For instance, the UPy unit is made in one step from commercially available compounds. Furthermore, the monomers for supramolecular polymers can be purified using the plethora of methods that have been developed for small organic molecules. This is an advantage over covalent polymers, which are more difficult to purify.

Can two or more kinds of monomer be combined?

Yes, simply by mixing different monomers together. The properties of the resulting supramolecular co-polymers can be easily tuned in a modular approach by changing the ratio of the monomers. For co-polymers with an alternating sequence of units, the two different monomers should have complementary binding motifs whereas, for random sequences, self-complementary monomers are preferred. But if the monomers strongly repel each other (for example, by hydrophobic interactions), they might not form mixed chains, so that the resulting material actually forms as a blend of two different polymers (Fig. 2). The modular approach to making supramolecular co-polymers is particularly useful for adjusting the properties of materials for medical applications, such as tissue engineering and drug delivery. These materials must be fine-tuned to meet the exact needs of biological systems.

What is the chief selling point of supramolecular polymers?

Perhaps their biggest advantage is the strong dependence of their melt viscosities on temperature — above the material's melting point (or glass transition temperature), small increases in temperature lead to a large reduction in viscosity. This allows the materials to be easily processed, and will be useful for several applications. For example, inks in ink-jet printers need to be runny when they're squirted out of the pen, but must become almost solid when they hit the paper. The properties of supramolecular polymers fit the bill well.

Why do the material properties depend so much on temperature?

Because of the mechanisms of stress relaxation that are available to supramolecular polymers. The dominant pathway for stress relaxation in conventional polymers is reptation — a snake-like movement of the polymeric chains. Stress relaxation in supramolecular polymers also occurs by reptation, but additionally, supramolecular chains can release strain by breaking and then recombining with other, strain-free chain-ends. This last process is unique to supramolecular polymers, and speeds up at higher temperatures. Finally, the chains in these polymers rapidly become shorter with increasing temperature, leading to liquid-like behaviour.

Can this temperature-dependent behaviour be modified?

Certainly. Second-generation supramolecular polymers are designed to incorporate additional interactions — usually lateral hydrogen bonds — between groups at the ends of the polymer chains to strengthen the materials. These end-groups are usually attached to oligomers that have molecular masses of only a few thousand daltons; typical polymer chains have molecular masses of several hundred thousand daltons. In this way, supramolecular polymers have been made that melt at higher temperatures, but that are elastic at temperatures below their melting point. The enhanced material properties of these polymers are a direct result of lateral hydrogen bonding between the end-groups, which causes the material to segregate into long nanofibres. Further studies on this phenomenon will show whether this is a general effect.

Are there other interesting properties?

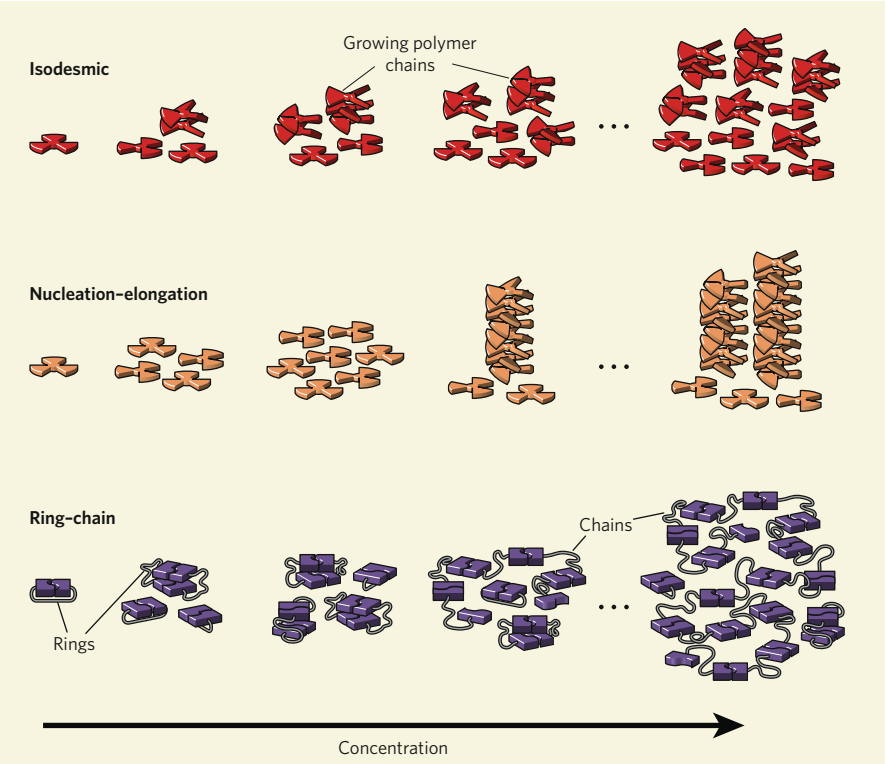
A supramolecular rubber was recently reported that can self-heal: fractures in the sample can be repaired simply by pressing the two broken edges of the material together. The inventors of the rubber propose that fractures occur when hydrogen bonds between monomers in the material are broken; bringing the edges of the fracture together allows the hydrogen bonds to re-form, sealing up the gap. Impressively, the dynamic behaviour of the hydrogen bonds — necessary for the self-healing process — does not compromise the properties of the material.

Which applications are closest to the marketplace?

Supramolecular polymers might find their first use as coatings for delicate, heat-sensitive substrates. For example, wood and fabrics

©2008 Nature Publishing Group

Box 1 | Mechanisms of supramolecular polymerization



There are three main growth mechanisms for supramolecular polymerization. The first, known as isodesmic growth, occurs when the strength of secondary interactions between monomers in the polymer chains is unaffected by the length of the chain. Each monomer added to a growing chain therefore increases the thermodynamic stability of the system by the same amount. Because each additional step is equivalent, no critical temperature or concentration of monomers is required for polymerization to occur. Instead, the concentration of polymer chains rises gradually as the concentration of monomers in the solution is increased, or as the temperature increases.

The second process — the nucleation–elongation mechanism — occurs in the growth of 'ordered' supramolecular polymers, such as those that form helices. This involves two distinct phases of self-assembly: a slow nucleation phase followed by a more rapid growth phase. In this mechanism, the non-covalent bonds between monomers in chains are weak, hindering the initial formation of the chains. But once a stable nucleus is formed, further monomer addition becomes favourable, at which point polymer growth accelerates. Long polymer chains will form only above a certain temperature and minimum concentration of monomer.

The third growth process is known as the ring–chain mechanism. This is a special case of nucleated polymerization that also occurs in covalent polymers; it is characterized by an equilibrium between closed rings of monomers and linear chains. In this mechanism, below a certain monomer concentration, the ends of any small polymer chains are more likely to encounter and react with each other (forming a ring) than they are to find another monomer that would result in chain elongation. Above the critical concentration, chain formation becomes more favourable and polymer growth occurs rapidly. As in the nucleation–elongation mechanism, the degree of polymerization changes abruptly once critical conditions are reached. **T.F.A.G. & E.W.M.**

can't usually be coated with covalent polymers because they would be damaged by the hot polymer melts. But this wouldn't be a problem for supramolecular polymers with their relatively low melting points. Supramolecular polymers can also be made into excellent adhesives, and their potential as inks was mentioned earlier. Finally, the materials are also attractive for any applications that require processing from a solution or a gel, such as cosmetics and personal-care products.

Are there any naturally occurring supramolecular polymers?

Yes, but not many. Some protein molecules form small networks in order to function, but don't make long enough chains to be classed as polymers. Others — such as G-actin and tubulin, which act as building-blocks for the scaffolding of cells — genuinely polymerize to form filaments, but the mechanism involves an irreversible reaction with phosphate-containing 'fuel' molecules. The polymerization is therefore not an equilibrium process, allowing the resulting materials to display some exciting properties. For example, tubulin dimers form long, rigid polymers known as microtubules, which can synchronize their growth (or shrinkage). But the greatest resemblance to synthetic supramolecular polymerization is seen in the self-assembly of the tobacco mosaic virus, as the assembly process occurs without the use of fuel molecules.

What will the future bring?

Supramolecular polymers will become an integral part of the general field of self-assembling polymers, and will find applications in areas ranging from electronics to medicine. The complexity of their molecular systems will increase rapidly, leading in turn to the development of yet more unconventional polymeric properties. But the most intriguing challenge will be the search for self-organizing (rather than self-assembling) supramolecular polymers.

What are self-organized polymers?

These materials will also be formed from monomers using secondary interactions, but the polymerization will be far from an equilibrium process — in self-organized structures, external energy is required to maintain a steady state, along with an open system that allows a continuous inflow of monomers. In theory, such a material would display both spatial and temporal order, on a much larger length- and timescale than seen in structures formed under equilibrium conditions. Self-organized polymers will thus mimic the dynamic behaviour of proteins such as tubulin. The critical requirements for self-organizing supramolecular polymers are auto-catalytic reactions (in which the products of reactions between monomers catalyse further polymerization reactions), a supply of chemical energy, and processes that involve competing diffusion of the molecules involved in the polymerization. These requirements have yet to be met in synthetic systems. With so much work still to be done, there has never been a more exciting time to work with supramolecular polymers. ■

Tom F. A. de Greef and E. W. Meijer are in the Laboratory of Macromolecular and Organic Chemistry, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, the Netherlands.
e-mails: t.f.a.d.greef@tue.nl; e.w.meijer@tue.nl

FURTHER READING
Brunsveld, L., Folmer, B. J. B., Meijer, E. W. & Sijbesma, R. P. *Chem. Rev.* **101,** 4071–4098 (2001).
Ciferri, A. (ed.) *Supramolecular Polymers* (CRC, New York, 2005).
Lehn, J.-M. *Polym. Int.* **51,** 825–839 (2002).
Binder, W. (ed.) *Hydrogen Bonded Polymers* (Springer, New York, 2007).
www.supramolix.com/index.php?page=supramolecular
Dankers, P. Y. W. & Meijer, E. W. *Bull. Chem. Soc. Jpn* **80,** 2047–2073 (2007).
Cordier, P., Tournilhac, T., Soulié-Ziakovic, C. & Leibler, L. *Nature* **451,** 977–980 (2008).

Competing financial interests: declared (see online article for details).

©2008 Nature Publishing Group

ARTICLES

Genome analysis of the platypus reveals unique signatures of evolution

A list of authors and their affiliations appears at the end of the paper

We present a draft genome sequence of the platypus, *Ornithorhynchus anatinus*. This monotreme exhibits a fascinating combination of reptilian and mammalian characters. For example, platypuses have a coat of fur adapted to an aquatic lifestyle; platypus females lactate, yet lay eggs; and males are equipped with venom similar to that of reptiles. Analysis of the first monotreme genome aligned these features with genetic innovations. We find that reptile and platypus venom proteins have been co-opted independently from the same gene families; milk protein genes are conserved despite platypuses laying eggs; and immune gene family expansions are directly related to platypus biology. Expansions of protein, non-protein-coding RNA and microRNA families, as well as repeat elements, are identified. Sequencing of this genome now provides a valuable resource for deep mammalian comparative analyses, as well as for monotreme biology and conservation.

The platypus (*Ornithorhynchus anatinus*) has always elicited excitement and controversy in the zoological world[1]. Some initially considered it to be a true mammal despite its duck-bill and webbed feet. The platypus was placed with the echidnas into a new taxon called the Monotremata (meaning 'single hole' because of their common external opening for urogenital and digestive systems). Traditionally, the Monotremata are considered to belong to the mammalian subclass Prototheria, which diverged from the therapsid line that led to the Theria and subsequently split into the marsupials (Marsupialia) and eutherians (Placentalia). The divergence of monotremes and therians falls into the large gap in the amniote phylogeny between the eutherian radiation about 90 million years (Myr) ago and the divergence of mammals from the sauropsid lineage around 315 Myr ago (Fig. 1). Estimates of the monotreme–theria divergence time range between 160 and 210 Myr ago; here we will use 166 Myr ago, recently estimated from fossil and molecular data[2].

The most extraordinary and controversial aspect of platypus biology was initially whether or not they lay eggs like birds and reptiles. In 1884, William Caldwell's concise telegram to the British Association announced "Monotremes oviparous, ovum meroblastic", not holoblastic as in the other two mammalian groups[3,4]. The egg is laid in an earthen nesting burrow after about 21 days and hatches 11 days later[5,6]. For about 4 months, when most organ systems differentiate, the young depend on milk sucked directly from the abdominal skin, as females lack nipples. Platypus milk changes in protein composition during lactation (as it does in marsupials, but not in most eutherians[5]). The anatomy of the monotreme reproductive system reflects its reptilian origins, but shows features typical of mammals[7], as well as unique specialized characteristics. Spermatozoa are filiform, like those of birds and reptiles, but, uniquely among amniotes, form bundles of 100 during passage through the epididymis. Chromosomes are arranged in defined order in sperm[8] as they are in therians, but not birds[9]. The testes synthesize testosterone and dihydrotestosterone, as in therians, but there is no scrotum and testes are abdominal[10].

Other special features of the platypus are its gastrointestinal system, neuroanatomy (electro-reception) and a venom delivery system, unique among mammals[11]. Platypus is an obligate aquatic feeder that relies on its thick pelage to maintain its low (31–32 °C) body temperature during feeding in often icy waters. With its eyes, ears and nostrils closed while foraging underwater, it uses an electro-sensory system in the bill to help locate aquatic invertebrates and other prey[12,13]. Interestingly, adult monotremes lack teeth.

The platypus genome, as well as the animal, is an amalgam of ancestral reptilian and derived mammalian characteristics. The platypus karyotype comprises 52 chromosomes in both sexes[14,15], with a few large and many small chromosomes, reminiscent of reptilian macro- and microchromosomes. Platypuses have multiple sex chromosomes with some homology to the bird Z chromosome[16]. Males have five X and five Y chromosomes, which form a chain at meiosis and segregate into 5X and 5Y sperm[17,18]. Sex determination and sex chromosome dosage compensation remain unclear.

Platypuses live in the waterways of eastern and southern Australia, including Tasmania. Its secretive lifestyle hampers understanding of its population dynamics and the social and family structure. Platypuses are still relatively common in the wild, but were recently reclassified as 'vulnerable' because of their reliance on an aquatic environment that is under stress from climate change and degradation by human activities. Water quality, erosion, destruction of habitat and food resources, and disease now threaten populations. Because the platypus has rarely bred in captivity and is the last of a long line of ornithorhynchid monotremes, their continued survival is of great importance. Here we describe the platypus genome sequence and compare it to the genomes of other mammals, and of the chicken.

Sequencing and assembly

All sequencing libraries were prepared from DNA of a single female platypus (Glennie; Glenrock Station, New South Wales, Australia) and were sequenced using established whole-genome shotgun (WGS) methods[19]. A draft assembly was produced from ~6× coverage of whole-genome plasmid, fosmid and bacterial artificial chromosome (BAC) reads (Supplementary Table 1) using the assembly program PCAP[20] (Supplementary Notes 1). A BAC-based physical map was developed in parallel with the sequence assembly and subsequently integrated with the WGS assembly to provide the primary means of scaffolding the assembly into larger ordered and oriented groupings (ultracontigs; Supplementary Notes 2 and 3 and Supplementary Table 2). Because there were no platypus linkage maps available, we used fluorescent *in situ* hybridization (FISH) to localize a subset of the sequence scaffolds to chromosomes following the agreed nomenclature[21]. Of the 1.84 gigabases (Gb) of assembled sequence, 437 megabases (Mb) were ordered and oriented along 20 of

©2008 Nature Publishing Group

the platypus chromosomes. We analysed numerous metrics of assembly quality (Supplementary Notes 4–11) and we conclude that despite the adverse contiguity, the existing platypus assembly, given its structural and nucleotide accuracy, provides a reasonable substrate for the analyses presented here.

Non-protein-coding genes

In general, the platypus genome contains fewer computationally predicted non-protein-coding (nc)RNAs (1,220 cases excluded high repetitive small nucleolar RNA (snoRNA) copies; see below) than do other mammalian species (for example, human with 4,421 Rfam hits), similar to observations in chicken[19] (655 Rfam-based ncRNAs). This is probably because of the extensive retrotransposition of ncRNAs in therian mammals and the apparent lack of L1-mediated retrotransposition in chicken and platypus. The exception to this is the platypus family of snoRNAs, which is markedly expanded (~2,000 matches to the Rfam covariant models) compared to that for therian mammals (~200). snoRNAs are involved in RNA modifications, in particular of ribosomal RNA, and are often located in introns of protein-coding genes[22]. Our investigations revealed a novel short-interspersed-element (SINE)-like, snoRNA-related retrotransposon—which we have labelled snoRTEs—that has duplicated in platypus to ~40,000 full-length or truncated copies. It is retrotransposed by means of retrotransposon-like non-LTR (long terminal repeat) transposable elements (RTE) as opposed to the L1-mediated transposition mechanism in therians[23]. We constructed a complementary DNA library of small, ncRNAs and identified 371 consensus sequences of small RNAs that included 166 snoRNAs[23] (Supplementary Table 3). Ninety-nine of these cloned snoRNAs are found in paralogous families, and 21 of them belong to the snoRTE class. The presence of both the structural requirements known to be important in snoRNA function[24] and evidence of their expression are consistent with these snoRTE elements being functional in the platypus. Similar to other unrelated ncRNAs that have proliferated in therian mammals (for example, 7SL RNA-derived primate *Alu* elements, tRNA-derived rodent identifier (ID) elements), this recent SINE-like expansion is probably due to chance events. However, given the RNA modification activity of snoRNAs, and our increasing awareness of the cellular importance of RNA molecules, it might be that some of the retrotranspositionally duplicated RNAs were exapted into new functions in this species.

Other small RNAs. Overall, we found commonalities with small RNA (sRNA) pathways of other mammals, but also features that are unique to monotremes. Components of the RNA interference machinery are conserved in platypus, including elements of biogenesis pathways (Dicer and Drosha) and RNA-interference effector complexes (argonaute proteins; Supplementary Table 4). Of 20,924,799 platypus and echidna sRNA reads derived from liver, kidney, brain, lung, heart and testis, 67% could be assigned to known microRNA (miRNA) families. Established patterns of miRNA expression were generally recapitulated in monotremes.

To determine the conservation patterns of miRNAs in platypus, we identified platypus miRNAs sharing at least 16-nucleotide identity with miRNAs in eutherian mammals (mouse/human) and chicken. Although most conserved miRNAs were identified across these vertebrate lineages (137 miRNAs), 10 miRNAs were shared only with eutherians (mouse/human) and 4 only with chicken (Fig. 2a). miRNAs can be classified into families based on identity of the functional 'seed' region at position 2–8 of the mature miRNA strand. We identified miRNA families that were shared between platypus and eutherians but not chicken (40 families), or between platypus and chicken but not eutherians (8 families), suggesting that for some miRNAs only

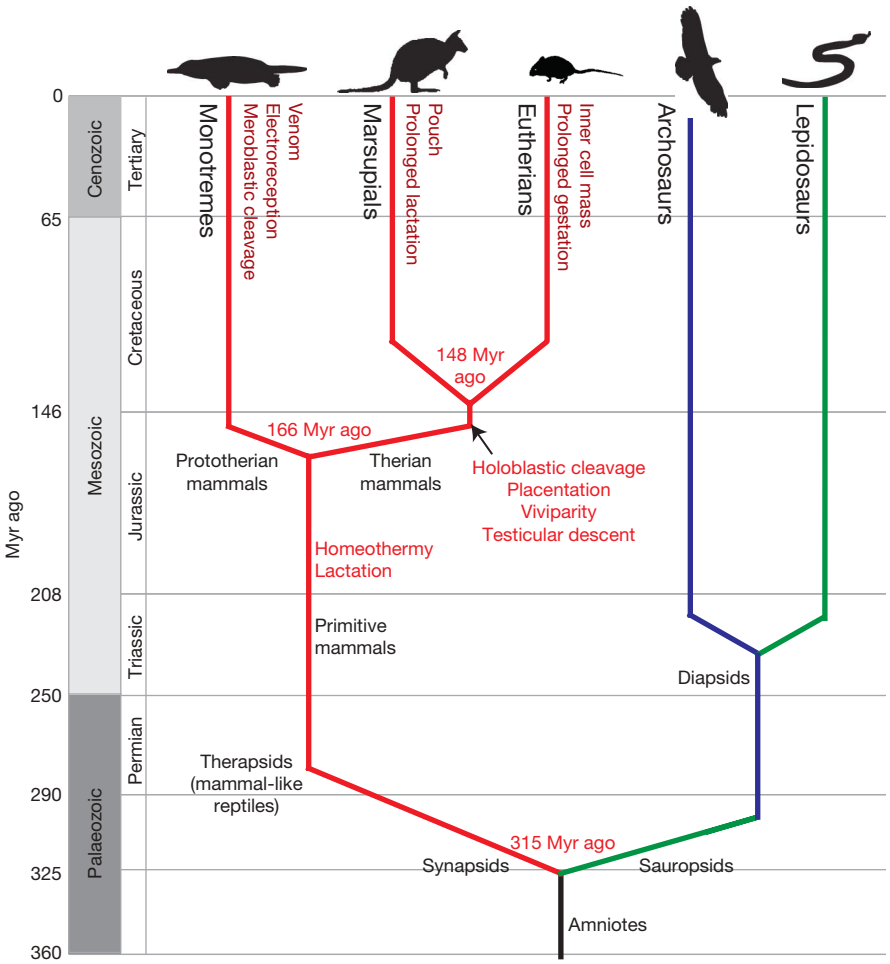


Figure 1 | Emergence of traits along the mammalian lineage. Amniotes split into the sauropsids (leading to birds and reptiles) and synapsids (leading to mammal-like reptiles). These small early mammals developed hair, homeothermy and lactation (red lines). Monotremes diverged from the therian mammal lineage ~166 Myr ago[2] and developed a unique suite of characters (dark-red text). Therian mammals with common characters split into marsupials and eutherians around 148 Myr ago[2] (dark-red text). Geological eras and periods with relative times (Myr ago) are indicated on the left. Mammal lineages are in red; diapsid reptiles, shown as archosaurs (birds, crocodilians and dinosaurs), are in blue; and lepidosaurs (snakes, lizards and relatives) are in green.

©2008 Nature Publishing Group

the seed region may have been selectively conserved (Fig. 2a). Conserved miRNAs tended to be more robustly expressed in the platypus tissues analysed than lineage-restricted miRNAs (Fig. 2b).

To identify miRNAs unique to monotremes we used a heuristic search that identifies miRNA candidates in deep-sequencing data sets[25]. This method predicted 183 novel miRNAs in platypus and echidna (Fig. 2a). Notably, 92 of these lay in 9 large clusters, on platypus chromosome X1 and contigs 1754, 7160, 7359, 8388, 11344, 22847, 198872 and 191065. Physical mapping confirmed that at least five of these contigs are linked to the long arm of chromosome X1 (ref. 25). These abundantly expressed clusters were sequenced almost exclusively from platypus and echidna testis (Fig. 2b). The expansion of this unique miRNA class and its expression domain suggest possible roles in monotreme reproductive biology[25].

Piwi-interacting RNAs (piRNAs) associate with a germline-expressed clade of argonaute proteins, known as Piwis[26], and have a role in transposon silencing and genome methylation[26]. Monotreme piRNAs bear strong structural similarity to those in eutherians. They are ~29 nucleotides in length and arise from large testis-specific genomic clusters with distinct genomic strand asymmetry, often with a typical 'bidirectional' organization. We identified 50 major platypus piRNA clusters as well as numerous smaller clusters[25]. In contrast to piRNAs in mouse, platypus piRNAs are repeat-rich and bear strong signatures of active transposon defence.

Gene evolution

We set out to define the protein-coding gene content of platypus to illuminate both the specific biology of the monotreme clade and for comparisons to eutherians and marsupials, or to chicken, the representative sauropsid. Protein-coding genes were predicted using the established Ensembl pipeline[27] suitably modified for platypus (Supplementary Notes 14), with a greater emphasis placed on similarity matches to mammalian genes. Overall this resulted in 18,527 protein-coding genes being predicted from the current platypus assembly. The number of platypus protein-coding genes thus is similar to estimates (18,600–20,800) for human and opossum[28,29].

We were interested first in identifying platypus genes that contribute most to core biological functions that are conserved across the mammals. These will typically be 'simple' 1:1 orthologues, genes that have remained as single copies without duplication or deletion in platypus, in Eutheria (specifically, in dog, human and mouse) and in opossum, a representative marsupial. Subsequently, we considered genes that have been duplicated or deleted in the monotreme lineage, or that have been lost in eutherian and/or marsupial lineages. Such genes are proposed to contribute most to the lineage-specific biological functions that distinguish individual mammals[30]. These studies required the use of an outgroup species, here chicken, a representative of the sauropsids.

As expected, the majority of platypus genes (82%; 15,312 out of 18,596) have orthologues in these five other amniotes (Supplementary Table 5). The remaining 'orphan' genes are expected to primarily reflect rapidly evolving genes, for which no other homologues are discernible, erroneous predictions, and true lineage-specific genes that have been lost in each of the other five species under consideration. Simple 1:1 orthologues, which have been conserved without duplication, deletion or non-functionalization across the five mammalian species, were greatly enriched in housekeeping functions, such as metabolism, DNA replication and mRNA splicing (Supplementary Table 6).

We then identified evolutionary lineages that experienced the most stringent purifying selection. The mouse terminal lineage exhibited a significantly higher degree of purifying selection (the ratio of amino acid replacement to silent substitution rates, $d_N/d_S = 0.105$, $P < 0.001$) than dog, opossum and chicken terminal branches (values of 0.123–0.128); human and platypus terminal lineages showed significantly reduced purifying selection (both 0.132, $P < 0.03$). These values probably reflect the increased efficiency of purifying selection in populations of larger effective size, such as that of mouse[31]. We find that at least one nucleotide substitution has occurred, on average, in synonymous sites of platypus and human orthologues since their last common ancestor (Supplementary Notes 17 and Supplementary Fig. 1). This means that most neutral sequence cannot be aligned accurately between monotreme and eutherian genomes.

Next, we determined the genetic distance of echidna (*Tachyglossus aculeatus*) from platypus. The median d_S value of 0.125 for the orthologues of echidna and platypus, when compared to the value for the monotreme lineage, predicts that platypus and echidna last shared a common ancestor 21.2 Myr ago. Although similar to previous estimates[32], this value seems to be at odds with fossil evidence, perhaps owing to relatively recent reductions of mutational rates in the monotreme lineage[33].

Monotreme biology

We next investigated whether the ancestral reptilian characters of monotremes are reflected in the set of genes that have been retained in platypus, sauropsids and other vertebrates from outside of the amniote clade (such as frogs and fish), but have been lost from eutherian and marsupial lineages (Fig. 1). These ancestral, sauropsid-like, characters of platypus include oviparity (egg laying) and the outward appearances of its spermatozoa and retina. Simultaneously, we sought genetic evidence within the platypus genome both for characteristics peculiar to monotremes, such as venom production and

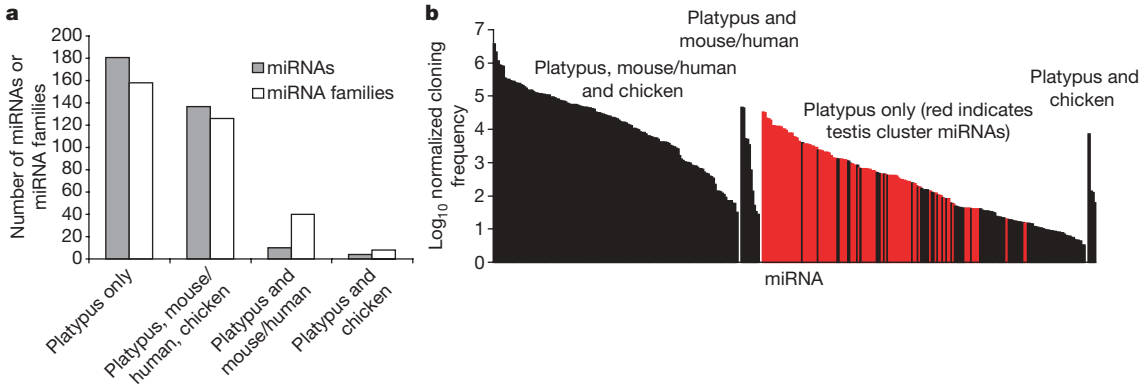


Figure 2 | Platypus miRNAs. a, Platypus has miRNAs shared with eutherians and chickens, and a set that is platypus-specific. miRNAs cloned from six platypus tissues were assigned to families based on seed conservation. Platypus miRNAs and families were divided into classes (indicated) based on their conservation patterns with eutherian mammals (mouse/human) and with chicken. **b**, Expression of platypus miRNAs. The cloning frequency of each platypus mature miRNA sequenced more than once is represented by a vertical bar and clustered by conservation pattern. miRNAs from a set of monotreme-specific miRNA clusters that are expressed in testis are shaded in red.

©2008 Nature Publishing Group

electro-reception, and for characteristics unique to mammals, in particular lactation. By investigating platypus homologues of genes already known to be involved in specific physiological processes (see Methods), we highlight those platypus genes for which evolution exemplifies the ancestral or derived physiological characters of monotremes.

Chemoreception. The semi-aquatic platypus was expected to sense its terrestrial, but not aquatic, environment by detecting airborne odorants using olfactory receptors and vomeronasal receptors (types 1 and 2: V1Rs, V2Rs). Nevertheless large numbers of odorant receptor, V1R and V2R homologues (approximately 700, 950 and 80, respectively) are apparent in the platypus genome assembly, although for each family only a minority lack frame disruptions (approximately 333, 270 and 15, respectively)[34]. Many of these platypus genes and pseudogenes are monophyletic, having arisen by duplication in the 166 Myr since the last common ancestor of monotremes and therians. Although mouse and rat genomes possess greater numbers of odorant receptors and V2Rs than the platypus genome[35,36], the platypus repertoire of V1Rs, showing undisrupted reading frames, is the largest yet seen, 50% more than for mouse (Fig. 3b). This is particularly noteworthy as the *Anolis carolinensis* lizard (sequence data used with the permission of the Broad Institute) and the chicken[19] seem to possess no such receptors. The large expansion of the platypus V1R gene family might reflect sensory adaptations for pheromonal communication or, more generally, for the detection of water-soluble, non-volatile odorants, during underwater foraging.

The platypus odorant receptor gene repertoire is roughly one-half as large as those in other mammals[37]. Nevertheless, platypus odorant receptors fall into class, family and subfamily structures that are well represented from across the mammals, with a few notable exceptions such as family 14 (Fig. 3a). Together with the finding that lizard contains only ~200 odorant receptor genes and pseudogenes, this indicates that the platypus olfactory repertoire is, as expected, more akin to other mammals than it is to sauropsids.

Eggs. Fertilization in the platypus exhibits both sauropsid and therian characteristics. Platypus ova are small (4 mm diameter) relative to comparably sized reptiles and birds, and eggs hatch at an early stage of development so that most growth of the embryo and infant is dependent on lactation, as in marsupials. Like all mammals and many other amniotes, when fertilization occurs the ovum is invested with a zona pellucida. The platypus genome encodes each of the four proteins of the human zona pellucida[38], as well as two ZPAX genes (Table 1) that previously were observed only in birds, amphibians and fish. The aspartyl-protease nothepsin is present in platypus, but has been lost from marsupial and eutherian genomes (Table 1). In zebrafish, this gene is specifically expressed in the liver of females under the action of oestrogens, and accumulates in the ovary[39]. These are the same characteristics as of the vitellogenins, indicating that nothepsin may be involved in processing vitellogenin or other egg-yolk proteins. We find that platypus has retained a single vitellogenin gene and pseudogene, whereas sauropsids such as chicken have three and the viviparous marsupials and eutherians have none.

Spermatozoa. Orthologues of many of the eutherian sperm membrane proteins related to fertilization[40] are present in platypus (and marsupial) genomes. These include the genes for a number of putative zona pellucida receptors and proteins implicated in sperm–oolemma fusion. Testis-specific proteases, which in eutherians participate in degradation of the zona pellucida during fertilization, are all absent from the platypus genome assembly.

Monotreme spermatozoa undergo some post-testicular maturational changes, including the acquisition of progressive motility, loss of cytoplasmic droplets and aggregation of single spermatozoa into bundles during passage through the epididymis[11]. Nevertheless, maturational changes in the sperm surface that are both unique and essential in other mammals for fertilization of the ovum have yet to be identified. Also, the epididymis of monotremes is not highly adapted for sperm storage as in most marsupial and eutherian mammals. Consistent with these findings is the absence of platypus genes for the epididymal-specific proteins that have been implicated in sperm maturation and storage in other mammals. The most abundant secreted protein in the platypus epididymis is a lipocalin, the homologues of which are the most secreted proteins in the reptilian epididymis[41]. Notably, *ADAM7*, a protease that is secreted in the epididymis of eutherians, has an orthologue in the platypus. This is a bona fide protease with a characteristic Zn^{2+}-coordinating sequence HExxH in the platypus, in the opossum and the tree shrew (*Tupaia belangeri*). However, loss of its proteolytic activity is predicted in eutherians[42] owing to a single point mutation within its active site (E to Q).

Lactation and dentition. Lactation is an ancient reproductive trait whose origin predates the origin of mammals. It has been proposed that early lactation evolved as a water source to protect porous parchment-shelled eggs from desiccation during incubation[43] or as a protection against microbial infection. Parchment-shelled egg-laying monotremes also exhibit a more ancestral glandular mammary patch or areola without a nipple that may still possess roles in egg protection. However, in common with all mammals, the milk of monotremes has evolved beyond primitive egg protection into a true milk that is a rich secretion containing sugars, lipids and milk proteins with nutritional, anti-microbial and bioactive functions. In a reflection of this eutherian similarity platypus casein genes are tightly clustered together in the genome, as they are in other mammals, although platypus contains a recently duplicated β-casein gene (Supplementary Fig. 2).

Mammalian casein genes are thought to have originally arisen by duplication of either enamelin or ameloblastin[44], both of which are tooth enamel matrix protein genes that are located adjacent to the casein gene cluster in eutherians and, we find, also in platypus. Adult platypuses, as well as echidnas, lack teeth but the conservation of these enamel protein genes is consistent with the presence of teeth and enamel in the juvenile, as well as the fossil platypuses[45].

Venom. Only a handful of mammals are venomous, but the male platypus is unique among them in delivering its poison not via a bite but from hind-leg spurs. Despite the obvious difficulties in obtaining

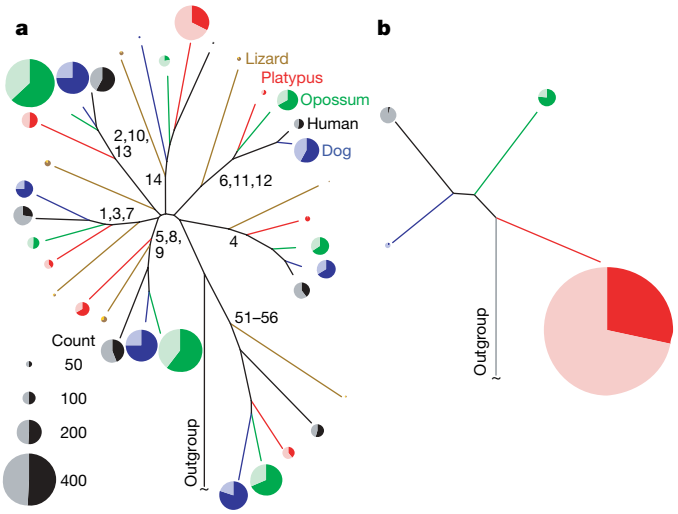


Figure 3 | The platypus chemosensory receptor gene repertoire. a, b, The platypus genome contains only few olfactory receptor genes from olfactory receptor families that are greatly expanded among therians (three other mammals and a reptile shown), but many genes in olfactory receptor family 14 (**a**), and relatively numerous vomeronasal type 1 (V1R) receptors (**b**). These schematic phylogenetic trees show relative family sizes and pseudogene contents of different gene families (enumerated beside internal branches) and the V1R repertoire in platypus. Pie charts illustrate the proportions of intact genes (heavily shaded) versus disrupted pseudogenes (lightly shaded).

©2008 Nature Publishing Group

samples, it is now known that platypus venom is a cocktail of at least 19 different substances[46] including defensin-like peptides (vDLPs), C-type natriuretic peptide (vCNP) and nerve growth factor (vNGF). When analysed phylogenetically and mapped to the platypus genome assembly, these sequences are revealed to have arisen from local duplications of genes possessing very different functions (Fig. 4). Notably, duplications in each of the β-defensin, C-type natriuretic peptide and nerve growth factor gene families have also occurred independently in reptiles during the evolution of their venom[47]. Convergent evolution has thus clearly occurred during the independent evolution of reptilian and monotreme venom[48].

Immunity. Although the major organs of the monotreme immune system are similar to those of other mammals[49], the repertoire of immunity molecules shows some important differences from those of other mammals. In particular, the platypus genome contains at least 214 natural killer receptor genes (Supplementary Notes 18) within the natural killer complex, a far larger number than for human (15 genes[50]), rat (45 genes[50]) or opossum (9 genes[51]).

Both platypus and opossum genomes contain gene expansions in the cathelicidin antimicrobial peptide gene family (Supplementary Fig. 3). Among eutherians, primates and rodents have a single cathelicidin gene[52,53], whereas sheep and cows have numerous genes that have been duplicated only recently[54]. The expanded repertoire of cathelicidin genes in both marsupials and monotremes may arm their immunologically naive young with a diverse arsenal of innate immune responses. In eutherians, with their increases in length of gestation and advances in development *in utero* of their immune systems, the diversity of antimicrobial peptide genes may have become less critical. The platypus genome also contains an expansion in the macrophage differentiation antigen *CD163* gene family (Supplementary Notes 18).

Genome landscape

First, we analyse the phylogenetic position of platypus and confirm that marsupials and eutherians are more closely related than either is to monotremes (Supplementary Notes 19). We then describe platypus chromosomes and observe some properties of platypus interspersed and tandem repeats. We also discuss a potential relationship between interspersed repeats and genomic imprinting and investigate how the extremely high G+C fraction in platypus affects the strong association seen in eutherians between CpG islands and gene promoters.

Platypus chromosomes. Platypus chromosomes provide clues to the relationship between mammal and reptile chromosomes, and to the origins of mammal sex chromosomes and dosage compensation. Our analysis provides further insight with the following findings: the 52 platypus chromosomes show no correlation between the position of orthologous genes on the small platypus chromosomes and chicken microchromosomes; for the unique 5X chromosomes of platypus we reveal considerable sequence alignment similarity to chicken Z and no orthologous gene alignments to human X, implying that the platypus X chromosome evolved directly from a bird-like ancestral reptilian system[55]; and the genes on the five platypus X chromosomes appear to be partially dosage compensated (Supplementary Fig. 5), perhaps parallel to the incomplete dosage compensation recently described in birds[56].

Repeat elements. About one-half of the platypus genome consists of interspersed repeats derived from transposable elements. The most abundant and still active repeats are (severely truncated) copies of the 5-kb long-interspersed-element (LINE2) and its non-autonomous SINE-companion mammalian-wide interspersed repeat (MIR, Mon-1 in monotremes) that became extinct in marsupials and in eutherians 60–100 Myr ago. We estimate that there are 1.9 and 2.75 million copies of LINE2 and MIR/Mon-1, respectively, in the 2.3-Gb platypus genome. DNA transposons and LTR retroelements are quite

Table 1 | Platypus genes that have been lost from the eutherian lineage

Description	Platypus Ensembl gene	Proposed function
Retinal guanylate cyclase activator 1A	ENSOANG00000012043	In zebrafish, expressed in retina
Enoyl-CoA hydratase/isomerase	ENSOANG00000012890	Involved in fatty acid metabolism
Ferric reductase/cytochrome *b*561	ENSOANG00000019725	Absorption of dietary iron
Nothepsin, aspartic proteinase	ENSOANG00000005955	Processes egg-yolk proteins
Glutamine synthetase	ENSOANG00000008089	Role in nitrogen metabolism
Vitellogenin II	Contig 10010	Major egg-yolk protein
Cytochrome P450, CYP2-like	ENSOANG00000004537	Toxin degradation
ATP6AP1 paralogue	ENSOANG00000004825	Retinal pigmentation
Organic solute transporter alpha (2 genes)	Ultracontig 462, Contig 159089	Bile acid transport
Neuropeptide Y7 receptor	ENSOANG00000014966	Regulator of food intake
Melatonin receptor 1C	ENSOANG00000011638	Circadian rhythm regulation
Epidermal differentiation-specific proteins (3 genes)	ENSOANG00000005335, ENSOANG00000003767, ENSOANG00000013512	Neural and epidermal differentiation
TRPV7/TRPV8 transient receptor potential cation channels	ENSOANG00000015080, ENSOANG00000015083	Novel epithelial calcium channels
Shortwave-sensitive-2 (SWS2) opsin gene	Ultracontig 401	Cone visual pigment
Opsin 5 paralogue	ENSOANG00000009478	Light-sensitive receptor
Indigoidine synthase A	Contig 29616	Pigmentation
ZPAX, egg envelope glycoprotein	ENSOANG00000007840, ENSOANG00000002187	Egg envelope protein
Galanin receptor	ENSOANG00000020606	Neuropeptide receptor
Kainate-binding protein	ENSOANG00000007006	Glutamate receptor
Anti-dorsalizing morphogenetic protein	ENSOANG00000002980	Patterning of the body axis during gastrulation
Retinal genes (2 genes)	ENSOANG00000001054, ENSOANG00000004065	Unknown function
Uteroglobin-like secretoglobins (3 genes)	ENSOANG00000020019, ENSOANG00000022350, ENSOANG00000021122	Unknown function
Testis homeobox C14-like proteins (>2 genes)	ENSOANG00000020069, ENSOANG00000022694	Unknown function
Parvalbumin	ENSOANG00000000764	Muscle function
Slc7a2-prov protein	ENSOANG00000009602	Cationic amino acid transporter
Cystine/glutamate transporter	ENSOANG00000005615	Amino acid transporter
SOUL protein	ENSOANG00000013998	Retina and pineal gland haem protein, oxygen sensing
Twin-pore potassium channel Talk-1-like	ENSOANG00000011839	Potassium channel
Alpha-aspartyl dipeptidase	ENSOANG00000009001	Unknown function
Monovalent cation/H^+ antiporter	ENSOANG00000012961	Unknown function; conserved in other metazoa and in yeast

Sequences without Ensembl nomenclature are found in Supplementary Information.

©2008 Nature Publishing Group

rare in platypus, but there are thousands of copies of an ancient gypsy-class LTR element (all LTR elements previously identified in mammals, birds, or reptiles belong to the retrovirus clade). Overall, the frequency of interspersed repeats (over 2 repeats per kb) is higher than in any previously characterized metazoan genome. Population analysis using LINE2/Mon-1 elements distinguished the Tasmanian population from three other mainland clusters (Supplementary Fig. 4a, b), in good agreement with tree-based analysis, physical proximity and previous knowledge of platypus population relationships[57].

Cluster analysis of all LINE2 copies revealed a phylogenetic relationship lacking branches, as if a single-locus, fast-evolving gene has steadily spread an exceptional number of pseudogenes over time (Supplementary Fig. 6). This 'master gene' appearance is, to a lesser degree, also observed for LINE1 in eutherians[58], but not to the same extent for MIR/Mon-1 or other retrotransposons in mammals. The phylogeny of LINE2 and Mon-1 was also supported by a genome-wide transposition-in-transposition (TinT) analysis[59] (Supplementary Tables 7 and 8). LINE2 density is similar on all chromosomes (Supplementary Fig. 7); it does not correlate with chromosome length (and recombination rate) as the CR1 LINE density does in the chicken genome[19], nor is it higher on sex chromosomes than on autosomes, as LINE1 density is in eutherians (which has led to postulations on a function in dosage compensation)[60].

We compared microsatellites in the platypus genome with those of representative vertebrates (Supplementary Notes 22). The mean microsatellite coverage of platypus genomic sequences assembled into chromosomes is 2.67 ± 0.34%; significantly lower than all other mammalian genomes sequenced so far and most similar to that observed in chicken (Supplementary Fig. 8). Microsatellites are on average shorter in platypus than in other genomes (Supplementary Table 9), but microsatellite coverage surpasses chicken owing to very long tri- and tetranucleotide repeats (Supplementary Fig. 9). The platypus has a higher proportion of microsatellites with high A+T content, in comparison to the other vertebrates examined, an abundance distribution that has more in common with reptiles than with mammals (Supplementary Fig. 10).

Genomic imprinting. Genomic imprinting is an epigenetic phenomenon that results in monoallelic gene expression. In the vertebrates, imprinting seems to have evolved recently and has only been confirmed in marsupials and eutherian mammals[61,62]. The autosomal localization of some imprinted orthologues in platypus is known[63]. However, we examined the conservation of synteny and the distribution of retrotransposed elements in all orthologous eutherian-imprinted clustered and non-clustered genes in the platypus genome. A representative cluster is shown in Fig. 5 (see also Supplementary Fig. 12).

Clusters that became imprinted in therians (with the exception of the Prader–Willi–Angelman locus[64]) have not been assembled recently and reside in ancient syntenic mammalian groups, although some regions have expanded by mechanisms such as gene duplication or transposition. There were significantly fewer LTR and DNA elements across all platypus orthologous regions relative to eutherian imprinted genes ($P < 0.04$ and 0.04, respectively), whereas there was a significant increase in the sequences masked by SINEs ($P < 0.03$). The chicken had fewer total repeats and no SINEs or sRNAs. Comparison of all regions in the platypus with the orthologous regions in opossum, mouse, dog and human demonstrates that accumulation of LTR, DNA elements, and simple and low complexity repeats coincides with, and may be a driving force in, the acquisition of imprinting in these regions in therian mammals.

The CpG fraction. The eutherian and chicken genomes generally average around 41% G+C content, although many intervals differ substantially from the average, particularly in humans (Supplementary Notes 23). In contrast, the platypus genome averages 45.5% G+C content and rarely deviates far from the average. The opossum genome averages only 38% G+C content and also has a narrow distribution (Supplementary Fig. 13). The source of the elevated G+C fraction in platypus remains unclear. It is explained only in part by monotreme interspersed repeat elements, as platypus DNA outside of known interspersed repeats is 44.7% G+C. Furthermore, tandem repeats of short DNA motifs (microsatellites) in platypus show an A+T bias, as with other mammals. Recombination-driven biased gene conversion may be a factor, in agreement with what has been shown for eutherians[65] and marsupials[66]. This is suggested by the observation that the six platypus chromosomes where the currently mapped DNA sequence averages over 45% G+C content (that is, 17, 20, 15, 14, 10 and 11 in order of decreasing G+C fraction) are among the 10 shortest (Supplementary Fig. 14), because short chromosomes have a higher recombination rate[67]. However, a direct test

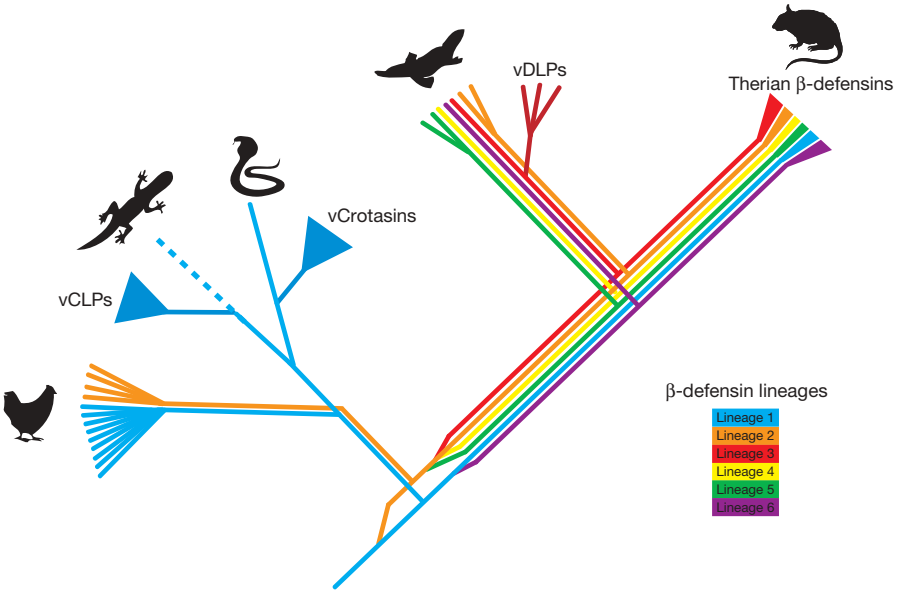


Figure 4 | The evolution of β-defensin peptides in platypus venom gland. The diagram illustrates separate gene duplications in different parts of the phylogeny for platypus venom defensin-like peptides (vDLPs), for lizard venom crotamine-like peptides (vCLPs) and for snake venom crotamines. These venom proteins have thus been co-opted from pre-existing non-toxin homologues independently in platypus and in lizards and snakes[48].

©2008 Nature Publishing Group

is currently lacking because platypus recombination rates have not been measured. A further examination of the CpG fraction, that associated with promoter elements, is found in Supplementary Notes 24 and Supplementary Fig. 15.

Conclusions

The egg-laying platypus is a remarkable species with many biological features unique among mammals. Our sequencing of the platypus genome now enables us to compare its sequence characteristics and organization with those of birds and therian mammals in order to address the questions of platypus biology and to date the emergence of mammalian traits. We report here that sequence characteristics of the platypus genome show features of reptiles as well as mammals.

Platypus contains a largely standard repertoire of non-protein-coding, ncRNAs, except for the snoRNAs, which exhibit a marked expansion associated with at least one retrotransposed subfamily. Some of these retrotransposed snoRNAs are expressed and thus may have functional roles. The platypus has fully elaborated piRNA and miRNA pathways, the latter including many monotreme-specific miRNAs and miRNAs that are shared with either mammals or chickens. Many functional assessments of these novel miRNAs remain to be carried out and will surely add to our knowledge of mammalian miRNA evolution.

The 18,527 protein-coding genes predicted from the platypus assembly fall within the range for therian genomes. Of particular interest are families of genes involved in biology that links monotremes to reptiles, such as egg-laying, vision and envenomation, as well as mammal-specific characters such as lactation, characters shared with marsupials such as antibacterial proteins, and platypus-specific characters such as venom delivery and underwater foraging. For instance, anatomical adaptations for chemoreception during underwater foraging are reflected in an unusually large repertoire of vomeronasal type 1 receptor genes. However, the repertoire of milk protein genes is typically mammalian, and the arrangement of milk protein genes seems to have been preserved since the last common ancestor of monotremes and therian mammals.

Since its initial description, the platypus has stood out as a species with a blend of reptilian and mammalian features, which is a characteristic that penetrates to the level of the genome sequence. The density and distribution of repetitive sequence, for example, reflects this fact. The high frequency of interspersed repeats in the platypus genome, although typical for mammalian genomes, is in contrast with the observed mean microsatellite coverage, which appears more reptilian. Additionally, the correlation of parent-of-origin-specific expression patterns in regions of reduced interspersed repeats in the platypus suggests that the evolution of imprinting in therians is linked to the accumulation of repetitive elements.

We find that the mixture of reptilian, mammalian and unique characteristics of the platypus genome provides many clues to the function and evolution of all mammalian genomes. The wealth of new findings and confirmation of existing knowledge immediately evident from the release of these data promise that the availability of

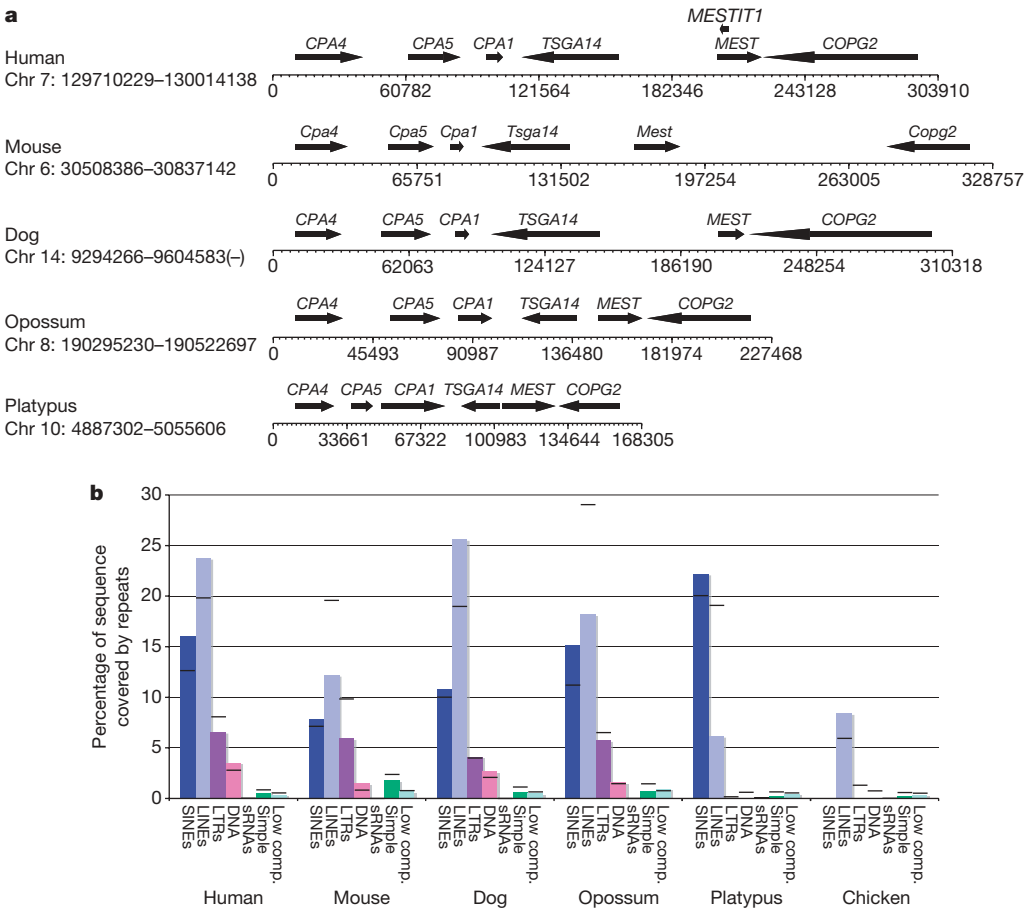


Figure 5 | Comparative mammalian analysis for a representative eutherian imprinted gene cluster (*PEG1/MEST*). **a**, The gene arrangement is conserved between mammals. However, non-coding regions are expanded in therians. Arrows indicate genes and the direction of transcription; the scale shows base pairs. **b**, Summary of repeat distribution for the *PEG1/MEST* cluster. Histograms represent the sequence (%) masked by each repeat element within the *MEST* cluster; black bars represent repeat distribution across the entire genome. With the exception of SINEs, platypus has fewer repeats of LINEs, LTRs, DNA and simple repeats (Simple) than eutherian mammals. Low comp., low complexity; sRNAs, small RNAs.

©2008 Nature Publishing Group

the platypus genome sequence will provide the critically needed background to inspire rapid advances in other investigations of mammalian biology and evolution.

METHODS SUMMARY

Tissue resources. Tissue was obtained from animals captured at the Upper Barnard River, New South Wales, Australia, during breeding season (AEEC permit number R.CG.07.03 to F. Grützner; Environment ACT permit number LI 2002 270 to J. A. M. Graves; NPWS permit number A193 to R. C. Jones; AEC permit number S-49-2006 to F. Grützner).

Sequence assembly. A total of 26.9 million reads was assembled using the PCAP software[20]. Attempts were made to assign the largest contiguous blocks of sequence to chromosomes using standard FISH techniques.

Non-coding RNAs. We used the established Rfam pipeline[68] and *de novo* sequencing to detect non-protein-coding RNAs (ncRNAs). Cloning, sequencing and annotation of sRNAs from platypus, echidna and chicken as well as miRNA sequences are described in ref. 25.

Genes. Protein-coding and non-protein-coding genes were computed using a modified version of the Ensembl pipeline (Supplementary Notes 14). Gene orthology assignment followed a procedure implemented previously[69]. Orthology rate estimation was performed with PAML[70] using the model of ref. 71. In all cases, codon frequencies were estimated from the nucleotide composition at each codon position (F3X4 model).

Genome landscape. Pairwise alignments between human and dog, mouse, opossum, platypus and chicken were projected from whole-genome alignments of 28 species (http://genome.cse.ucsc.edu/). These alignments were the basis for phylogeny, chromosome synteny, interspersed repeats, imprinting and CpG fraction analyses.

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

Received 14 September 2007; accepted 25 March 2008.

1. Home, E. A. A description of the anatomy of *Ornithorynchus paradoxus*. *Phil. Trans. R. Soc. Lond. B* **92,** 67–84 (1802).
2. Bininda-Emonds, O. R. P. *et al.* The delayed rise of present-day mammals. *Nature* **446,** 507–512 (2007).
3. Caldwell, H. The embryology of Monotremata and Marsupialia. *Phil. Trans. R. Soc. Lond. B* **178,** 463–486 (1887).
4. Griffiths, M. *Echidnas* (Pergamon, Oxford, 1968).
5. Griffiths, M. *The Biology of the Monotremes* (Academic, New York, 1978).
6. Renfree, M. B. Monotreme and marsupial reproduction. *Reprod. Fertil. Dev.* **7,** 1003–1020 (1995).
7. Renfree, M. B. *Mammal Phylogeny* (Springer, New York, 1993).
8. Watson, J. M., Meyne, J. & Graves, J. A. M. Ordered tandem arrangement of chromosomes in the sperm heads of monotreme mammals. *Proc. Natl Acad. Sci. USA* **93,** 10200–10205 (1996).
9. Greaves, I. K., Rens, W., Ferguson-Smith, M. A., Griffin, D. & Graves, J. A. M. Conservation of chromosome arrangement and position of the X in mammalian sperm suggests functional significance. *Chromosome Res.* **11,** 503–512 (2003).
10. Temple-Smith, P. & Grant, T. Uncertain breeding: a short history of reproduction in monotremes. *Reprod. Fertil. Dev.* **13,** 487–497 (2001).
11. Temple-Smith, P. D. *Seasonal Breeding Biology of the Platypus,* Ornithorynchus anatinus *(Shaw 199), with Special Reference to Male*. PhD thesis, Australian National University (1973).
12. Scheich, H., Langner, G., Tidemann, C., Coles, R. B. & Guppy, A. Electroreception and electrolocation in platypus. *Nature* **319,** 401–402 (1986).
13. Pettigrew, J. D. Electroreception in monotremes. *J. Exp. Biol.* **202,** 1447–1454 (1999).
14. Bick, Y. A. E. & Sharman, G. B. The chromosomes of the platypus (*Ornithorynchus*: Monotremata). *Cytobios* **14,** 17–28 (1975).
15. Wrigley, J. M. & Graves, J. A. M. Two monotreme cell lines, derived from female platypuses (*Ornithorhynchus anatinus*; Monotremata, Mammalia). *In Vitro* **20,** 321–328 (1984).
16. El-Mogharbel, N. *et al. DMRT* gene cluster analysis in the platypus: new insights into genomic organization and regulatory regions. *Genomics* **89,** 10–21 (2007).
17. Rens, W. *et al.* Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution. *Proc. Natl Acad. Sci. USA* **101,** 16257–16261 (2004).
18. Grutzner, F. *et al.* In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. *Nature* **432,** 913–917 (2004).
19. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432,** 695–716 (2004).
20. Huang, X. *et al.* Application of a superword array in genome assembly. *Nucleic Acids Res.* **34,** 201–205 (2006).
21. McMillan, D. *et al.* Characterizing the chromosomes of the platypus (*Ornithorhynchus anatinus*). *Chromosome Res.* **15,** 961–974 (2007).
22. Maxwell, E. S. & Fournier, M. J. The small nucleolar RNAs. *Annu. Rev. Biochem.* **64,** 897–934 (1995).
23. Schmitz, J. *et al.* Retroposed SNOfall—A mammalian-wide comparison of platypus snoRNAs. *Genome Res.* doi:10.1101/gr.7177908 (in the press).
24. Zemann, A., op de Bekke, A., Kiefmann, M., Brosius, J. & Schmitz, J. Evolution of small nucleolar RNAs in nematodes. *Nucleic Acids Res.* **34,** 2676–2685 (2006).
25. Murchison, E. P. *et al.* Conservation of small RNA pathways in platypus. *Genome Res.* doi:10.1101/gr.73056.107 (in the press).
26. Aravin, A. A., Hannon, G. J. & Brennecke, J. The Piwi/piRNA pathway provides adaptive defense for the transposon arms race. *Science* **318,** 761–764 (2007).
27. Curwen, V. *et al.* The Ensembl automatic gene annotation system. *Genome Res.* **14,** 942–950 (2004).
28. Goodstadt, L. & Ponting, C. P. Phylogenetic reconstruction of orthology, paralogy, and conserved synteny for dog and human. *PLoS Comput. Biol.* **2,** e133 (2006).
29. Goodstadt, L., Heger, A., Webber, C. & Ponting, C. P. An analysis of the gene complement of a marsupial, *Monodelphis domestica*: evolution of lineage-specific genes and giant chromosomes. *Genome Res.* **17,** 969–981 (2007).
30. Emes, R. D., Goodstadt, L., Winter, E. E. & Ponting, C. P. Comparison of the genomes of human and mouse lays the foundation of genome zoology. *Hum. Mol. Genet.* **12,** 701–709 (2003).
31. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246,** 96–98 (1973).
32. Kirsch, J. A. & Mayer, G. C. The platypus is not a rodent: DNA hybridization, amniote phylogeny and the palimpsest theory. *Phil. Trans. R. Soc. Lond. B* **353,** 1221–1237 (1998).
33. Rowe, T., Rich, T. H., Vickers-Rich, P., Springer, M. & Woodburne, M. O. The oldest platypus and its bearing on divergence timing on platypus and echidna clades. *Proc. Natl Acad. Sci. USA* **105,** 1238–1242 (2008).
34. Grus, W. E., Shi, P. & Zhang, J. Largest vertebrate vomeronasal type 1 receptor (V1R) gene repertoire in the semi-aquatic platypus. *Mol. Biol. Evol.* **24,** 2153–2157 (2007).
35. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).
36. Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428,** 493–521 (2004).
37. Aloni, R., Olender, T. & Lancet, D. Ancient genomic architecture for mammalian olfactory clusters. *Genome Biol.* **7,** R88 (2006).
38. Jovine, L., Qi, H., Williams, Z., Litscher, E. S. & Wassarman, P. M. Features that affect secretion and assembly of zona pellucida glycoproteins during mammalian oogenesis. *Soc. Reprod. Fertil.* **63** (suppl.), 187–201 (2007).
39. Riggio, M., Scudiero, R., Filosa, S. & Parisi, E. Sex- and tissue-specific expression of aspartic proteinases in *Danio rerio* (zebrafish). *Gene* **260,** 67–75 (2000).
40. Nixon, B., Aitken, R. J. & McLaughlin, E. A. New insights into the molecular mechanisms of sperm-egg interaction. *Cell. Mol. Life Sci.* **64,** 1805–1823 (2007).
41. Morel, L., Dufaure, J. P. & Depeiges, A. The lipocalin sperm coating lizard epididymal secretory protein family: mRNA structural analysis and sequential expression during the annual cycle of the lizard, *Lacerta vivipara*. *J. Mol. Endocrinol.* **24,** 127–133 (2000).
42. Schlomann, U. *et al.* The metalloprotease disintegrin ADAM8. Processing by autocatalysis is required for proteolytic activity and cell adhesion. *J. Biol. Chem.* **277,** 48210–48219 (2002).
43. Oftedal, O. The origin of lactation as a water source for parchment-shelled eggs. *J. Mammary Gland Biol. Neoplasia* **7,** 253–266 (2002).
44. Kawasaki, K. & Weiss, K. M. Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. *Proc. Natl Acad. Sci. USA* **100,** 4060–4065 (2003).
45. Lester, K. S., Boyde, A., Gilkeson, C. & Archer, M. Marsupial and monotreme enamel structure. *Scanning Microsc.* **1,** 401–420 (1987).
46. de Plater, G., Martin, R. L. & Milburn, P. J. A pharmacological and biochemical investigation of the venom from the platypus (*Ornithorhynchus anatinus*). *Toxicon* **33,** 157–169 (1995).
47. Fry, B. G. From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res.* **15,** 403–420 (2005).
48. Whittington, C. *et al.* Defensins and the convergent evolution of platypus and reptile venom genes. *Genome Res.* doi:10.1101/gr.7149808 (in the press).
49. Diener, E. & Ealey, E. H. Immune system in a monotreme: studies on the Australian echidna (*Tachyglossus aculeatus*). *Nature* **208,** 950–953 (1965).
50. Kelley, J., Walter, L. & Trowsdale, J. Comparative genomics of natural killer cell receptor gene clusters. *PLoS Genet.* **1,** 129–139 (2005).
51. Belov, K. *et al.* Characterization of the opossum immune genome provides insights into the evolution of the mammalian immune system. *Genome Res.* **17,** 982–991 (2007).
52. Durr, U. H., Sudheendra, U. S. & Ramamoorthy, A. LL-37, the only human member of the cathelicidin family of antimicrobial peptides. *Biochim. Biophys. Acta* **1758,** 1408–1425 (2006).
53. Bals, R. & Wilson, J. M. Cathelicidins—a family of multifunctional antimicrobial peptides. *Cell. Mol. Life Sci.* **60,** 711–720 (2003).

©2008 Nature Publishing Group

54. Zanetti, M. Cathelicidins, multifunctional peptides of the innate immunity. *J. Leukoc. Biol.* **75,** 39–48 (2004).
55. Veyrunes, F. *et al.* Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res.* doi:10.1101/gr.7101908 (in the press).
56. Itoh, Y. *et al.* Dosage compensation is less effective in birds than in mammals. *J. Biol.* **6,** 2 (2007).
57. Gemmell, N. J. *et al.* Determining platypus relationships. *Aust. J. Zool.* **43,** 283–291 (1995).
58. Smit, A. F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9,** 657–663 (1999).
59. Kriegs, J. O. *et al.* Waves of genomic hitchhikers shed light on the evolution of gamebirds (Aves: Galliformes). *BMC Evol. Biol.* **7,** 190 (2007).
60. Ross, M. T. *et al.* The DNA sequence of the human X chromosome. *Nature* **434,** 325–337 (2005).
61. Killian, J. K. *et al.* M6P/IGF2R imprinting evolution in mammals. *Mol. Cell* **5,** 707–716 (2000).
62. Suzuki, S. *et al.* Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. *PLoS Genet.* **3,** e55 (2007).
63. Edwards, C. A. *et al.* The evolution of imprinting: chromosomal mapping of orthologues of mammalian imprinted domains in monotreme and marsupial mammals. *BMC Evol. Biol.* **7,** 157 (2007).
64. Rapkins, R. W. *et al.* The Prader-Willi/Angelman imprinted domain was assembled recently from non-imprinted components. *PLoS Genet.* **2,** 1–10 (2006).
65. Meunier, J. & Duret, L. Recombination drives the evolution of GC-content in the human genome. *Mol. Biol. Evol.* **21,** 984–990 (2004).
66. Mikkelsen, T. S. *et al.* Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. *Nature* **447,** 167–177 (2007).
67. Coop, G. & Przeworski, M. An evolutionary view of human recombination. *Nature Rev. Genet.* **8,** 23–34 (2007).
68. Griffiths-Jones, S. *et al.* Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33,** D121–D124 (2005).
69. Heger, A. & Ponting, C. P. Evolutionary rate analyses of orthologues and paralogues from twelve *Drosophila* genomes. *Genome Res.* **17,** 1837–1849 (2007).
70. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13,** 555–556 (1997).
71. Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11,** 725–736 (1994).

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

Acknowledgements The sequencing of platypus was funded by the National Human Genome Research Institute (NHGRI). This research was supported by grant HG002238 from the NHGRI (W.M.), NGFN (0313358A; to J.S. and J.B.), the DFG (SCHM 1469; to J.S. and J.B.), National Science Foundation BCS-0218338 (M.A.B.) and EPS-0346411 (M.A.B.), National Institutes of Health RO1 GM59290 (M.A.B.), National Institutes of Health RO1HG02385 (E.E.E), Australian Research Council (F.G.), UK Medical Research Council (C.P.P. and A.H.), Ministry of Science-Spain (X.S.P. and C.L.-O.) and the State of Louisiana Board of Regents Support Fund (M.A.B.). We thank T. Grant, S. Akiyama, P. Temple-Smith, R. Whittington and the Queensland Museum for platypus sample collection and DNA, and Macquarie Generation and Glenrock station for providing access and facilities during sampling. Approval to collect animals was granted by the New South Wales National Parks and Wildlife Services, New South Wales. Funding support for some platypus samples was provided by Australian Research Council and W.V. Scott Foundation. We thank M. Shelton, I. Elton and the Healesville Sanctuary for platypus pictures. We thank L. Duret for assistance on genome landscape analysis; G. Shaw for use of the silhouettes on Fig. 1; and Z.-X. Luo, M. Archer and R. Beck for advice on the Fig. 1 phylogeny. We acknowledge the approved use of the green anole lizard sequence data provided by the Broad Institute. Resources for exploring the sequence and annotation data are available on browser displays available at UCSC (http://genome.ucsc.edu), Ensembl (http://www.ensembl.org) and the NCBI (http://www.ncbi.nlm.nih.gov).

Author Information The *Ornithorhynchus anatinus* whole-genome shotgun project has been deposited in DDBJ/EMBL/GenBank under the project accession AAPN00000000. The version described in this paper is the first version, AAPN01000000. The SNPs have been deposited in the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/) with Submitter Method IDs PLATYPUS-ASSEMBLY_SNPS_200801 and PLATYPUS-READS_SNPS_200801.

Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. Correspondence and requests for materials should be addressed to W.C.W. (wwarren@wustl.edu) or R.K.W. (rwilson@wustl.edu).

Wesley C. Warren[1], LaDeana W. Hillier[1], Jennifer A. Marshall Graves[2], Ewan Birney[3], Chris P. Ponting[4], Frank Grützner[5], Katherine Belov[6], Webb Miller[7], Laura Clarke[8], Asif T. Chinwalla[1], Shiaw-Pyng Yang[1], Andreas Heger[4], Devin P. Locke[1], Pat Miethke[2], Paul D. Waters[2], Frédéric Veyrunes[2,9], Lucinda Fulton[1], Bob Fulton[1], Tina Graves[1], John Wallis[1], Xose S. Puente[10], Carlos López-Otín[10], Gonzalo R. Ordóñez[10], Evan E. Eichler[11], Lin Chen[11], Ze Cheng[11], Janine E. Deakin[2], Amber Alsop[2], Katherine Thompson[2], Patrick Kirby[2], Anthony T. Papenfuss[12], Matthew J. Wakefield[12], Tsviya Olender[13], Doron Lancet[13], Gavin A. Huttley[14], Arian F. A. Smit[15], Andrew Pask[16], Peter Temple-Smith[16,17], Mark A. Batzer[18], Jerilyn A. Walker[18], Miriam K. Konkel[18], Robert S. Harris[7], Camilla M. Whittington[6], Emily S. W. Wong[6], Neil J. Gemmell[19], Emmanuel Buschiazzo[19], Iris M. Vargas Jentzsch[19], Angelika Merkel[19], Juergen Schmitz[20], Anja Zemann[20], Gennady Churakov[20], Jan Ole Kriegs[20], Juergen Brosius[20], Elizabeth P. Murchison[21], Ravi Sachidanandam[21], Carly Smith[21], Gregory J. Hannon[21], Enkhjargal Tsend-Ayush[5], Daniel McMillan[2], Rosalind Attenborough[2], Willem Rens[9], Malcolm Ferguson-Smith[9], Christophe M. Lefèvre[22,23], Julie A. Sharp[23], Kevin R. Nicholas[23], David A. Ray[24], Michael Kube[25], Richard Reinhardt[25], Thomas H. Pringle[26], James Taylor[27], Russell C. Jones[28], Brett Nixon[28], Jean-Louis Dacheux[29], Hitoshi Niwa[30], Yoko Sekita[30], Xiaoqiu Huang[31], Alexander Stark[32], Pouya Kheradpour[32], Manolis Kellis[32], Paul Flicek[3], Yuan Chen[3], Caleb Webber[4], Ross Hardison[7], Joanne Nelson[1], Kym Hallsworth-Pepin[1], Kim Delehaunty[1], Chris Markovic[1], Pat Minx[1], Yucheng Feng[1], Colin Kremitzki[1], Makedonka Mitreva[1], Jarret Glasscock[1], Todd Wylie[1], Patricia Wohldmann[1], Prathapan Thiru[1], Michael N. Nhan[1], Craig S. Pohl[1], Scott M. Smith[1], Shunfeng Hou[1], Marilyn B. Renfree[16], Elaine R. Mardis[1] & Richard K. Wilson[1]

[1]Genome Sequencing Center, Washington University School of Medicine, Campus Box 8501, 4444 Forest Park Avenue, St Louis, Missouri 63108, USA. [2]Australian National University, Canberra, Australian Capital Territory 0200, Australia. [3]EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [4]MRC Functional Genetics Unit, University of Oxford, Department of Human Physiology, Anatomy and Genetics, South Parks Road, Oxford OX1 3QX, UK. [5]Discipline of Genetics, School of Molecular & Biomedical Science, The University of Adelaide, 5005 South Australia, Australia. [6]Faculty of Veterinary Science, The University of Sydney, Sydney, New South Wales 2006, Australia. [7]Center for Comparative Genomics and Bioinformatics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA. [8]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [9]Cambridge University, Department of Veterinary Medicine, Madingley Road, Cambridge CB3 0ES, UK. [10]Instituto Universitario de Oncologia, Departamento de Bioquimica y Biologia Molecular, Universidad de Oviedo, 33006-Oviedo, Spain. [11]Department of Genome Sciences, Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. [12]The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3050, Australia. [13]Crown Human Genome Center, Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. [14]John Curtin School of Medical Research, The Australian National University, Canberra, Australian Capital Territory 0200, Australia. [15]Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103-8904, USA. [16]Department of Zoology, The University of Melbourne, Victoria 3010, Australia. [17]Monash Institute of Medical Research, 27–31 Wright Street, Clayton, Victoria 3168, Australia. [18]Department of Biological Sciences, Center for Bio-Modular Multi-Scale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, Louisiana 70803, USA. [19]School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand. [20]Institute of Experimental Pathology, University of Muenster, Von-Esmarch-Strasse 56, D-48149 Muenster, Germany. [21]Cold Spring Harbor Laboratory, Howard Hughes Medical Institute, Cold Spring Harbor, New York 11724, USA. [22]Victorian Bioinformatics Consortium, Monash University, Clayton, Victoria 3080, Australia. [23]CRC for Innovative Dairy Products, Department of Zoology, University of Melbourne, Victoria 3010, Australia. [24]Department of Biology, West Virginia University, Morgantown, West Virginia 26505, USA. [25]MPI Molecular Genetics, D-14195 Berlin-Dahlem, Ihnestrasse 73, Germany. [26]Sperling Foundation, Eugene, Oregon 97405, USA. [27]Courant Institute, New York University, New York, New York 10012, USA. [28]Discipline of Biological Sciences, School of Environmental and Life Sciences, University of Newcastle, New South Wales 2308, Australia. [29]UMR INRA-CNRS 6073, Physiologie de la Reproduction et des Comportements, Nouzilly 37380, France. [30]Laboratory for Pluripotent Cell Studies, RIKEN Center for Developmental Biology (CDB), 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe 6500047, Japan. [31]Department of Computer Science, Iowa State University, 226 Atanasoff Hall, Ames, Iowa 50011, USA. [32]The Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02139, USA.

©2008 Nature Publishing Group

METHODS

Sequence assembly. A total of 26.9 million reads was assembled using the PCAP software[20]. Assembly quality assessment accounted for read depth, chimaeric reads, repeat content, cloning bias, G+C content and heterozygosity (Supplementary Notes 4–11). We identified a total of ~1.2 million single nucleotide polymorphisms (SNPs) within the 1.84-Gb sequenced female platypus genome using two independent analyses, SSAHA2 (SSAHA: a fast search method for large DNA databases[72]) and PCAP output[20] (Supplementary Notes 11).

Non-coding RNAs. snoRNA annotation is as described in ref. 23. miRNAs sharing a heptamer at nucleotide position 2–8 were defined as a family. Homology with mouse/human miRNAs was based on annotated miRNAs in Rfam (http://microrna.sanger.ac.uk/sequences/index.shtml). piRNA sequences have been submitted to GEO (http://www.ncbi.nlm.nih.gov/geo/). miRNA total cloning frequency was normalized across tissue libraries by scaling cloning frequency per library by a factor representing total number of miRNA reads per library.

Genes. Orthologue groups were selected based on whether they contained genes predicted only from the platypus, and not from the chicken, opossum, dog, mouse or human genome assemblies (Supplementary Notes 15–17). Other groups were selected where the number of in-paralogous platypus genes exceeded the numbers of the other (chicken, opossum, dog, mouse and human) terminal lineages. Some of these groups represent erroneous gene predictions where, for example, protein-coding sequence predictions represented instead transposed element or highly repetitive sequence, or overlapped, on the reverse strand, other well-established coding sequence. Such instances were discarded. Lineage-specific gene loss was detected by inspection of BLASTZ alignment chains and nets at the UCSC Genome Browser (http://genome.cse.ucsc.edu/); by the interrogation of all known cDNA, EST and protein sequences held in GenBank using BLAST; and by attempting to predict orthologous genes within genomic intervals flanked by syntenic anchors.

Genome landscape. To establish phylogeny we extended the basic data sampling approach described previously[73] to protein-coding genes, and used established techniques to analyse protein-coding indels[74] and retrotransposon insertions[75] (Supplementary Notes 19).

The population structure of 90 platypuses from different regions in Australia was determined using Structure software v2.1 (ref. 76) using genotypes of 57 polymorphic Mon-1 and LINE2 loci. Five thousand replications were examined (Supplementary Notes 21).

Microsatellites were identified across the platypus genome (ornAna1) combining two programs: Tandem Repeat Finder (TRF)[77] and Sputnik[78] (Supplementary Notes 22).

For the imprinting cluster of *PEG1/MEST*, comparative maps were complied from Vega annotations for the mouse and human, and Ensembl gene builds for other species. Multiple alignments of each region for repeat distribution analyses were constructed using MLAGAN[79] with translated anchoring.

We examined genomic assemblies for human (hg18), mouse (musMus8), dog (canFam2), opossum (monDom4), platypus (ornAna1) and chicken (galGal3), downloaded from the UCSC Genome Browser (http://genome.ucsc.edu), and computed the fraction of G+C nucleotides in each non-overlapping 10,000-bp window free of ambiguous bases. Bases in repeats were not distinguished and were counted along with non-repeat bases. For platypus all assembled sequence was analysed; for the other species only bases assigned to chromosomes were used.

72. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11,** 1725–1729 (2001).
73. Huttley, G. A., Wakefield, M. J. & Easteal, S. Rates of genome evolution and branching order from whole genome analysis. *Mol. Biol. Evol.* **24,** 1722–1730 (2007).
74. Murphy, W. J., Pringle, T. H., Crider, T. A., Springer, M. S. & Miller, W. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res.* **17,** 413–421 (2007).
75. Kriegs, J. O. *et al.* Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol.* **4,** e91 (2006).
76. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155,** 945–959 (2000).
77. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27,** 573–580 (1999).
78. La Rota, M., Kantety, R. V., Yu, J. K. & Sorrells, M. E. Nonrandom distribution and frequencies of genomic and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* **6,** 23–29 (2005).
79. Brudno, M. *et al.* LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13,** 721–731 (2003).

©2008 Nature Publishing Group

CORRIGENDUM

doi:10.1038/nature07253

Genome analysis of the platypus reveals unique signatures of evolution

Wesley C. Warren, LaDeana W. Hillier, Jennifer A. Marshall Graves, Ewan Birney, Chris P. Ponting, Frank Grützner, Katherine Belov, Webb Miller, Laura Clarke, Asif T. Chinwalla, Shiaw-Pyng Yang, Andreas Heger, Devin P. Locke, Pat Miethke, Paul D. Waters, Frédéric Veyrunes, Lucinda Fulton, Bob Fulton, Tina Graves, John Wallis, Xose S. Puente, Carlos López-Otín, Gonzalo R. Ordóñez, Evan E. Eichler, Lin Chen, Ze Cheng, Janine E. Deakin, Amber Alsop, Katherine Thompson, Patrick Kirby, Anthony T. Papenfuss, Matthew J. Wakefield, Tsviya Olender, Doron Lancet, Gavin A. Huttley, Arian F. A. Smit, Andrew Pask, Peter Temple-Smith, Mark A. Batzer, Jerilyn A. Walker, Miriam K. Konkel, Robert S. Harris, Camilla M. Whittington, Emily S. W. Wong, Neil J. Gemmell, Emmanuel Buschiazzo, Iris M. Vargas Jentzsch, Angelika Merkel, Juergen Schmitz, Anja Zemann, Gennady Churakov, Jan Ole Kriegs, Juergen Brosius, Elizabeth P. Murchison, Ravi Sachidanandam, Carly Smith, Gregory J. Hannon, Enkhjargal Tsend-Ayush, Daniel McMillan, Rosalind Attenborough, Willem Rens, Malcolm Ferguson-Smith, Christophe M. Lefèvre, Julie A. Sharp, Kevin R. Nicholas, David A. Ray, Michael Kube, Richard Reinhardt, Thomas H. Pringle, James Taylor, Russell C. Jones, Brett Nixon, Jean-Louis Dacheux, Hitoshi Niwa, Yoko Sekita, Xiaoqiu Huang, Alexander Stark, Pouya Kheradpour, Manolis Kellis, Paul Flicek, Yuan Chen, Caleb Webber, Ross Hardison, Joanne Nelson, Kym Hallsworth-Pepin, Kim Delehaunty, Chris Markovic, Pat Minx, Yucheng Feng, Colin Kremitzki, Makedonka Mitreva, Jarret Glasscock, Todd Wylie, Patricia Wohldmann, Prathapan Thiru, Michael N. Nhan, Craig S. Pohl, Scott M. Smith, Shunfeng Hou, Mikhail Nefedov[1], Pieter J. de Jong[1], Marilyn B. Renfree, Elaine R. Mardis & Richard K. Wilson

[1]Children's Hospital Oakland Research Institute, Bruce Lyon Research Building, 747 52nd Street, Oakland, California 94609, USA.

Nature 453, 175–183 (2008)

In this Article, Mikhail Nefedov and Pieter J. de Jong were omitted from the author list.

ERRATUM

doi:10.1038/nature07316

Tumour invasion and metastasis initiated by microRNA-10b in breast cancer

Li Ma, Julie Teruya-Feldstein & Robert A. Weinberg

Nature 449, 682–688 (2007)

In Fig. 4e of this Article, the two E-box sequences were inadvertently exchanged. E-box 1, which is near the stem-loop (at −313 bp), should be CACTTG instead of CACCTG, and E-box 2 (at −2,422 bp), which is distal to the stem-loop, should be CACCTG instead of CACTTG.

©2008 Macmillan Publishers Limited. All rights reserved

ARTICLES

Dynamic binding orientations direct activity of HIV reverse transcriptase

Elio A. Abbondanzieri[1]*, Gregory Bokinsky[1]*, Jason W. Rausch[4], Jennifer X. Zhang[1], Stuart F. J. Le Grice[4] & Xiaowei Zhuang[1,2,3]

The reverse transcriptase of human immunodeficiency virus (HIV) catalyses a series of reactions to convert the single-stranded RNA genome of HIV into double-stranded DNA for host-cell integration. This task requires the reverse transcriptase to discriminate a variety of nucleic-acid substrates such that active sites of the enzyme are correctly positioned to support one of three catalytic functions: RNA-directed DNA synthesis, DNA-directed DNA synthesis and DNA-directed RNA hydrolysis. However, the mechanism by which substrates regulate reverse transcriptase activities remains unclear. Here we report distinct orientational dynamics of reverse transcriptase observed on different substrates with a single-molecule assay. The enzyme adopted opposite binding orientations on duplexes containing DNA or RNA primers, directing its DNA synthesis or RNA hydrolysis activity, respectively. On duplexes containing the unique polypurine RNA primers for plus-strand DNA synthesis, the enzyme can rapidly switch between the two orientations. The switching kinetics were regulated by cognate nucleotides and non-nucleoside reverse transcriptase inhibitors, a major class of anti-HIV drugs. These results indicate that the activities of reverse transcriptase are determined by its binding orientation on substrates.

Virtually all RNA-processing and DNA-processing enzymes show selectivity for backbone compositions or base sequences of their nucleic-acid substrates. This substrate selectivity is especially crucial for the HIV-1 reverse transcriptase (RT), which binds and discriminates between a variety of nucleic-acid duplexes for distinct catalytic functions[1,2]. RT is a heterodimer consisting of a p51 and a p66 subunit, the latter of which contains catalytically active DNA polymerase and RNase H domains[3,4], catalysing a complex, multi-step reaction to convert the single-stranded RNA genome into double-stranded DNA[1,2]. First, RT uses the viral RNA genome as a template and a host-cell transfer RNA as a primer to synthesize a minus-strand DNA, producing an RNA–DNA hybrid[5–7]. This duplex becomes the substrate of the RNase H domain of RT, which cleaves the RNA strand at numerous points, leaving behind short RNA segments hybridized to the nascent DNA[8–10]. Among these RNAs, two specific purine-rich sequences, known as the polypurine tracts (PPTs), serve as unique primers to initiate the synthesis of plus-strand DNA[11–13], thereby creating the double-stranded DNA viral genome. Specific cleavage by RNase H then removes the PPT primers and exposes the integration sequence to facilitate the insertion of the viral DNA into the host chromosome[14]. Inappropriate initiation of synthesis of the plus-strand DNA at other RNA segments prevents integration[2,15]. RT must therefore obey the following primer-selection rules: first, DNA primers readily engage the polymerase activity of RT; second, generic RNA primers are not efficiently extended by RT but readily engage the RNase H activity of RT when annealed with DNA; third, the PPT RNA can direct both the DNA polymerase activity and a site-specific RNase H activity of RT. The mechanism by which RT discriminates between these substrates and executes the appropriate catalytic function is, however, poorly understood. Although RNase H cleavage analysis suggests the presence of different interaction modes of RT with substrates[16,17], crystal structures have so far revealed only one enzyme-binding orientation[4,18–22].

Single-molecule assay for enzyme–substrate interactions

For a better understanding of how RT interacts with substrates, we designed a single-molecule assay to measure the enzyme orientation relative to its substrate by using fluorescence resonance energy transfer (FRET)[23,24], a method well suited for probing dynamic interactions between proteins and nucleic acids[25–27]. Static FRET measurements have also been used previously to characterize the pre-translocation and post-translocation states of RT on a DNA duplex[28]. Because RT accommodates 19–22 base pairs of nucleic-acid duplex within its primer–template-binding cleft[19,22,29] (Fig. 1a), we constructed several duplex substrates with different backbone compositions and base sequences, each consisting of a 50-nucleotide (nt) oligonucleotide mimicking the template and a complementary 19–21-nt oligonucleotide emulating the primer (Fig. 1b and Supplementary Fig. 1). A Cy5 fluorophore was specifically attached to one of the single-stranded overhang regions on the template to serve as the FRET acceptor. We refer to the labelling schemes with Cy5 near the 5′ and 3′ ends of the primer as the 5* and 3* labels, respectively.

Surface-immobilized substrates were immersed in a solution containing RT molecules labelled with a FRET donor dye, Cy3, either at the RNase H domain (H-labelled) or at the fingers domain (F-labelled) of the p66 subunit (Fig. 1a). The two dye-labelling sites were located on opposite poles of the enzyme and separated by about 8 nm. An E478Q mutation was introduced into the RNase H domain to abolish its RNA cleavage activity so as to prevent the RT-induced degradation of nucleic-acid substrates during observation[30]. Experiments were conducted with this RNase-H-inactive variant unless otherwise mentioned. Neither dye attachment nor surface immobilization significantly altered the polymerase activity of RT (Supplementary Fig. 2).

Fluorescence of individual duplex substrates on the slide was monitored by using total-internal-reflection fluorescence (TIRF) microscopy with alternating laser excitations[25] at 532 and 635 nm

[1]Department of Chemistry and Chemical Biology, [2]Department of Physics, and [3]Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA. [4]HIV Drug Resistance Program, National Cancer Institute, Frederick, Maryland 21702, USA.
*These authors contributed equally to this work.

©2008 Nature Publishing Group

(Fig. 1c). The 532-nm light excites the FRET donor Cy3 without significantly exciting the acceptor Cy5, allowing us to detect the FRET between the Cy3-labelled RT and the Cy5-labelled substrate. The 635-nm light directly excites the Cy5 dye, providing a means to probe the presence of the nucleic-acid substrate and FRET acceptor independently of RT binding. Freely diffusing RT was observed to bind and dissociate from the substrates in real time. Each binding event caused an increase in the total fluorescence signal collected from both Cy3 and Cy5 channels under the 532-nm excitation without affecting the signal obtained under the 635-nm excitation (Fig. 1d). The observed FRET value allowed the enzyme orientation of each binding event to be determined.

RT binds DNA and RNA primers in opposite orientations

We first examined the binding orientation of RT on a 19-nt DNA primer hybridized to a 50-nt DNA template. When H-labelled RT was added to the 5*-labelled substrates, binding events consistently yielded high FRET values (centred at about 0.94; Fig. 2a), indicating an overwhelming tendency for the enzyme to bind with its RNase H domain close to the 5′ terminus of the primer. Conversely, predominantly low FRET values (about 0.14) were observed when H-labelled RT bound to an isogenic 3*-labelled substrate (Fig. 2b). Furthermore, F-labelled RT bound primarily to the 3*-labelled substrate with high FRET values (about 0.90; Fig. 2c), indicating that the DNA polymerase domain was located near the 3′ end of the primer. As a control, when both Cy3 and Cy5 were placed on the substrates, either near the same end of or flanking the duplex region, no significant change in FRET was observed on the addition of RT, suggesting that RT does not cause a sizable change in the photophysical properties of the dyes or the end-to-end distance of the duplex (Supplementary Fig. 3). Taken together, these results indicate that RT binds to the DNA–DNA primer–template complex with its polymerase active site between the fingers and palm domains close to the 3′ end of the primer and the RNase H domain near the 5′ end—an orientation that matches the polymerization-competent binding mode observed in RT–substrate co-crystal structures[19–22]. A virtually identical binding orientation was observed for RT on a 19-nt DNA primer annealed to a 50-nt RNA template (Supplementary Fig. 4). The same binding orientation was also observed for the RNase-H-active RT (without the E478Q mutation) on the DNA–DNA primer–template complex (Supplementary Fig. 5a).

Next we examined binding to an RNA primer annealed to a DNA template. The primer and template sequences were identical to those used above. Again, RT adopted a predominantly single binding configuration, but now with a drastically different orientation: H-labelled RT bound to the 5*-labelled substrates with low FRET values (about 0.27; Fig. 2d) but to the 3*-labelled substrates with primarily high FRET values (about 0.95; Fig. 2e); F-labelled RT bound to 5*-labelled substrate with high FRET values (about 0.88; Fig. 2f). These results unambiguously define a binding orientation on the RNA primer that is opposite to that on the DNA primer, with the DNA polymerase domain adjacent to the 5′ terminus of the primer and the RNase H domain close to the 3′ end. The same binding orientation was also found for the RNase-H-active RT (Supplementary Fig. 5b). This orientation clearly cannot support primer extension activity but directly explains the primary RNase-H-cleavage mode observed previously on similar substrates, in which the cleavage site is 18 nt from the 5′ terminus of the RNA[17]. The two opposite binding orientations on DNA and RNA primers were also observed on primers encoding an alternative sequence (Supplementary Fig. 6).

To identify which features were most important in discriminating between DNA and RNA primers and directing RT orientation, we

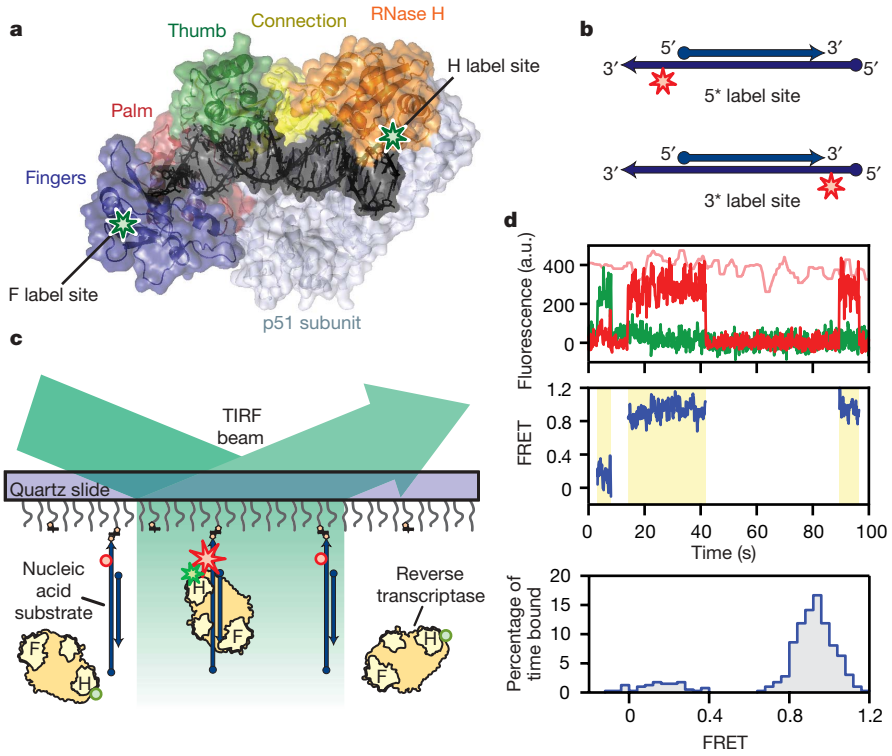


Figure 1 | Single-molecule FRET assay for probing the orientational dynamics of RT. a, The structure of HIV-1 RT bound to a DNA–DNA substrate[21]. Labelling sites for Cy3 on RT are highlighted by green stars. **b**, Nucleic-acid substrates consisted of a 19–21-nt primer strand annealed to a 50-nt template strand containing a Cy5 label (red star). Cy5 was either 3 nt from the 5′ end (circle) or 4–6 nt from the 3′ end (arrow) of the primer. **c**, Single-molecule detection of Cy3 (green star or sphere)-labelled RT binding to and dissociating from the surface-immobilized nucleic-acid substrates labelled with Cy5 (red star or sphere). The stars and spheres indicate dyes that do and do not emit fluorescence, respectively. **d**, FRET analysis for RT binding to a single primer–template complex. Top: fluorescence time traces from Cy3 (green) and Cy5 (red) under excitation at 532 nm and that from Cy5 (pink) under excitation at 635 nm. Middle: FRET value calculated over the duration of the binding events (yellow shaded regions). Bottom: FRET distribution histogram created for the binding events.

©2008 Nature Publishing Group

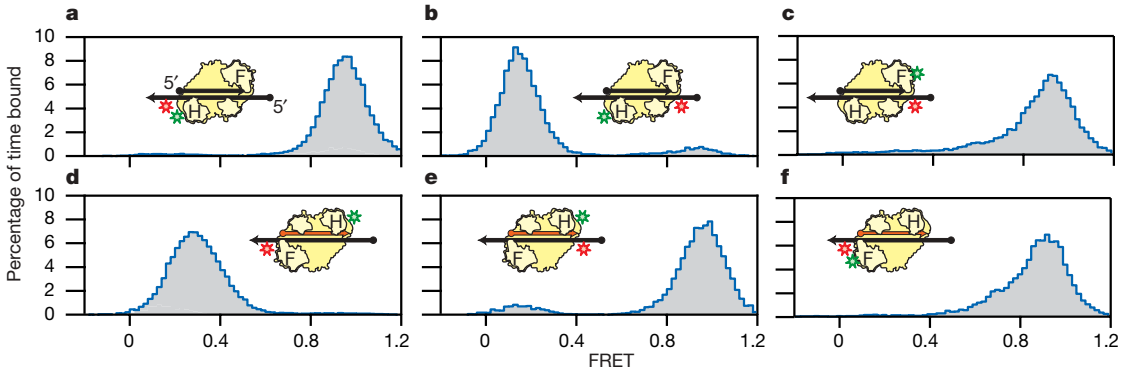


Figure 2 | FRET distributions of RT bound to nucleic-acid substrates reveal distinct RT binding orientations on RNA and DNA primers. **a**, RT with Cy3 (green star) attached in the H-labelling scheme was allowed to bind substrates consisting of DNA primer and template (black arrows), with Cy5 (red star) attached in the 5*-labelling scheme. **b**, H-labelled RT bound to 3*-labelled DNA duplex substrates. **c**, F-labelled RT bound to 3*-labelled DNA duplex substrates. **d**, H-labelled RT bound to 5*-labelled hybrid duplex substrates consisting of a RNA primer (orange arrow) and a DNA template (black arrow). **e**, H-labelled RT bound to 3*-labelled hybrid duplex substrates. **f**, F-labelled RT bound to 5*-labelled hybrid duplex substrates. The RT binding orientations consistent with the FRET distributions are depicted.

designed a series of 19-nt chimaeric primers containing different compositions of RNA and DNA nucleotides denoted by *x*R:*y*D (that is, *x* RNA nt at the 5′ end and *y* DNA nt at the 3′ end). These chimaeras were annealed to a 50-nt DNA template and incubated with H-labelled RT (Fig. 3a and Supplementary Fig. 7). Whereas pure DNA and RNA primers bound to RT predominantly in a single orientation, most chimaeric primers supported both high and low FRET orientations (Supplementary Fig. 7). The free-energy difference (ΔG) between the two states was most sensitive to the sugar composition of the four or five nucleotides located at each end of the 19-nt primer (Fig. 3b), suggesting that the interactions between RT and nucleic acid at opposite ends of the primer–template binding cleft were most important in determining the binding orientation. This observation is consistent with the crystal structures, which show RT–substrate contacts primarily clustered in two regions near the DNA polymerase and RNase H active sites[19,22]. A single nucleotide provided the strongest determinant of binding orientation: changing the sugar content of the fifth nucleotide from the primer 5′ terminus alone caused a nearly $2k_BT$ (where k_B is the Boltzmann constant) change in ΔG (Fig. 3b). This position makes specific contacts with RT residues T473 and Q475 located within the RNase H primer grip[19,22]. These residues are conserved between RNases H found in viruses, bacteria and humans[22,31,32]. Replacement of these residues with alanine in HIV-1 RT decreases the DNA synthesis rate and inhibits virus infectivity[33]. Overall, the backbone composition of nucleotides near the 5′ end of the primer had a greater influence on enzyme binding orientation than those near the 3′ end (Fig. 3b). To test this notion further, we constructed a new chimaeric primer 9D:10R with 5′-proximal DNA and 3′-proximal RNA, which had the same DNA/RNA ratio as the 10R:9D primer but with a different 5′-end backbone composition. As expected, the 9D:10R and 10R:9D primers supported opposite binding orientations of RT that closely resembled the orientational distributions of RT bound to pure DNA and RNA primers, respectively (Fig. 3a).

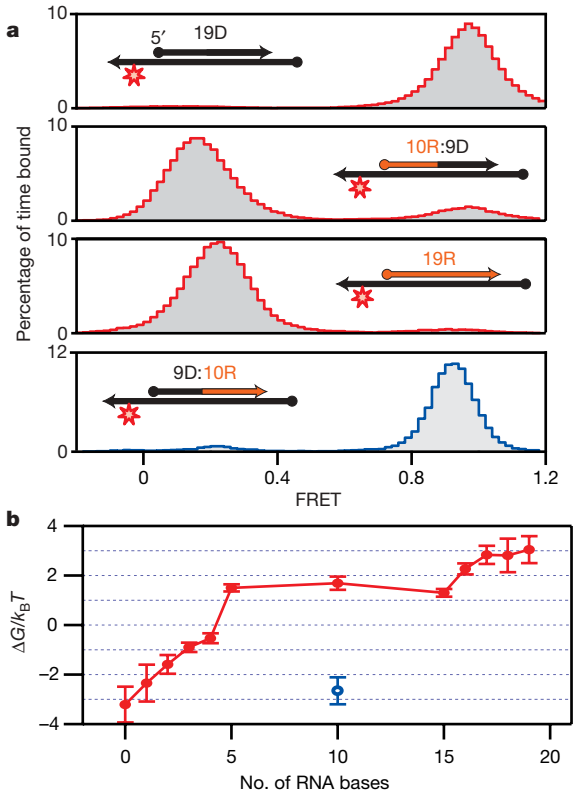


Figure 3 | Binding orientation of RT on chimaeric substrates. **a**, Selected FRET distributions of H-labelled RT bound to 5*-labelled substrates containing various 19-nt chimaeric RNA:DNA primers hybridized to a DNA template. FRET distributions of other *x*R:*y*D substrates are shown in Supplementary Fig. 7. **b**, The free-energy difference ΔG between the high-FRET and low-FRET orientations plotted as a function of RNA content for both *x*R:*y*D (red) and 9D:10R (blue) chimaeras. Error bars indicate s.e.m. ($n = 3$).

Binding orientation determines enzymatic activity of RT

The observation that RT bound to the DNA and RNA primers with opposite orientations suggests a hypothesis: that primer extension activity is determined by the binding orientation of the enzyme. To test this model, we probed the DNA polymerase activity of RT on the DNA (19D) and RNA (19R) primers as well as the chimaeric primers 9D:10R and 10R:9D, each annealed to a 50-nt DNA template (Fig. 4a and Supplementary Fig. 8). RT was capable of extending the 3′ end of both 19D and 9D:10R rapidly, with a rate comparable to previously reported steady-state extension rates of DNA primers[34,35], even though the 9D:10R primer contained a ribonucleotide sugar backbone at its 3′ terminus. Polymerase activity was strongly inhibited for both 19R and 10R:9D. Furthermore, the rate of primer extension correlated with the fraction of time for which the RT enzyme bound in the polymerase-component orientation (Fig. 4b). These results indicate that the binding orientation is the strongest determinant of the primer-extension activity, whereas the content of the sugar-phosphate backbone contacting the active site of DNA polymerase is less important for synthesis activity. Our results also suggest a surprising allosteric effect in which contacts between the 5′ end of the primer and the RNase H primer grip regulate the DNA

©2008 Nature Publishing Group

polymerase activity by determining the orientation of the enzyme on the substrate.

Dynamic binding orientations of RT on PPT substrates

Although RNA primers do not generally support the initiation of DNA synthesis by RT, two copies of 15-nt RNA purine sequences, referred to as the PPTs, uniquely serve as primers for plus-strand DNA synthesis[2,11–13,15]. During infection, RT cleaves precisely at the 3′ terminus of the PPT, allowing DNA synthesis to be initiated at this position. The enzyme then removes the PPT primer by cleaving at its junction with the nascent DNA[2,15]. How RT interacts with the PPT to support the activities of both DNA polymerase and RNase H is still unknown. To address this question, we constructed three oligonucleotides encoding the PPT sequence to mimic different stages in plus-strand DNA synthesis. To simulate a PPT sequence that has not yet been cleaved at its 3′ terminus, we introduced a 2-nt RNA extension, creating the PPT:r2 RNA. Similarly, the PPT:d2 chimaera (containing a 2-nt DNA extension) was used to emulate a plus-strand primer from which DNA synthesis has already started. These primers were annealed to a 50-nt DNA template and assayed for RT binding (Fig. 5a–c). The FRET distribution of RT bound to the PPT:r2 primer–template complex was quantitatively similar to that observed for a non-specific RNA primer (compare Fig. 5a with Fig. 2d), suggesting that RT was predominantly bound in a cleavage orientation. By contrast, on the PPT and PPT:d2 substrates, RT spent a substantial portion of time in the high-FRET, polymerization-competent orientation (Fig. 5b, c). These data suggest that the priming activity of the PPT for plus-strand DNA synthesis originates from its specific ability among RNA sequences to direct RT binding in a polymerase-competent orientation.

On the substrates that support both DNA-polymerase-competent and RNase-H-competent orientations, including the PPT, PPT:d2 and chimaeric RNA:DNA primers, RT exhibited spontaneous transitions between these two orientations, and flipping transitions were observed with different labelling schemes (Fig. 5d and Supplementary Fig. 9). The flipping transition did not seem to require the binding of multiple enzymes, because the flipping kinetics were independent of the RT concentration. The observation of flipping transitions within a single binding event was unexpected, considering the extensive network of contacts between the RT and its substrates[19,22].

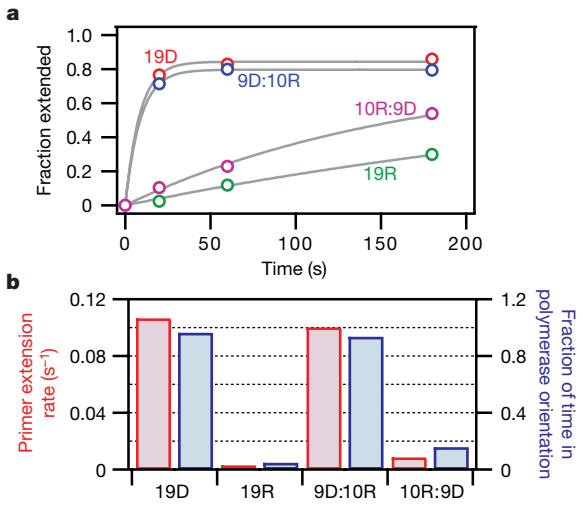


Figure 4 | The DNA polymerase activity of RT correlates with its binding orientation on substrates. **a**, Primer extension activity of RT assayed on four selected primers (19D, 19R, 9D:10R and 10R:9D) annealed to a DNA template. The fraction of primers that had been extended by more than one base is plotted as a function of time for the four primers (coloured circles). The data were fitted to a single-exponential decay (grey lines) to deduce the primer extension-rate constants. **b**, Rate constants of primer extension (red) correlate with the fraction of time that the RT bound in the high-FRET orientation conducive to polymerization (blue).

Small-molecule ligands regulate RT binding orientation

To explore the flipping mechanism, we investigated the effect of small molecules, including dNTP and non-nucleoside RT inhibitors (NNRTIs), on the equilibrium and rate constants of the flipping transitions by using the PPT:d2 primer or a modified primer containing a chain-terminating dideoxyribonucleotide (PPT:dd2). The addition of dTTP, the next cognate nucleotide for primer extension, stabilized the high-FRET, polymerase-competent orientation of RT (Fig. 5d, e). The stabilization magnitude increased with dTTP concentration over a physiologically relevant range (Fig. 5e). Kinetically, the addition of 1 mM dTTP decreased the rate constant of flipping from the high-FRET to the low-FRET orientation, $k_{high–low}$, 20-fold without substantially affecting the reverse rate, $k_{low–high}$ (Supplementary Fig. 10). By contrast, the addition of a mismatched nucleotide (dCTP) did not induce a similar effect (Supplementary Fig. 11).

NNRTIs are clinically approved anti-HIV drugs[36] that bind to a hydrophobic pocket[4] near the polymerase active site of RT to inhibit DNA synthesis allosterically[37]. We examined one such NNRTI, nevirapine, for its effects on the orientational dynamics of RT. Nevirapine seemed to have an opposite effect to that of cognate dNTP. The addition of nevirapine significantly destabilized the high-FRET,

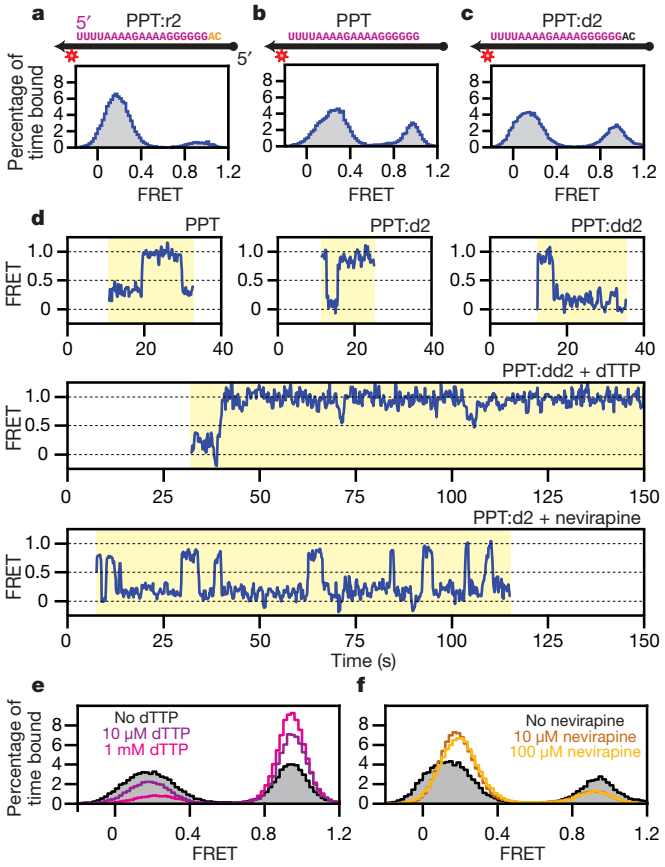


Figure 5 | Dynamic binding orientations of RT on PPT substrates. **a**–**c**, FRET histograms of H-labelled RT bound to substrates containing 5*-labelled PPT:r2 (**a**), PPT (**b**) or PPT:d2 (**c**) primers annealed to DNA templates. The PPT sequence is shown in violet letters and the 2-nt RNA and DNA extensions are coloured orange and black, respectively. The DNA template is shown as a black arrow. **d**, FRET time traces of RT bound to PPT, PPT:d2 and PPT:dd2 substrates, showing spontaneous flipping transitions between the two binding orientations. **e**, FRET histograms of RT bound to PPT:dd2 substrates in the presence of 0, 10 μM and 1 mM dTTP. **f**, FRET histograms of RT bound to PPT:d2 substrates in the presence of 0, 10 and 100 μM nevirapine.

©2008 Nature Publishing Group

polymerase-competent orientation (Fig. 5d, f): the presence of the drug caused a 3.5-fold increase in the flipping rate from the high-FRET to the low-FRET orientation, $k_{high-low}$, without significantly altering the reverse rate, $k_{low-high}$ (Supplementary Fig. 10). A similar effect was observed for a different NNRTI, efavirenz. These results provide a structural basis for the previously observed specific inhibition of the PPT-initiated plus-strand DNA synthesis by NNRTIs, which occurs at a 40-fold lower concentration of NNRTI than that required for the inhibition of minus-strand DNA synthesis[38].

Discussion

We have developed a single-molecule FRET assay to monitor the interactions between HIV RT and its nucleic-acid substrates in real time. These experiments directly revealed two opposite orientations with which the RT enzyme binds to DNA and RNA primers. The primary determinant of the enzyme orientation is the sugar backbone composition of the four or five nucleotides at each end of the primer, located within the polymerase and RNase H primer grip regions of the RT binding cleft. The primer-extension activity of RT is quantitatively correlated with the enzyme orientation, providing a structural basis for the primer-selection rule of RT.

Remarkably, the enzyme can bind to the special PPT RNA sequence, which directs the transition from minus-strand to plus-strand synthesis, in both orientations. Furthermore, the enzyme can flip spontaneously between the two states despite the extensive contact between RT and its nucleic-acid substrates. The flipping kinetics were altered by both cognate nucleotides and non-nucleoside RT inhibitors, but the two types of small molecule had opposite effects. Whereas the addition of cognate nucleotides caused a drastic decrease in the flipping rate from the polymerase-competent to the RNase H-competent orientation, the NNRTI substantially increased the same rate constant. NNRTIs and dNTPs have been shown to have opposite effects on the structural dynamics of the fingers and thumb subdomains[20]: whereas dNTPs bring these regions closer together to form a tighter clamp on the nucleic-acid substrate, NNRTIs cause further separation of the two subdomains. Our data therefore suggest a potential pathway for the flipping transition that requires relaxation of the 'grip' formed by the fingers and thumb subdomains around the nucleic-acid substrate. This spontaneous structural reorganization of the RT–substrate complex potentially allows the enzyme to rapidly explore multiple binding orientations that support distinct functions, thereby increasing replication efficacy.

METHODS SUMMARY

For single-molecule measurements of RT interactions with nucleic acids, the dye-labelled nucleic-acid substrates were immobilized on poly(ethylene glycol)-coated fused quartz slides through a biotin–streptavidin linkage. Binding of the dye-labelled RT molecules in solution to the immobilized substrates was monitored with the TIRF imaging geometry with alternating 532-nm and 635-nm excitations[25]. FRET histograms were constructed from binding events of hundreds of molecules. To calculate the difference in free energy between high-FRET and low-FRET states, these histograms were fitted to a double gaussian function and ΔG was determined from the ratio between the areas under the two gaussian peaks. Within each binding event, high-FRET and low-FRET substates were identified, and the lifetimes of the binding events and the substates were recorded. These lifetimes were combined with a simple kinetic model to derive the rate constants of transition between the substates as well as the rate of dissociation from each substate.

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

Received 12 November 2007; accepted 20 March 2008.

1. Goff, S. P. in *Fields Virology* Vol. 2 (ed. D. M. Knipe & P. M. Howley) 1871–1940 (Lippincott Williams & Wilkins, Philadelphia, 2001).
2. Champoux, J. J. in *Reverse Transcriptase* (ed. A. M. Skalka & S. P. Goff) 103–118 (Cold Spring Harbor Laboratory Press, New York, 1993).
3. Hostomsky, Z., Hostomska, Z., Fu, T. B. & Taylor, J. Reverse transcriptase of human immunodeficiency virus type 1: functionality of subunits of the heterodimer in DNA synthesis. *J. Virol.* **66,** 3179–3182 (1992).
4. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256,** 1783–1790 (1992).
5. Baltimore, D. Viral RNA-dependent DNA polymerase: RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226,** 1209–1211 (1970).
6. Temin, A. M. & Mizutani, S. Viral RNA-dependent DNA polymerase: RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* **226,** 1211–1213 (1970).
7. Aiyar, A., Cobrinik, D., Ge, Z., Kung, H. J. & Leis, J. Interaction between retroviral U5 RNA and the TψC loop of the $tRNA^{Trp}$ primer is required for efficient initiation of reverse transcription. *J. Virol.* **66,** 2464–2472 (1992).
8. Leis, J. P., Berkower, I. & Hurwitz, J. RNA-dependent DNA polymerase activity of RNA tumor viruses. 5. Mechanism of action of ribonuclease H isolated from avian myeloblastosis virus and *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **70,** 466–470 (1973).
9. Hansen, J., Schulze, T., Mellert, W. & Moelling, K. Identification and characterization of HIV-specific RNase H by monoclonal antibody. *EMBO J.* **7,** 239–243 (1988).
10. Tanese, N., Telesnitsky, A. & Goff, S. P. Abortive reverse transcription by mutants of Moloney murine leukaemia virus deficient in the reverse transcriptase-associated RNase H function. *J. Virol.* **65,** 4387–4397 (1991).
11. Finston, W. I. & Champoux, J. J. RNA-primed initiation of moloney murine leukemia-virus plus strands by reverse-transcriptase *in vitro*. *J. Virol.* **51,** 26–33 (1984).
12. Omer, C. A., Resnick, R. & Faras, A. J. Evidence for involvement of an RNA primer in initiation of strong-stop plus DNA-synthesis during reverse transcription *in vitro*. *J. Virol.* **50,** 465–470 (1984).
13. Huber, H. E. & Richardson, C. C. Processing of the primer for plus strand DNA-synthesis by human immunodeficiency virus-1 reverse-transcriptase. *J. Biol. Chem.* **265,** 10565–10573 (1990).
14. Schultz, S. J., Zhang, M. H., Kelleher, C. D. & Champoux, J. J. Analysis of plus-strand primer selection, removal, and reutilization by retroviral reverse transcriptases. *J. Biol. Chem.* **275,** 32299–32309 (2000).
15. Rausch, J. W. & Le Grice, S. F. 'Binding, bending and bonding': polypurine tract-primed initiation of plus-strand DNA synthesis in human immunodeficiency virus. *Int. J. Biochem. Cell Biol.* **36,** 1752–1766 (2004).
16. Gopalakrishnan, V., Peliska, J. A. & Benkovic, S. J. Human immunodeficiency virus type 1 reverse transcriptase: spatial and temporal relationship between the polymerase and RNase H activities. *Proc. Natl Acad. Sci. USA* **89,** 10763–10767 (1992).
17. Wisniewski, M., Balakrishnan, M., Palaniappan, C., Fay, P. J. & Bambara, R. A. Unique progressive cleavage mechanism of HIV reverse transcriptase RNase H. *Proc. Natl Acad. Sci. USA* **97,** 11978–11983 (2000).
18. Arnold, E. *et al.* Structure of HIV-1 reverse transcriptase/DNA complex at 7 Å resolution showing active site locations. *Nature* **357,** 85–89 (1992).
19. Ding, J. *et al.* Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. *J. Mol. Biol.* **284,** 1095–1111 (1998).
20. Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282,** 1669–1675 (1998).
21. Sarafianos, S. G. *et al.* Structures of HIV-1 reverse transcriptase with pre- and post-translocation AZTMP-terminated DNA. *EMBO J.* **21,** 6614–6624 (2002).
22. Sarafianos, S. G. *et al.* Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J.* **20,** 1449–1461 (2001).
23. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc. Natl Acad. Sci. USA* **58,** 719–726 (1967).
24. Ha, T. *et al.* Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl Acad. Sci. USA* **93,** 6264–6268 (1996).
25. Kapanidis, A. N. *et al.* Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314,** 1144–1147 (2006).
26. Stone, M. D. *et al.* Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature* **446,** 458–461 (2007).
27. Myong, S., Bruno, M. M., Pyle, A. M. & Ha, T. Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. *Science* **317,** 513–516 (2007).
28. Rothwell, P. J. *et al.* Multiparameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase:primer/template complexes. *Proc. Natl Acad. Sci. USA* **100,** 1655–1660 (2003).
29. Metzger, W., Hermann, T., Schatz, O., Le Grice, S. F. & Heumann, H. Hydroxyl radical footprint analysis of human immunodeficiency virus reverse transcriptase–template•primer complexes. *Proc. Natl Acad. Sci. USA* **90,** 5909–5913 (1993).

©2008 Nature Publishing Group

30. Schatz, O., Cromme, F. V., Gruninger-Leitch, F. & Le Grice, S. F. Point mutations in conserved amino acid residues within the C-terminal domain of HIV-1 reverse transcriptase specifically repress RNase H function. *FEBS Lett.* **257,** 311–314 (1989).
31. Nowotny, M., Gaidamakov, S. A., Crouch, R. J. & Yang, W. Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **121,** 1005–1016 (2005).
32. Nowotny, M. *et al.* Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription. *Mol. Cell* **28,** 264–276 (2007).
33. Julias, J. G., McWilliams, M. J., Sarafianos, S. G., Arnold, E. & Hughes, S. H. Mutations in the RNase H domain of HIV-1 reverse transcriptase affect the initiation of DNA synthesis and the specificity of RNase H cleavage *in vivo*. *Proc. Natl Acad. Sci. USA* **99,** 9515–9520 (2002).
34. Hsieh, J. C., Zinnen, S. & Modrich, P. Kinetic mechanism of the DNA-dependent DNA polymerase activity of human immunodeficiency virus reverse transcriptase. *J. Biol. Chem.* **268,** 24607–24613 (1993).
35. Kati, W. M., Johnson, K. A., Jerva, L. F. & Anderson, K. S. Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.* **267,** 25988–25997 (1992).
36. El Safadi, Y., Vivet-Boudou, V. & Marquet, R. HIV-1 reverse transcriptase inhibitors. *Appl. Microbiol. Biotechnol.* **75,** 723–737 (2007).
37. Spence, R. A., Kati, W. M., Anderson, K. S. & Johnson, K. A. Mechanism of inhibition of HIV-1 reverse transcriptase by nonnucleoside inhibitors. *Science* **267,** 988–993 (1995).
38. Grobler, J. A. *et al.* HIV-1 reverse transcriptase plus-strand initiation exhibits preferential sensitivity to non-nucleoside reverse transcriptase inhibitors *in vitro*. *J. Biol. Chem.* **282,** 8005–8010 (2007).

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

Acknowledgements We thank J. Miller for providing initial samples of RT, and S. Liu for discussions. This work is supported in part by grants from the National Institutes of Health (GM 068518) and the Packard Foundation to X.Z., and from the Intramural Research Program of the Centre for Cancer Research, National Cancer Institute (to S.F.J.L.G.). X.Z. is a Howard Hughes Medical Institute investigator. E.A.A. is a Jane Coffin Childs postdoctoral fellow. Nevirapine and efavirenz were provided through the AIDS Research and Reference Reagent Program of the National Institutes of Health.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to X.Z. (zhuang@chemistry.harvard.edu).

©2008 Nature Publishing Group

METHODS

Preparation and labelling of RT and nucleic-acid substrates. The H-labelled RT was derived from a mutant enzyme in which a single-residue modification E478Q (ref. 30) was introduced to the RNase H domain to inhibit the RNase H activity. This modification was employed to prevent substrate cleavage during the single-molecule measurement. To engineer the H-labelled RT, native cysteine residues located at positions 38 and 280 were changed to serine, and a unique cysteine residue was introduced at the carboxy terminus of the p66 subunit to allow specific dye labelling through a thiol-maleimide reaction[39]. F-labelled RT was created from a similar mutant containing a unique cysteine residue at position 38 of the p66 subunit. Purified RT was incubated with Cy3-maleimide (GE Healthcare) and allowed to react for 60 min in 100 mM phosphate buffer (pH 7). The Cy3-labelled RT was then purified by dialysis for more than 48 h to remove the unreacted dye molecules. The p51 subunit was unlabelled.

Synthetic DNA (Qiagen Operon) and synthetic RNA (Dharmacon) oligonucleotides were purified by PAGE. All template strands (50 nt long) contained a biotin moiety attached to the 3′ end as well as an internal amino modifier (dT C6). The amine group was labelled with a monoreactive Cy5 in accordance with the manufacturer's instructions (GE Healthcare) and the labelled template strands were HPLC-purified by reverse-phase chromatography on a C_8 column (GE Healthcare). The primer strands (19–21-nt long) were annealed at 60 °C for 10 min to the template strands (in 10 mM Tris-HCl pH 8, 80 mM NaCl, 1 mM EDTA) at a roughly 15:1 ratio, cooled to less than 30 °C over about 1 h, and stored at −20 °C. Removal of unannealed primer stands was not necessary because they did not contain either dye molecules or biotin groups for surface immobilization.

Single-molecule FRET measurements of RT-substrate binding orientation. Quartz slides were cleaned with the use of argon plasma (Harrick Scientific), treated with 1% (w/v) Vectabond (Vector Laboratories) in acetone, rinsed, and then incubated with 20% (w/v) methoxy-PEG (M_r 5,000; Nektar Therapeutics) and 0.2% biotin-PEG (M_r 5,000; Nektar Therapeutics) in 0.1 M sodium bicarbonate (pH 8.4) for at least 3 h. Streptavidin (0.2 mg ml^{-1}; Molecular Probes) and BSA (0.5 mg ml^{-1}, New England Biolabs) in 10 mM Tris-HCl (pH 8), and 10 mM NaCl was applied to the slide before immobilization of the biotinylated primer–template complexes (50 pM).

Donor and acceptor fluorescence signals were collected on a prism-type TIRF microscope. The FRET donor Cy3 was excited by a 532-nm YAG laser (Crystal Laser) and direct excitation of Cy5 was performed by a 635-nm laser (Coherent). Emissions from donor and acceptor were separated with dichroic mirrors (Chroma Technology) and imaged onto the two halves of an Andor Ixon 887 back-illuminated electron-multiplying charge-coupled device. The FRET value is defined as $I_A/(I_A + I_D)$, where I_A and I_D are the fluorescence signals detected from the acceptor and donor channels, respectively, under 532-nm excitation.

During image acquisition, Cy3-labelled RT (6–24 nM) was added to the sample containing surface-immobilized primer–template complexes in an imaging buffer containing 40 mM NaCl, 50 mM Tris-HCl pH 8.0, 6 mM $MgCl_2$, 0.1 mg ml^{-1} BSA, 10% w/v glucose and about 1.5 mM Trolox (Sigma Aldrich)[40]. An oxygen scavenger system (300 μg ml^{-1} glucose oxidase, 40 μg ml^{-1} catalase) was also added to the sample to decrease photobleaching. Positions of Cy5-labelled primer–template complexes were initially located by direct excitation of Cy5 with the 635-nm laser. FRET between Cy3 and Cy5 and fluorescence of Cy5 from direct excitation were then monitored by alternating the 532-nm and 635-nm excitations.

By increasing the intensity of the 532-nm excitation, photobleaching of Cy3 on the p66 subunit could be induced, shortening the duration of the observed fluorescence signal during a binding event. Under these conditions, Cy3 photobleaching was observed to occur in a single step, as manifested by the one-step decrease in the fluorescence signal to the background level, indicating that only one Cy3 dye was present on the enzyme–substrate complex. We therefore conclude that most of the binding events involved one p66 subunit. Because the RT concentrations used in the single-molecule imaging experiments (10–20 nM) were far below the equilibrium constant K_d for p66–p51 dimer formation (200–300 nM)[41], an excess of unlabelled p51 was added (about 300 nM) to ensure that most p66 subunits formed dimers with p51. Because the p66–p51 heterodimers are more stable than the p66–p66 homodimers, the probability of potential p66–p66 dimer formation during the experiment was thus minimal under these conditions. Finally, because the p66 and p51 subunits alone showed much lower affinities than the p66–p51 dimers for the nucleic-acid substrates, we conclude that most of the binding events observed in the experiments involve a single p66–p51 dimer.

DNA polymerase activity measurements by single-molecule FRET. To confirm that RT retains DNA polymerase activity on surface-immobilized primer–template complexes, we performed an *in situ* primer extension assay by using single-molecule FRET. We designed a primer–template complex capable of reporting primer extension with FRET: the duplex region was labelled with Cy5, whereas Cy3 was placed at the 5′ end of the template, 10 nt from the 3′ end of the primer and 19 nt from the Cy5 (Supplementary Fig. 2a). Extension of the primer by the DNA polymerase activity of RT converted the single-stranded region of the template to double-stranded, stretching the template and lowering the FRET from about 0.8 to about 0.5 (Supplementary Fig. 2b). To monitor primer extension in real time, FRET time traces from the primer–template complexes in a single field were recorded while 100 nM unlabelled RT and 250 μM dNTPs were added. The *in situ* primer extension rate was determined by averaging all trajectories that showed a strong fluorescent signal and non-zero FRET (to avoid complexes with bleached dyes) and fitting the average trace with a single-exponential decay. This rate was compared with the results obtained from the unimmobilized primer–template complexes incubated with unlabelled or Cy3-labelled RT with the gel electrophoresis assay, as described below (Supplementary Fig. 2c).

DNA polymerase activity measurements by gel electrophoresis. To measure the DNA polymerase activity of Cy3-labelled RT in comparison with unlabelled RT, 100 nM Cy3-labelled or unlabelled RT was preincubated for 10–15 min in 150 μl of imaging buffer (described above) containing each dNTP at 250 μM. Extension of Cy5 end-labelled primer was initiated by the addition of 1.5 μl of 500 nM annealed primer–template complexes to the preincubated solution of RT and dNTP, and aliquots were removed at various time points, mixed in 90% v/v formamide, 1 × TBE buffer and 10 mM EDTA to quench the reaction, and heated to 90 °C for 1 min to denature the primer–template complexes. The products were then loaded on a precast 10% polyacrylamide gel (8 M urea; Bio-Rad). Intensities of fluorescent bands were measured with a Typhoon gel scanner (GE Healthcare). This condition was used to provide a better comparison with the single-molecule *in situ* data where RT and dNTP were added to the surface-immobilized substrate simultaneously. We note that the apparent primer extension rates measured in these experiments reflects the convolution of the binding rate of RT to the substrate and the true rate of elongation by RT, and thus should be slower than the true elongation rate as probed in the following experiments.

To measure the DNA polymerase activity of the RT on substrates containing the 19D, 19R, 10R:9D or 9D:10R primers, 3 μM unlabelled RT was incubated in the imaging buffer with 500 nM primer–template complexes. The 5′ end of each primer strand was labelled with ^{32}P. Primer extension was initiated by the addition of 1 mM dNTP. The reaction was quenched at various time points by the addition of 500 mM EDTA and formamide. The reaction products were heated to 90 °C for 1 min then separated on an 8% polyacrylamide gel. Radiolabelled primers were imaged on a PhosphoImager cassette with a Typhoon gel scanner. The fraction of extended primers was quantified as a function of time. These data were fitted to single-exponential decays with an offset to deduce the primer extension rate constant for each substrate. For the 19R and 10R:9D primers, the fit was constrained to asymptote at a value of 0.8, which was the saturated fraction of extended primers for 19D and 9D:10R.

39. Rausch, J. W., Sathyanarayana, B. K., Bona, M. K. & Le Grice, S. F. Probing contacts between the ribonuclease H domain of HIV-1 reverse transcriptase and nucleic acid by site-specific photocross-linking. *J. Biol. Chem.* **275,** 16015–16022 (2000).
40. Rasnik, I., McKinney, S. A. & Ha, T. Nonblinking and long-lasting single-molecule fluorescence imaging. *Nature Methods* **3,** 891–893 (2006).
41. Venezia, C. F., Howard, K. J., Ignatov, M. E., Holladay, L. A. & Barkley, M. D. Effects of efavirenz binding on the subunit equilibria of HIV-1 reverse transcriptase. *Biochemistry* **45,** 2779–2789 (2006).

©2008 Nature Publishing Group

Vol 453|8 May 2008|doi:10.1038/nature06879

ARTICLES

Kemp elimination catalysts by computational enzyme design

Daniela Röthlisberger[1]*, Olga Khersonsky[4]*, Andrew M. Wollacott[1]*, Lin Jiang[1,2], Jason DeChancie[6], Jamie Betker[3], Jasmine L. Gallaher[3], Eric A. Althoff[1], Alexandre Zanghellini[1,2], Orly Dym[5], Shira Albeck[5], Kendall N. Houk[6], Dan S. Tawfik[4] & David Baker[1,2,3]

The design of new enzymes for reactions not catalysed by naturally occurring biocatalysts is a challenge for protein engineering and is a critical test of our understanding of enzyme catalysis. Here we describe the computational design of eight enzymes that use two different catalytic motifs to catalyse the Kemp elimination—a model reaction for proton transfer from carbon—with measured rate enhancements of up to 10^5 and multiple turnovers. Mutational analysis confirms that catalysis depends on the computationally designed active sites, and a high-resolution crystal structure suggests that the designs have close to atomic accuracy. Application of *in vitro* evolution to enhance the computational designs produced a >200-fold increase in k_{cat}/K_m (k_{cat}/K_m of 2,600 $M^{-1}s^{-1}$ and k_{cat}/k_{uncat} of $>10^6$). These results demonstrate the power of combining computational protein design with directed evolution for creating new enzymes, and we anticipate the creation of a wide range of useful new catalysts in the future.

Naturally occurring enzymes are extraordinarily efficient catalysts[1]. They bind their substrates in a well-defined active site with precisely aligned catalytic residues to form highly active and selective catalysts for a wide range of chemical reactions under mild conditions. Nevertheless, many important synthetic reactions lack a naturally occurring enzymatic counterpart. Hence, the design of stable enzymes with new catalytic activities is of great practical interest, with potential applications in biotechnology, biomedicine and industrial processes. Furthermore, the computational design of new enzymes provides a stringent test of our understanding of how naturally occurring enzymes work. In the past several years, there has been exciting progress in designing new biocatalysts[2,3].

Here we describe the use of our recently developed computational enzyme design methodology[4] to create new enzyme catalysts for a reaction for which no naturally occurring enzyme exists: the Kemp elimination[5,6]. The reaction, shown in Fig. 1a, has been extensively studied as an activated model system for understanding the catalysis of proton abstraction from carbon—a process that is normally restricted by high activation-energy barriers[7,8].

Computational design method

The first step in our protocol for designing new enzymes is to choose a catalytic mechanism and then to use quantum mechanical transition state calculations to create an idealized active site with protein functional groups positioned so as to maximize transition state stabilization (Fig. 1b). The key step for the Kemp elimination is deprotonation of a carbon by a general base. We chose two different catalytic bases for this purpose: first, the carboxyl group of an aspartate or glutamate side chain, and, second, the imidazole of a histidine positioned and polarized by the carboxyl group of an aspartate or glutamate (we refer to this combination as a His–Asp dyad). The two choices have complementary strengths and weaknesses. The advantage of the carboxylate is that it is likely to be in the basic (deprotonated) form, but partial desolvation of the charged group in an apolar environment (to increase its relatively weak basicity) could destabilize the protein and further desolvation by the substrate could oppose binding. Although histidine is a better general base than a carboxylate, it is necessary to regulate both its pK_a and its tautomeric state. Coupling the histidine with a base such as aspartate in a dyad serves to both position the histidine and increase its basicity. If the pK_a of histidine is raised too high, however, it can become doubly protonated, rendering it ineffective as a base.

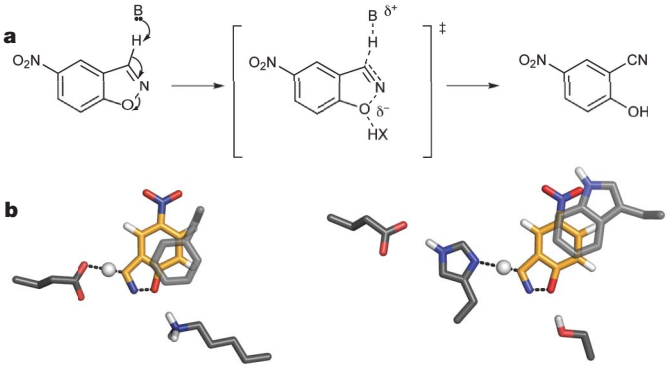


Figure 1 | Reaction scheme and catalytic motifs used in design. a, The Kemp elimination proceeds by means of a single transition state, which can be stabilized by a base deprotonating the carbon and the dispersion of the resulting negative charge; a hydrogen bond donor can also be used to stabilize the partial negative charge on the phenolic oxygen. **b**, Examples of active site motifs highlighting the two choices for the catalytic base (a carboxylate (left) or a His–Asp dyad (right)) used for deprotonation, and a π-stacking aromatic residue for transition state stabilization. For each catalytic base, all combinations of hydrogen bond donor groups (Lys, Arg, Ser, Tyr, His, water or none) and π-stacking interactions (Phe, Tyr, Trp) were input as active site motifs into RosettaMatch.

[1]Department of Biochemistry, [2]Biomolecular Structure and Design, and [3]Howard Hughes Medical Institute, University of Washington, Seattle, Washington 98195, USA. [4]Department of Biological Chemistry, and [5]Israel Structural Proteomics Center, Weizmann Institute of Science, Rehovot 76100, Israel. [6]Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90095, USA.
*These authors contributed equally to this work.

©2008 Nature Publishing Group

For both the carboxylate- and histidine-based mechanisms, we included additional functional groups in the idealized active sites to further facilitate catalysis using both quantum mechanical and classical methods[9]. A hydrogen bond donor was used to stabilize the developing negative charge on the phenolic oxygen in the otherwise hydrophobic active site. Catalytic motifs lacking the H-bond donor were also tested, because the developing negative charge is relatively small in the transition state and can be easily solvated by water[9,10]. For each choice of catalytic site composition, density functional theory quantum-mechanical methods[11–13] were used to optimize the placement and orientations of the catalytic groups around the transition state for maximal stabilization (see Methods). Finally, because stabilization of the transition state by charge delocalization is a key factor in catalysis of the Kemp elimination[5–7,10,14], we chose to stack aromatic amino acid side chains on the planar transition state (Fig. 1b) using idealized π-stacking geometries[15].

We next used the RosettaMatch hashing algorithm[4] to search for constellations of protein backbone positions capable of supporting these idealized active sites in a large set of stable protein scaffolds with ligand-binding pockets and high-resolution crystal structures. As described in the Methods, the His–Asp dyad required generalizing RosettaMatch to handle side chains, such as the Asp, for which the range of allowed positions are referenced to another catalytic side chain rather than to the transition state; this was accomplished by identifying, for each His rotamer in a scaffold, the set of Asp rotamers that can provide the supporting hydrogen bond. The scaffold set spans a broad range of protein folds, including TIM barrels, β-propellers, jelly rolls, Rossman folds and lipocalins, amongst others (Supplementary Table 3). In a typical search, more than 100,000 possible realizations of the input idealized active site were found in the scaffold set. For each of these 'matches', gradient-based minimization[16] was used to optimize the rigid body orientation of the transition state and the torsional degrees of freedom of the catalytic side chains to best satisfy all catalytic geometrical constraints. Subsequently, residues surrounding the transition state were redesigned both to maximize the stability of the active site conformation and the affinity to the transition state and to maintain protein stability using the Rosetta design methodology for proteins[17] and small molecules[18]. Designs were screened for compatibility with substrate and product and were ranked on the basis of the catalytic geometry and the computed transition-state-binding energy.

A steady enrichment of the fraction of designs in the TIM barrel scaffold was observed throughout the enzyme design process. TIM barrel scaffolds represent 25% of the proteins in the input scaffold set, 43% of the initial matches, and 71% of the low-energy designs. Inspection of the designs suggests that the binding pockets in TIM barrel scaffolds were favoured because of the large number of take-off positions (all positions around the barrel pointing towards the cavity) for both the catalytic residues and the additional transition-state-binding residues optimized in the design process; the former favoured TIM barrel matches, and the latter favoured low-energy designs in TIM scaffolds. The TIM barrel is the most widespread and catalytically diverse fold in naturally occurring enzymes; our *in silico* design process seems to be drawn towards the same structural features as naturally occurring enzyme evolution.

Experimental characterization

Following the active site design, a total of 59 designs in 17 different scaffolds were selected for experimental characterization. Out of the 59 designs, 39 use an Asp or Glu as the generalized base and 20 use a His–Asp or His–Glu dyad. Eight of the designs showed measurable activity in Kemp elimination assays in an initial activity screen (Table 1; see Supplementary Table 4 for sequence information and Methods for experimental details). For each of these eight designs, mutation of the catalytic base (to Ala or Gln/Asn) markedly decreased the activity or abolished catalysis completely, suggesting that the observed activity results from the designed active site (Table 1; for some examples, see Fig. 2a). The designs have k_{cat}/K_m values in the range of 6 to 160 $M^{-1} s^{-1}$ (Table 1 and Fig. 2b); it was not possible to obtain saturation kinetics in all cases (for example, see KE10 (open squares) and KE61 (open triangles) in Fig. 2b) owing to low substrate solubility. Both catalytic motifs were used in active designs; of the two most active catalysts, which show a rate acceleration of roughly 10^5 and a k_{cat}/K_m of about 100, one uses the Glu as the base and the other uses the His–Asp dyad. All designs exhibited multiple turnovers (≥7)—a prerequisite for efficient catalysis.

Models for these two most active designs are shown in Fig. 3. In the KE59 design (Fig. 3a), which is in a TIM barrel scaffold, Glu 231 is the catalytic base and Trp 110 facilitates charge delocalization by π-stacking to the transition state. Additionally, Leu 108, Ile 133, Ile 178, Val 159 and Ala 210 create a tightly packed hydrophobic pocket that envelops the non-polar substrate. The polar residues Ser 180 and Ser 211 provide hydrogen-bonding interactions with the nitro group of the transition state. Mutation of the catalytic base Glu 231 to Gln abolished catalytic activity (Table 1 and Fig. 2a, open triangles). Attempts to add a hydrogen bond donor to stabilize the negative charge developing at the phenolic oxygen through a Gly 131

Table 1 | Kinetic parameters of designed enzymes

Design	PDB code of template	Base	Hydrogen-bond donor	π-stack	k_{cat} (s^{-1}; mean ± s.d.)	K_m (mM)	k_{cat}/K_m ($M^{-1} s^{-1}$)	k_{cat}/k_{uncat}*
KE07	1thf	E101	K222	W50	0.018 ± 0.001	1.4 ± 0.1	12.2 ± 0.1	1.6×10^4
KE07	1thf	E101A	K222	W50	0.0009 ± 0.0004	0.29 ± 0.14	3.5 ± 1.0	7.8×10^2
KE07	1thf	E101	K222A	W50	0.030 ± 0.004	1.3 ± 0.2	22.7 ± 2.7	2.6×10^4
KE10	1a53	E178	None	W210	NA	NA	51.6 ± 4.0	NA
KE10	1a53	E178Q	None	W210	NA	NA	0.17 ± 0.17	NA
KE15	1thf	D48	None	Y126	0.022 ± 0.003	0.63 ± 0.09	35.1 ± 4.8	1.9×10^4
KE15	1thf	D48A	None	Y126	NA	NA	0.05 ± 0.07	NA
KE16	1thf	D48	K201	Y126	0.006 ± 0.001	4.2 ± 0.1	13.4 ± 1.4	5.2×10^3
KE16	1thf	D48A	K201	Y126	ND	ND	ND	ND
KE59	1a53	E231	None	W110	0.29 ± 0.11	1.8 ± 0.6	163 ± 21	2.5×10^5
KE59	1a53	E231Q	None	W110	NA	NA	0.003 ± 0.002	NA
KE59	1a53	E231	G131S	W110	NA	NA	17.9 ± 0.7	NA
KE61	1h61	E100	None	W184	NA	NA	7.9 ± 0.6	NA
KE61	1h61	E100A	None	W184	NA	NA	0.7 ± 0.1	NA
KE70	1jcl	H16–D44	S137	Y47	0.16 ± 0.05	2.1 ± 0.8	78.3 ± 13.7	1.4×10^5
KE70	1jcl	H16A–D44	S137	Y47	NA	NA	0.02 ± 0.03	NA
KE70	1jcl	H16–D44N	S137	Y47	0.06 ± 0.03	2.3 ± 1.5	29.8 ± 4.6	5.2×10^4
KE71	1a53	H89–D85	Y210	W184	NA	NA	5.9 ± 0.2	NA
KE71	1a53	H89A–D85	Y210	W184	NA	NA	0.4 ± 0.2	NA
KE71	1a53	H89–D85N	Y210	W184	NA	NA	1.0 ± 0.2	NA

* k_{uncat} (1.16×10^{-6} s^{-1}) was determined in HEPES buffer at pH 7.25, and extrapolated to zero buffer concentration. The means and standard deviations of the kinetic parameters were calculated from at least three independent measurements. NA, not applicable; ND, not determined.

©2008 Nature Publishing Group

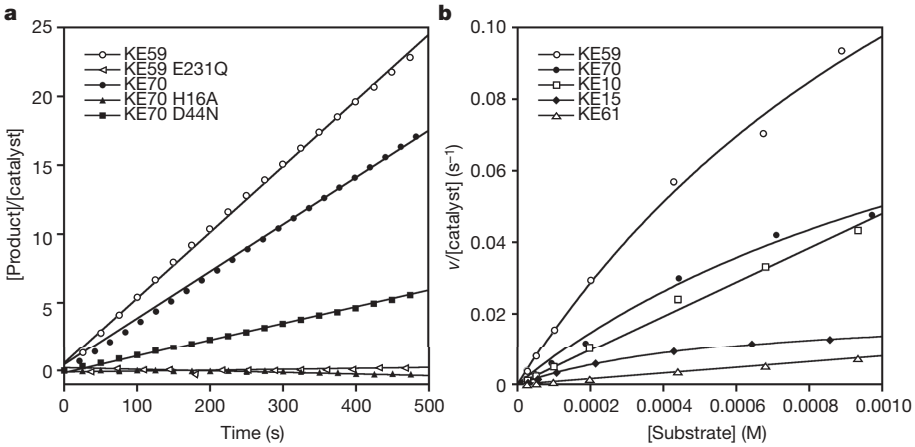


Figure 2 | Kinetic characterization of designed catalysts. **a**, Catalytic activity was measured by monitoring the product formation over time for KE59 (open circles) and KE70 (filled circles) at 400 μM substrate concentration. The *y* axis is the product concentration divided by the catalyst concentration that corresponds to the number of substrate turnovers. Deleting the catalytic base in both designs largely eliminates catalytic activity (open and filled triangles). Mutating Asp 44 of the catalytic dyad of KE70 to Asn (filled squares) causes a 2.5-fold reduction in activity. **b**, Michaelis–Menten plots for a representative selection of designed catalysts. The reaction velocity *v* divided by catalyst concentration is plotted on the *y* axis and the substrate concentration on the *x* axis. Some designs (for example, KE10 (open squares) and KE61 (open triangles)) show no saturation up to the maximal substrate solubility.

to Ser mutation caused a ninefold reduction in k_{cat}/K_m (Table 1), perhaps owing to unfavourable electrostatic interactions between the oxygen atoms on the serine and substrate; this large effect suggests that the transition-state-binding site is quite well defined. The aromatic-rich pocket and carboxylate base are reminiscent of the active site of the Kemp catalytic antibody 34E4 (ref. 10).

The KE70 design (Fig. 3b) uses the His–Asp dyad mechanism. Asp 44 positions and polarizes His 16 to optimally deprotonate the substrate. Tyr 47 π-stacks above the transition state, and together with Ile 201, Ile 139, Val 167, Ala 18, Ala 102 and Trp 71 creates a tight hydrophobic pocket around the transition state. The active site is again in a TIM barrel scaffold with the His–Asp dyad near the bottom of the site. Mutation of the catalytic base His 16 to Ala abolished catalytic activity (Table 1 and Fig. 2a, filled triangles), whereas mutating Asp 44 of the catalytic dyad to Asn produced an approximately 2.5-fold reduction (Table 1 and Fig. 2a, filled squares). In another design using a His–Asp dyad as general base (KE71), the analogous Asp-to-Asn mutation reduced activity sixfold (Table 1) whereas the His-to-Ala mutation abolished catalysis (Table 1).

High-resolution structural information on designed proteins is essential to validate the accuracy of the design methodology. We were able to grow crystals and obtain a high-resolution structure of one of the early Glu-based designs, KE07 (see Supplementary Information for details). As shown in Fig. 4, the crystal structure and design model are virtually superimposable, with an active site (6.0 Å around the transition state) root mean squared deviation (r.m.s.d.) of 0.95 Å mostly reflecting modest side-chain rearrangements. The similarity between the design model and the crystal structure suggests that the active sites in our new enzymes resemble those in the corresponding design models. The subtle deviations in the backbone indicate loop regions in which explicitly modelling backbone flexibility may yield improved designs.

The crystal structure also revealed that Lys 222 makes a salt bridge to the catalytic Glu 101 in the absence of substrate, whereas in the

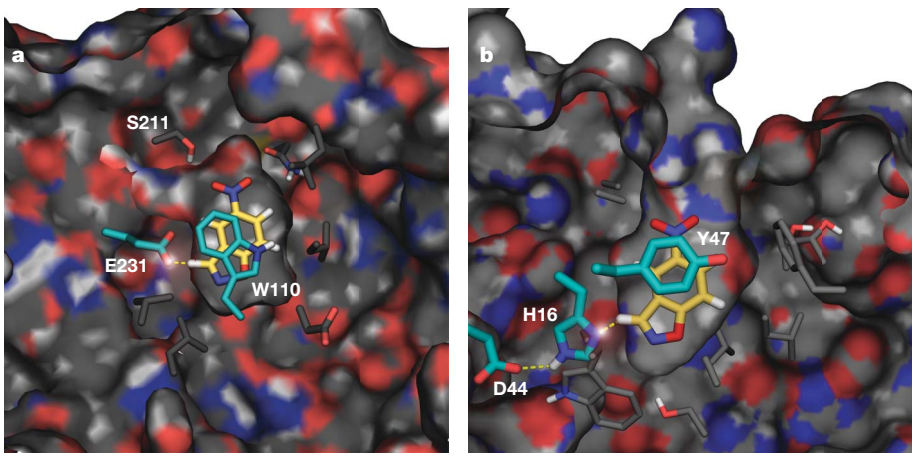


Figure 3 | Computational design models of the two most active catalysts. **a**, KE59 uses indole-3-glycerolphosphate synthase from *Sulfolobus solfataricus* as a scaffold. The transition state model is almost completely buried, with loops covering the active site. The mostly hydrophobic residues in the active site pocket pack the transition state model tightly, providing high shape complementarity (shape complementarity = 0.84; ref. 29). The polar residue Ser 211 interacts with the nitro group of the transition state to promote binding. The key catalytic residues (Glu 231 and Trp 110) are depicted in cyan. **b**, The deoxyribose-phosphate aldolase from *E. coli* is the scaffold for KE70. The shorter loops leave the active-site pocket freely accessible for the substrate. The transition state is surrounded by hydrophobic residues that provide high shape complementarity (shape complementarity = 0.77; ref. 29). His 16 and Asp 44 (in cyan) constitute the catalytic dyad whereas Tyr 47 (in cyan) provides π-stacking interactions.

©2008 Nature Publishing Group

designed model the ammonium of the lysine stabilizes the developing phenoxide in the transition state. Forming the productive transition state complex thus requires breaking of the salt bridge, and therefore elimination of the salt bridge in the unbound state would be expected to improve catalysis. We tested this prediction by substituting the lysine with an alanine, and this resulted in a 2.5-fold increase in k_{cat}/K_m (Table 1).

Directed evolution

In vitro evolution has been shown to markedly improve the stability, expression and activity of enzymes, and is currently the most widely used and successful approach for refining biocatalysts[19]. However, *in vitro* enzyme evolution generally requires a starting point with at least a low level of the desired activity, which is then optimized by repeated rounds of mutation and selection (for a notable exception, see ref. 20). We reasoned that *in vitro* evolution would be an excellent complement to our computational design efforts. The design calculations ensure that key catalytic functional groups are correctly positioned around the transition state, and, as demonstrated above, can generate active catalysts without requiring any starting activity. Thus, computational design can potentially provide excellent starting points for *in vitro* evolution. In contrast, the design process does not explicitly model configurational entropy changes, longer range second-shell interactions, and dynamics effects that can be important for efficient turnover; these shortcomings can potentially be remedied by directed evolution. Directed evolution can be valuable both in improving the designed catalysts and in stimulating improvements in the computational design methodology by shedding light on what is missing from the designs.

To investigate the extent to which *in vitro* evolution methods can improve computationally designed enzymes, we initiated evolution experiments on KE07—the early design for which the crystal structure was determined. Seven rounds of random mutagenesis and shuffling (also including synthetic oligonucleotides that expanded the diversity at selected residues), followed by screens in microtitre plates, yielded variants that had 4–8 mutations relative to KE07 and an improvement of >200-fold in k_{cat}/K_m (Table 2). Notably, the key aspects of the computational design, including the identities of the catalytic side chains, were not altered by the evolutionary process (indeed, mutating the catalytic base Glu 101 abolished the catalytic activity of both the designed template and its evolved variants; Table 2). Instead, the mutations were often seen in residues adjacent to designed positions (for example, Val 12, Ile 102, Gly 202), and thus provide subtle fine-tuning of the designed enzyme. Some mutations, such as Gly202Arg, are likely to increase the flexibility of regions neighbouring the active site. The hydrophobic residues Ile 7 and Ile 199 at the bottom of the active site were frequently mutated to polar or charged residues (the most common mutation being Ile7Asp), which may hold Lys 222 in position to stabilize the developing negative charge in the transition state while preventing interaction of Lys 222 with Glu 101. Consistent with this idea, the pK_a of the catalytic Glu 101 shifts from <4.5 to 5.9 in the evolved variant with the Ile7Asp mutation (for details, see Supplementary Information). Although the Lys222Ala mutation increases the activity of the original KE07, it significantly decreases the activity of the evolved variants, perhaps owing to the uncompensated additional negative charge.

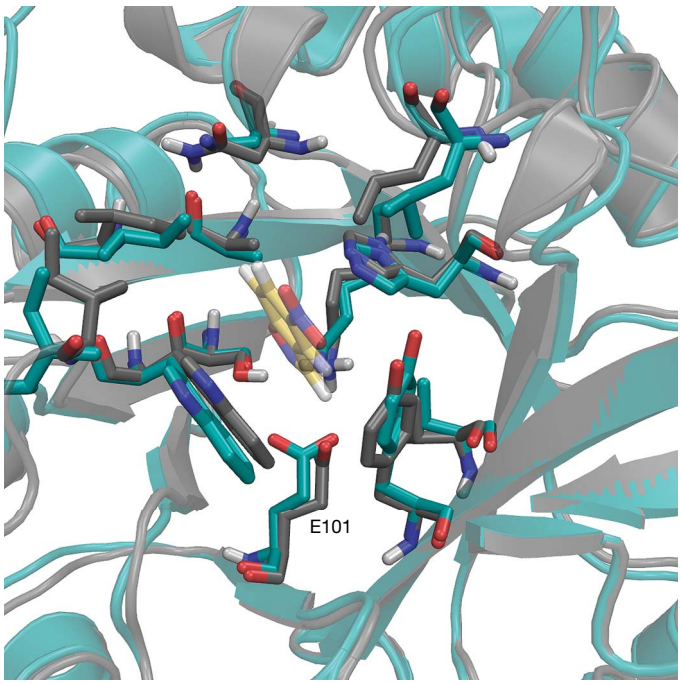


Figure 4 | Comparison of the designed model of KE07 and the crystal structure. The crystal structure (cyan) was solved in the unbound state and shows only modest rearrangement of active site side chains compared to the designed structure (grey) modelled in the presence of the transition state (yellow, transparent). (Backbone r.m.s.d. for the active site is 0.32 Å versus 0.95 Å for the active site including the side chains.) The observed electron density around relevant amino acids in the active site is shown in Supplementary Fig. 6. KE07 contains 13 mutations compared to the starting template scaffold (PDB code 1thf).

Table 2 | Kinetic parameters of KE07 variants

Variant	Mutations	k_{cat} (s^{-1})	K_m, (mM)	k_{cat}/K_m ($M^{-1}s^{-1}$)	k_{cat}/k_{uncat}*
KE07 WT	-	0.018 ± 0.001	1.4 ± 0.1	12.2 ± 0.1	1.55×10^4
R2 11/10D†	K19E Q123R K146T G202R N224D	0.021 ± 0.001	0.31 ± 0.02	66 ± 2	8.38×10^4
R3 I3/10A	I7Q F86L K146T G202R N224D F229S	0.206 ± 0.003	0.48 ± 0.03	425 ± 16	1.78×10^5
R3 I3/10A E101A				≤3.9	
R4 1E/11H	I7D K146E G202R N224D	0.699 ± 0.001	2.40 ± 0.07	291 ± 9	6.02×10^5
R4 1E/11H E101A				≤2.4	
R5 10/3B	I7D V12M G202R N224D	0.49 ± 0.01	0.59 ± 0.03	836 ± 18	4.22×10^5
R6 3/7F	I7D K19E K146T G202R N224D	0.60 ± 0.07	0.69 ± 0.09	872 ± 25	5.17×10^5
R7 2/5B	I7D F77I G202R N224D	1.20 ± 0.08	0.86 ± 0.08	1,388 ± 44	1.03×10^6
R7 10/11G	I7D V12M F77I I102F K146T G202R N224D F229S	1.37 ± 0.14	0.54 ± 0.12	2,590 ± 302	1.18×10^6

* k_{uncat} (1.16×10^{-6} s^{-1}) was determined in HEPES buffer at pH 7.25, and extrapolated to zero buffer concentration. The means and standard deviations of the kinetic parameters were calculated from at least three independent measurements.
† The nomenclature of the evolved variants states the directed evolution round (R2, R3, and so on), the microtitre plate number and the location in the microtitre plate.

©2008 Nature Publishing Group

Conclusions

The marked increase in catalytic activity and in turnover (>1,000 catalytic cycles were observed for the evolved variants), achieved through screening a relatively small number of variants (800–1,600 clones per round) by molecular evolution standards bodes well for future combinations of computational design and molecular evolution. In particular, the *in vitro* evolution of the most active of the computational designs, for example, KE59 or KE70, has the potential to yield highly active catalysts for the Kemp elimination reaction. We anticipate the successful use of the combination of computational design and molecular evolution that we have described here for a wide range of important reactions in the years to come.

The challenge of generating new biocatalysts has led to several successful experimental strategies[20–22]. In particular, the Kemp elimination comprises a well-defined model for catalysis of proton transfer from carbon—a highly demanding reaction and a rate-determining step in numerous enzymes. It has therefore been the subject of several attempts to generate enzyme-mimics and models (such as catalytic antibodies[23], promiscuous protein catalysts[24] and enzyme-like polymers[14]). The catalytic parameters of the new enzymes described here are comparable to the most active catalysts of the Kemp elimination of 5-nitro-benzisoxazole described thus far, and provide further insights into the makings of an enzyme. Comparison with the catalytic antibodies[23] highlights the major shortcoming of many of the designs noted above—that is, their relatively weak binding of the substrate. Although the computational design methodology has the advantage of being able to explicitly place key catalytic residues, this may come at a cost of overall substrate and transition-state binding affinity. Consistently achieving high affinity to the transition state and high turnover numbers is a challenge that we are currently approaching by introducing scaffold backbone flexibility into the design process. This should enable us to create higher affinity binding sites formed by more precisely positioned constellations of binding and catalytic residues.

The computational methodology described here can be readily generalized to design catalysts for more complex multistep reactions[25]. The combination of computational enzyme design to create the overall active site framework for catalysing a synthetic chemical reaction with molecular evolution to fine-tune and incorporate subtleties not yet modelled in the design methodology is a powerful route to create new enzyme catalysts for the very wide range of chemical reactions for which naturally occurring enzymes do not exist. Equally importantly, computational design provides a critical testing ground for evaluating and refining our understanding of how enzymes work.

METHODS SUMMARY

Computational design. Transition state geometries were computed at the B3LYP/6-31G(d) level for idealized active sites containing either a carboxylate or an imidazole-carboxylate dyad as the general base. Aromatic side chains were placed above and below the transition state using idealized π-stacking geometries[15]. A six-dimensional hashing procedure[4] was applied to find transition state placements in a large set of protein scaffolds (Supplementary Table 3) that were consistent with the catalytic geometry. Residues surrounding the catalytic side chains and transition state were repacked and redesigned[17,18] to optimize steric, coulombic and hydrogen-bonding interactions with the transition state and associated catalytic residues.

Experimental characterization. The proteins were expressed in *Escherichia coli* BL21(DE3) using pET29b (Novagen) and purified over a Ni-NTA column (Qiagen). The proteins (1 μM to 10 μM) were assayed in 25 mM HEPES (pH 7.25) and 100 mM NaCl at 250 μM substrate concentration for the initial screening, and substrate dilutions from 1 mM to 11 μM were used for kinetic characterization. Kinetic parameters were determined in at least three independent measurements. Fitted K_m values above 1 mM (and their corresponding k_{cat} values) are necessarily approximate. Site-directed mutagenesis of catalytic residues and independent protein purifications by different protocols/laboratories were carried out to exclude possible contaminating enzymes (Supplementary Information).

***In vitro* evolution.** Gene libraries of KE07 were created by random mutagenesis using error-prone PCR with 'wobble' base analogues dPTP and 8-oxo-dGTP[26] using the Genemorph PCR mutagenesis kit (Stratagene), and by DNA shuffling of the most active variants[27]. In certain rounds, shuffling included the spiking of synthetic oligonucleotides that expanded the diversity at selected residues[28]. In each round, the cleared lysates of 800–1,600 individual colonies were assayed for hydrolysis of 5-nitrobenzisoxazole (0.125 mM) by following product formation at 380 nm. The most active clones were sub-cloned and sequenced, and the encoding plasmids were used as templates for subsequent rounds of mutagenesis and screening.

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

Received 25 October 2007; accepted 3 March 2008.
Published online 19 March 2008.

1. Radzicka, A. & Wolfenden, R. A proficient enzyme. *Science* **267,** 90–93 (1995).
2. Bolon, D. N. & Mayo, S. L. Enzyme-like proteins by computational design. *Proc. Natl Acad. Sci. USA* **98,** 14274–14279 (2001).
3. Kaplan, J. & DeGrado, W. F. *De novo* design of catalytic proteins. *Proc. Natl Acad. Sci. USA* **101,** 11566–11570 (2004).
4. Zanghellini, A. *et al.* New algorithms and an *in silico* benchmark for computational enzyme design. *Protein Sci.* **15,** 2785–2794 (2006).
5. Casey, M. L., Kemp, D. S., Paul, K. G. & Cox, D. D. The physical organic chemistry of benzisoxazoles I. The mechanism of the base-catalyzed decomposition of benzisoxazoles. *J. Org. Chem.* **38,** 2294–2301 (1973).
6. Kemp, D. S. & Casey, M. L. Physical organic chemistry of benzisoxazoles II. Linearity of the brønsted free energy relationship for the base-catalyzed decomposition of benzisoxazoles. *J. Am. Chem. Soc.* **95,** 6670–6680 (1973).
7. Hu, Y., Houk, K. N., Kikuchi, K., Hotta, K. & Hilvert, D. Nonspecific medium effects versus specific group positioning in the antibody and albumin catalysis of the base-promoted ring-opening reactions of benzisoxazoles. *J. Am. Chem. Soc.* **126,** 8197–8205 (2004).
8. Hollfelder, F., Kirby, A. J., Tawfik, D. S., Kikuchi, K. & Hilvert, D. Characterization of proton-transfer catalysis by serum albumins. *J. Am. Chem. Soc.* **122,** 1022–1029 (2000).
9. Na, J., Houk, K. N. & Hilvert, D. Transition state of the base-promoted ring-opening of isoxazoles. Theoretical prediction of catalytic functionalities and design of haptens for antibody production. *J. Am. Chem. Soc.* **118,** 6462–6471 (1996).
10. Debler, E. W. *et al.* Structural origins of efficient proton abstraction from carbon by a catalytic antibody. *Proc. Natl Acad. Sci. USA* **102,** 4984–4989 (2005).
11. Lee, C., Yang, W. & Parr, R. G. Development of the Colle–Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter* **37,** 785–789 (1988).
12. Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38,** 3098–3100 (1988).
13. Frisch, M. J. *et al. Gaussian 03, revision C. 02* (Gaussian, Inc., Wallingford, Connecticut, 2004).
14. Hollfelder, F., Kirby, A. J. & Tawfik, D. S. Efficient catalysis of proton transfer by synzymes. *J. Am. Chem. Soc.* **119,** 9578–9579 (1997).
15. Misura, K. M., Morozov, A. V. & Baker, D. Analysis of anisotropic side-chain packing in proteins and application to high-resolution structure prediction. *J. Mol. Biol.* **342,** 651–664 (2004).
16. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C++* 2nd edn (Cambridge Univ. Press, Cambridge, UK, 2002).
17. Kuhlman, B. *et al.* Design of a novel globular protein fold with atomic-level accuracy. *Science* **302,** 1364–1368 (2003).
18. Meiler, J. & Baker, D. ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. *Proteins* **65,** 538–548 (2006).
19. Chica, R. A., Doucet, N. & Pelletier, J. N. Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* **16,** 378–384 (2005).
20. Seelig, B. & Szostak, J. W. Selection and evolution of enzymes from a partially randomized non-catalytic scaffold. *Nature* **448,** 828–831 (2007).
21. Cesaro-Tadic, S. *et al.* Turnover-based *in vitro* selection and evolution of biocatalysts from a fully synthetic antibody library. *Nature Biotechnol.* **21,** 679–685 (2003).
22. Varadarajan, N., Gam, J., Olsen, M. J., Georgiou, G. & Iverson, B. L. Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc. Natl Acad. Sci. USA* **102,** 6855–6860 (2005).
23. Thorn, S. N., Daniels, R. G., Auditor, M. T. & Hilvert, D. Large rate accelerations in antibody catalysis by strategic use of haptenic charge. *Nature* **373,** 228–230 (1995).
24. Hollfelder, F., Kirby, A. J. & Tawfik, D. S. Off-the-shelf proteins that rival tailor-made antibodies as catalysts. *Nature* **383,** 60–62 (1996).
25. Jiang, L. *et al. De novo* computational design of retro-aldol enzymes. *Science* **319,** 1387–1391 (2008).
26. Vartanian, J. P., Henry, M. & Wain-Hobson, S. Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. *Nucleic Acids Res.* **24,** 2627–2631 (1996).
27. Abecassis, V., Pompon, D. & Truan, G. High efficiency family shuffling based on multi-step PCR and *in vivo* DNA recombination in yeast: statistical and functional

©2008 Nature Publishing Group

analysis of a combinatorial library between human cytochrome P450 1A1 and 1A2. *Nucleic Acids Res.* **28,** E88 (2000).
28. Herman, A. & Tawfik, D. S. Incorporating synthetic oligonucleotides via gene reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng. Des. Sel.* **20,** 219–226 (2007).
29. The. CCP4 suite: programs for protein crystallography. *Acta Crystallogr.* **50,** 760–763 (1994).

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

Acknowledgements We thank R. Stanfield and I. Wilson for providing D-2-deoxyribose-5-phosphate aldolase wild-type protein (PDB code 1jcl) and W. A. Greenberg and C.-H. Wong for providing the expression plasmid. We thank Rosetta@home participants for their contributions of computing power. This work was supported by a postdoctoral fellowship from the Swiss National Science Foundation to D.R., an Adams Fellowship (Israel Academy of Science) to O.K., research grants from the Minerva Foundation and the Fannie Sherr Estate to D.S.T., and NSF and NIH-CBI grants to K.N.H. We are also thankful for financial support from the Defense Advances Research Projects Agency (DARPA) and the Howard Hughes Medical Institute (HHMI) for this research.

Author Contributions D.R. performed computational design using carboxylate and the His–Asp motif, and purified and experimentally characterized designed catalysts; O.K. synthesized the substrate, performed *in vitro* evolution and experimentally characterized evolved variants; A.M.W. performed computational design on the His–Asp motif; L.J. performed initial computational design on the carboxylate motif; J.D. and K.N.H. computed idealized active sites using quantum mechanics; J.B. and J.L.G. expressed and purified designed catalysts; E.A.A. helped with enzyme design set-up; A.Z. wrote RosettaMatch and helped with computational set-up; O.D. and S.A. crystallized KE07; and D.R., A.M.W., D.B., K.N.H., O.K. and D.S.T. designed the experiment and wrote the manuscript.

Author Information The crystal structure of KE07 has been deposited in the RCSB Protein Data Bank (www.rcsb.org) under the accession number 2rkx. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.B. (dabaker@u.washington.edu) or D.S.T. (dan.tawfik@weizmann.ac.il).

©2008 Nature Publishing Group

METHODS

Quantum mechanical transition state calculation. Quantum mechanical calculations using density functional theory with the B3LYP functional and the 6-31G(d) basis set[11,12] were used to locate transition structures (confirmed by vibrational frequency analysis) for the acetate- and imidazole/acetate-catalysed reactions in the gas phase. Lysine, serine, threonine and tyrosine functional groups were included in the calculations as hydrogen bond donors to stabilize the developing negative charge on the phenolic oxygen of the transition state. All calculations were carried using Gaussian03 (ref. 13).

Aromatic side chains (Phe, Tyr and Trp) were also modelled to stabilize charge delocalization of the transition state and to provide favourable π-stacking interactions. These side chains were placed using idealized π-stacking geometries[15] in a parallel configuration (4 Å separation) with the aromatic centre offset from the transition state rings by 1 Å. The aromatic groups were placed above either the five- or the six-membered ring and were allowed on both the top and the bottom faces of the transition state. Full rotation about the normal to the aromatic plane was permitted, allowing for variable Cβ–Cγ bond vector placement. The optimal catalytic geometry and the associated constraints for both reaction mechanisms are shown in Supplementary Fig. 1.

Scaffold selection. A large set of protein scaffolds were chosen as candidates for transition state placement. The selection criteria for these scaffolds were as follows: that a high-resolution crystal structure is available; that expression in *E. coli* is possible; that they are stable proteins; that they contain a preexisting pocket; and that they span a variety of protein folds. The protein scaffolds used in this study are listed in Supplementary Table 3.

For each scaffold, a three-dimensional grid representing the pre-existing pocket was mapped out using an in-house pymol plugin (Supplementary Fig. 2). This was used to reduce the extremely large search space for transition state placement (see below). The positions of potential catalytic residues near the active site were then compiled for each scaffold. In addition, a three-dimensional grid representing the protein backbone was created for each scaffold to allow for a fast clash check.

Transition state placement. To find active site placements in the input scaffolds, it is necessary to consider many alternative geometries for each catalytic motif. As described below, by varying the precise orientations of the catalytic side chains relative to the transition state, we generated very large ensembles of active site geometries. For each of these active site geometric variants, RosettaMatch[4] was used to position simultaneously transition state and catalytic residues into the set of pre-selected protein scaffolds so as to satisfy all catalytic constraints without steric overlap (only scaffold backbone atoms were modelled for clashes). Supplementary Figs 3–5 show the geometric descriptors used for catalytic side chain–transition state placement and the corresponding number of alternative conformations to be sampled. The His-based mechanism is shown as an example. The Glu/Asp-based mechanism was diversified similarly.

The geometric parameters for the catalytic base–transition state interaction were sampled much more finely because the relative geometry of the general base was considered to be more important than π-stacking or negative charge stabilization. Using the geometric parameters specified in Supplementary Fig. 3, there were 77,472,288 histidine–transition state conformations per position, 52,488 serine–transition state conformations per position, and 27,216 π-stacking–transition state conformations per position.

For a typical matching run, such as the TIM barrel protein scaffold, histidines were sampled at 41 positions around the barrel, and serine and π-stacking residues were placed at 119 residues to allow for catalytic side chains at second-shell residues. For this example, there are more than 1.5×10^{21} possible combinations for creating the catalytic motif, which would be computationally intractable to enumerate. By using the linear-scaling RosettaMatch algorithm, this number was reduced to a much more manageable number (8.7×10^{7}). The use of three-dimensional grids allows for rapid pruning of this large number of transition state conformations, as described above.

For the catalytic mechanism using histidine as the base, we prefiltered each scaffold to identify pairs of positions at which histidine and aspartate/glutamate rotamers can be placed to achieve the dyad geometry. Rotamer pairs with a van der Waals repulsive energy less than 2.0 kcal mol^{-1} and hydrogen bonding energy less than -0.5 kcal mol^{-1} were stored in an 'interaction graph'. Matching was carried out using histidine as the catalytic residue, iterating only over histidine rotamers in the interaction graph of His–Asp and His–Glu pairs. For a given match, each Asp or Glu rotamer in the interaction graph that interact with the matched His rotamer was grafted onto the match, and the result screened to remove clashes between the transition state and the backing-up residue. Using the interaction graph decreases the number of potential histidine rotamers that must be modelled in the active site, and thus allows for even finer sampling of ligand rotamer sets. In the TIM barrel example described above, the number of histidine rotamers sampled at the 41 residue positions was decreased from 3,321 (81×41) to 253 by precalculating and filtering only the subset of histidine rotamers that can form hydrogen bonds to Asp/Glu. This reduces the number of histidine–transition state conformations from 7.7×10^{7} to 5.9×10^{6}.

Geometric filters were applied to remove matches unlikely to produce good designs. Matches for which transition state poses clashed with more than four modelled Cβ atoms were removed as they would require too many Gly mutations to be introduced to accommodate the bound pose, potentially destabilizing the folded state. Matches with an insufficient number of neighbouring residues around the transition state would be expected to lead to underpacking during the design stage and were also removed.

Protein design. Residues surrounding the transition state and catalytic residues were selected for redesign, and the Rosetta protein design methodology[17,18,30] was used to create a pocket with high affinity for the transition state. Residue selection was carried out using a shell-based method. Residues with Cβ atoms within 8 Å were redesigned, those within 10 Å for which the Cα–Cβ bond vector pointed towards the transition state were redesigned, and all other residues within 12 Å were repacked. A rigid-body minimization of the transition state as well as side-chain relaxation of the protein was performed for each designed model.

Design filtering. A geometric filter was applied to choose models for which catalytic geometry was consistent with the specified constraints (tables in Supplementary Figs 3–5). The van der Waals interaction energy for the transition state and catalytic residues was a useful filter for choosing designs that were roughly well packed; designs with a transition state–protein van der Waals energy greater than -5.0 kcal mol^{-1} were removed. Filters were used to select for high transition state–protein shape complementarity[29], and to choose models with minimal small cavities surrounding the transition state (W. Sheffler and D.B., submitted). Solvent accessibility measures were used to remove models that completely buried the transition state. For the His–Asp dyad mechanism, an additional filter was added, requiring that the His–Asp hydrogen bond remain on repacking of all residues in the presence of the transition state.

Protein expression and purification. Genes encoding the designs in the pET29b expression vector (Novagen) were purchased from Codon Devices, Inc. The catalytic-side-chain knockout mutations to Ala or Asn/Gln were introduced by site-directed mutagenesis as described[31]. After transformation into BL21 Star (Invitrogen), a one litre culture of auto-induction media[32] was inoculated with a single colony and shaken at 37 °C for 8 h. Expression was continued at 18 °C for 24 h. The cells were harvested, resuspended in 25 mM HEPES (pH 7.5) and 100 mM NaCl, and lysed by sonication. The soluble fraction was applied to a Ni-NTA column (Qiagen), washed with 20 mM imidazole, and the protein was eluted with 250 mM imidazole. The proteins were concentrated and the buffer was exchanged to 25 mM HEPES (pH 7.25) and 100 mM NaCl using a 5 ml HiTrap desalting column (GE Healthcare). For KE59, an additional size-exclusion chromatography step (Superdex75 10/300 GL from GE Healthcare) was performed. Protein concentrations were determined by measuring the absorbance at 280 nm using the calculated extinction coefficient[33]. To eliminate the possibility of observing the activity from a contaminating natural enzyme, further purification steps were carried out for KE07 and the evolved KE07 variants, for KE59 and for KE70 as described in the Supplementary Information section 10, validating the Kemp elimination activity of the designed and evolved enzymes.

Kinetic measurements. For the initial activity screen, 100 μl of the designed proteins (10 μM final concentration) were mixed with 100 μl of 500 μM substrate (freshly diluted from a 50 mM stock solution in acetonitrile) in 25 mM HEPES (pH 7.25) and 100 mM NaCl in a 96-well plate. For the kinetic characterization, the reactions were started by adding 150 μl of substrate dilutions (1 mM to 11 μM final concentration) in 25 mM HEPES (pH 7.25), 100 mM NaCl and 2% acetonitrile to 50 μl of protein (1 μM to 10 μM final concentration) in 25 mM HEPES (pH 7.25) and 100 mM NaCl (or no protein for the background reaction) in a 96-well plate. Product formation was followed at 380 nm in a SpectraMax M5e (Molecular Devices) plate reader at 27 °C in at least three independent experiments. The initial rates divided by the catalyst concentration were plotted against substrate concentration, and k_{cat} and K_m were determined by fitting the data to the Michaelis–Menten equation (equation (1)) using Kaleidagraph.

$$v/[\text{catalyst}] = k_{cat} \bullet [\text{substrate}]/(K_m + [\text{substrate}]) \quad (1)$$

If saturation kinetics were not observed, k_{cat}/K_m values were calculated from a linear fit from the data.

Screening procedure. The libraries were screened by growing the cultures of *E. coli* BL21 cells in 96-deep-well plates and checking the activity of the lysates with 5-nitrobenzisoxazole. In brief, *E. coli* BL21 cells transformed with the libraries were grown on luria broth (LB) agar plates (containing 100 μg ml^{-1}

©2008 Nature Publishing Group

kanamycin). Individual colonies were inoculated into 2YT supplemented with 50 μg ml^{-1} kanamycin (300 μl) in 96-deep-well plates, and grown for ~15 hours at 37 °C. Overnight cultures (20 μl) were inoculated into 2YT supplemented with 50 μg ml^{-1} kanamycin (500 μl) in 96-deep-well plates and grown to $A_{600\,nm}$ of ~0.6. Overexpression was induced by adding 1 mM isopropyl-β-D-thiogalactoside (IPTG), and the cultures were grown for another 5 h, centrifugated, and the pellet frozen overnight at −20 °C. The cells were lysed with lysis buffer, 250 μl well^{-1} (50 mM HEPES (pH 7.25), 0.2% Triton, 0.1 mg ml^{-1} lysozyme), and the lysates were cleared by centrifugation and assayed for hydrolysis of 5-nitrobenzisoxazole (0.125 mM) by following the release of the phenol product at 380 nm (Power HT microtitre scanning spectrophotometer). Overnight cultures of the most active clones were plated on LB agar plates containing 100 μg ml^{-1} kanamycin. To ensure monoclonality, and to verify the activity of the selected variants, the hydrolysis rates were re-assayed after growing two sub-clones from each original colony in the same conditions. Plasmids were extracted and used for sequencing and as templates for subsequent mutagenesis and screening rounds. Variants subjected to detailed analysis were re-transformed into *E. coli* BL21 cells, and the protein overexpressed and purified as described above.

Round 1. First-generation libraries were constructed from the designed KE07 gene by an error-prone PCR method using the 'wobble' base analogues dPTP and 8-oxo-dGTP[26]. The rate of mutations was 5 ± 3 per gene, and mutations were mainly of the transition type. The first round of KE07 evolution yielded active variants with lysate activity up to fivefold higher than of that of the starting point KE07.

Round 2. The 23 most active variants isolated in the first round of screening were subjected to DNA shuffling in the presence of the designed template (20%)[27] to yield second-generation libraries. The most active variants of round 2 had lysate activities up to 15-fold higher than that of the KE07. Analysis by SDS–PAGE demonstrated that the improvements in the activity were partially caused by enhancing the expression of KE07. Four active variants from round 2 were purified, and their kinetic parameters determined (Supplementary Table 1). Several dominant mutations in round 2 clones were identified; these can be divided into three groups: 1) Lys19Glu/Thr or Lys146Glu/Thr—mutations on the surface of the protein that seem to increase the expression levels of KE07. 2) Gly202Arg or Asn224Asp—mutations at the active site, probably interacting with the substrate-binding residues. Two other mutations in the helix 223–233 (Val226Ala and Phe229Ser), which are adjacent to Asp 224, were obtained. 3) Ile7Thr or Ile199Thr—residues located at the bottom of the active site, but not in direct contact with the substrate.

Round 3. The third-generation libraries were created by shuffling the four active variants of round 2 while randomizing various positions by incorporating spiking oligonucleotides during assembly of DNA fragments[28]:

Library 1: positions Ile 7 and Ile 199 were randomized (to Ile, Thr, Val, Ala, Phe, Ser, Glu, Asp, Gln, His), with the aim of finding the optimal combination of these residues at the bottom of the active site.

Library 2: positions Tyr 128 and His 201 were randomized (His 201 to Cys, Ser, Tyr, Ser, Thr, Asn; Tyr 128 to Leu, Pro, Ile, Thr, Val, Ala, Phe, Ser) to probe other residues at these designed positions that are responsible for benzisoxazole ring stacking.

Library 3: one or two amino acids were inserted between residues 224–225 and 225–226 to probe the variations of the helix 223–233, which seemed to be a target of many round 2 mutations.

Library 1 yielded clones with lysate activity up to 70-fold higher than that of KE07. Libraries 2 and 3 did not yield any improved variants, thus demonstrating that the designed stacking residues His 201 and Tyr 128 are at their optimal configurations, and that the length of the helix 223–233 does not need to be further optimized.

Round 4. At round 4, randomization of Ile 199 was continued because it was not changed in most of the clones of round 3. Positions Ile 173 and Leu 176 were randomized as well (to Ala, Asp, Glu, Val, Leu, Ile, Thr, Asn, Lys, Pro, His and Gln) because these residues interact with Gly 202, which in most of the improved variants was mutated into arginine.

Round 4 yielded active variants with crude lysate activities up to 200-fold higher than that of KE07. The most active variants of rounds 3 and 4 were purified, and their catalytic parameters determined (Supplementary Table 1).

Sequencing of round 3 and 4 variants confirmed the importance of the mutations found in round 2. Lys19Glu/Thr and Lys146Glu/Thr mutations increased the expression levels, and Gly202Arg and Asn224Asp optimized the top part of the active site. Randomization of positions Ile 7 and Ile 199 at the bottom of the active site demonstrated that, in the optimal combination, Ile 7 is changed to a more polar residue and Ile 199 remains intact. In several improved variants, the residues Ile 173 and Leu 176 were mutated as well, but their effect is relatively minor.

Because the mutation Asn224Asp was found in all the improved variants of rounds 3 and 4 (with the exception of R4 2F/2G), we wanted to ensure that this mutation did not alter the initial design, by acting, for example, as a general base, thus replacing the designed base Glu 101. Thus, we created Glu101Ala mutants of the variants R3 I3/10A, R4 1E/11H and R4 2F/2G, and of the KE07. Mutagenesis of Glu 101 caused a significant decrease in the activity of all the variants (up to 1%). These results demonstrated that the initial design, in which Glu 101 acts as a general base, was maintained (Supplementary Table 2).

Round 5. The active variants from round 4 were subjected to random mutagenesis by error-prone PCR with mutazyme (Genemorph PCR mutagenesis kit, Stratagene[34]) to yield the fifth-generation libraries, which contained 1 ± 1 mutations per gene and a large portion of shuffled genes. Mild lysate activity improvements (up to 1.5-fold) were observed, and the 12 most active variants from round 5 were subjected to another round of mutagenesis at a higher mutational load.

Round 6. At round 6, the 12 most active variants from round 5 were subjected to random mutagenesis by error-prone PCR with mutazyme (Genemorph PCR mutagenesis kit, Stratagene[34]) to yield the sixth-generation libraries, which contained 3 ± 1 mutations per gene and a large portion of shuffled genes. Lysate activity improvements of up to 1.5-fold were observed.

Round 7. Seventh-generation libraries were created by shuffling the 20 active variants of round 6, and lysate activity improvements of up to threefold were observed.

The xyz coordinates of the design KE07, KE59 and KE70 are available in the Supplementary Information.

30. Dantas, G., Kuhlman, B., Callender, D., Wong, M. & Baker, D. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J. Mol. Biol.* **332,** 449–460 (2003).
31. Kunkel, T. A., Roberts, J. D. & Zakour, R. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* **154,** 367–382 (1987).
32. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 (2005).
33. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4,** 2411–2423 (1995).
34. Barlow, M. & Hall, B. G. Predicting evolutionary potential: *in vitro* evolution accurately reproduces natural evolution of the tem β-lactamase. *Genetics* **160,** 823–832 (2002).

©2008 Nature Publishing Group

LETTERS

Semi-annual oscillations in Saturn's low-latitude stratospheric temperatures

Glenn S. Orton[1], Padma A. Yanamandra-Fisher[1], Brendan M. Fisher[2], A. James Friedson[1], Paul D. Parrish[3], Jesse F. Nelson[4], Amber Swenson Bauermeister[5], Leigh Fletcher[1], Daniel Y. Gezari[6], Frank Varosi[7], Alan T. Tokunaga[8], John Caldwell[9], Kevin H. Baines[2], Joseph L. Hora[10], Michael E. Ressler[11], Takuya Fujiyoshi[12], Tetsuharu Fuse[12], Hagop Hagopian[13], Terry Z. Martin[1], Jay T. Bergstralh[14], Carly Howett[15], William F. Hoffmann[16], Lynne K. Deutsch‡, Jeffrey E. Van Cleve[17], Eldar Noe[18], Joseph D. Adams[19], Marc Kassis[20] & Eric Tollestrup[21]

Observations of oscillations of temperature and wind in planetary atmospheres provide a means of generalizing models for atmospheric dynamics in a diverse set of planets in the Solar System and elsewhere. An equatorial oscillation similar to one in the Earth's atmosphere[1,2] has been discovered in Jupiter[3–6]. Here we report the existence of similar oscillations in Saturn's atmosphere, from an analysis of over two decades of spatially resolved observations of its 7.8-μm methane and 12.2-μm ethane stratospheric emissions, where we compare zonal-mean stratospheric brightness temperatures at planetographic latitudes of 3.6° and 15.5° in both the northern and the southern hemispheres. These results support the interpretation of vertical and meridional variability of temperatures in Saturn's stratosphere[7] as a manifestation of a wave phenomenon similar to that on the Earth and in Jupiter. The period of this oscillation is 14.8 ± 1.2 terrestrial years, roughly half of Saturn's year, suggesting the influence of seasonal forcing, as is the case with the Earth's semi-annual oscillation[1].

These conclusions are based on a sequence of filtered mid-infrared maps or images of Saturn, through narrow- to medium-band spectral filters that are sensitive to upwelling radiance emerging from Saturn's stratosphere. As in our study of Jupiter[6], we preferred to use the emission of stratospheric methane at wavelengths of around 7.8 μm to detect the stratospheric temperature field near the 20-mbar pressure level in the atmosphere, because methane is expected to be well mixed in Saturn's stratosphere. Thus, all variations in the thermal radiance must be attributed to variations in temperature, rather than in the methane abundance. However, because 7.8-μm methane emission is much fainter for Saturn than it is for Jupiter, most of our earliest observations with lengthy raster scans consist only of observations of much brighter stratospheric emission from ethane at wavelengths of around 12.2 μm (see the Supplementary Information), because only these images had sufficient signal-to-noise ratios to be useful. Figure 1 shows examples of 7.8-μm methane emission observed from the NASA Infrared Telescope Facility (IRTF) in two different phases of the oscillation. Details of the observations are given in the Supplementary Information.

The angular resolution of scans and images at the IRTF was limited by diffraction to no better than 0.7 arcsec (at latitude 4°) for 7.8-μm methane emission and 1.1 arcsec (at latitude 7°) for 12.2-μm ethane emission, with some additional blurring arising from seeing (that is, distortion due to terrestrial atmospheric turbulence). (Here and below, latitude values without an explicit attribution refer to either the northern or the southern hemisphere.) It is possible to resolve differences between emission at planetographic latitudes of 3.6° and 15.5° (planetocentric latitudes of 3.0° and 13.0°) in all the images used in this study, which is a requirement for this investigation. We ignored regions of the planet that were potentially obscured by the rings, defined for this purpose as extending from 74,500 km (the inner edge of the C ring) out to 154,000 km (the outer edge of the A ring) from the centre of the planet. We did not consider images that were sufficiently close to ring-plane crossing times that substantial ring obscuration of the atmosphere at latitude 3.6° might occur. Potential small obscurations were included in the uncertainties associated with the 1997 data (see the Supplementary Information). Brightness temperatures were averaged over ±15° of longitude around the central meridian, minimizing limb-brightening effects while maximizing the latitude extent of our sample. Figure 2 shows representative plots of zonally averaged brightness temperature versus latitude for methane and ethane for each year of observation.

In Figs 1 and 2 there is clear evidence of low-latitude oscillations between near-equatorial and higher latitudes, corresponding to the differences seen in the vertical structures observed by the Cassini Composite Infrared Spectrometer (CIRS)[7]. These differences are maximized by selecting brightness temperatures at latitude 3.6° (3.0° planetocentric) north or south and subtracting temperatures at latitude 15.5° (13.0° planetocentric) in the same hemisphere. Measurements of brightness temperatures at latitude 3.6° north or

[1]MS 169-237, [2]MS 183-601, Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA. [3]School of Geosciences, University of Edinburgh, Crew Building, The King's Buildings, West Mains Road, Edinburgh EH9 3JN, UK. [4]Department of Physics and Astronomy, University of Maine, 5709 Bennett Hall, Orono, Maine 04469, USA. [5]Astronomy Department, 601 Campbell Hall, University of California, Berkeley, California 94720-3411, USA. [6]Code 667, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. [7]Department of Astronomy, University of Florida, 211 Bryant Space Science Center, PO Box 112055, Gainesville, Florida 32611-2055, USA. [8]Institute for Astronomy, University of Hawaii, 2880 Woodlawn Drive, Honolulu, Hawaii 96822, USA. [9]Department of Physics and Astronomy, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada. [10]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [11]MS 79-5, Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA. [12]Subaru Telescope, National Astronomical Observatory of Japan, 650 North A'ohoku Place, Hilo, Hawaii 96720, USA. [13]Department of Computer Science, 4732 Boelter Hall, University of California, Los Angeles, California 90095, USA. [14]Langley Research Center, Hampton, Virginia 23681-2199, USA. [15]Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, Oxford University, Parks Road, Oxford, Oxfordshire OX1 3PU, UK. [16]Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721, USA. [17]Ball Aerospace & Technologies Corp., Civil Space Systems, PO Box 1062, Boulder, Colorado 80306, USA. [18]MS 183-501, Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, California 91109, USA. [19]Department of Radiophysics & Space Research, 206 Space Sciences Building, Ithaca, New York 14583, USA. [20]W. M. Keck Observatory, 62-1120 Mamalahoa Highway, Kamuela, Hawaii 96743, USA. [21]NASA Infrared Telescope Facility, Institute for Astronomy, 640 North A'ohoku Place, Hilo, Hawaii 96822, USA.
‡Deceased.

©2008 Nature Publishing Group

south sample almost the same vertical temperature structure as do measurements directly at the equator[7], but this selection provides a significantly larger number of samples that are not precluded by potential ring obscuration of the equator itself.

In Figure 3 we plot these differences as a function of time, which makes clear the cyclic nature of the temperature differences between these latitudes over the past 22 years. The variability of equatorial emission between Voyager 1 observations in 1980 and the 1989 IRTF observations[8] is consistent with this periodicity. Figure 3 also validates our first-order assumption that the oscillation is essentially equal in each hemisphere. Fitting the ethane emission data, which cover more than one cycle, results in a 'peak-to-peak' amplitude of just over 2 K in the 12.2-μm brightness temperature difference variation (8 K at 7.8 μm) and a period of 14.8 ± 1.2 yr.

We did not attempt to derive stratospheric kinetic temperatures using our measured brightness temperatures, unlike in the investigation of Jupiter's stratospheric temperature variability[6] with similar data, primarily because of the intractability of separating latitudinal variations of stratospheric temperatures from ethane abundances. However, the altitude and latitude variability of temperatures derived by the Cassini CIRS experiment[7] provides a sound basis for the latitudinal variability of the temperature and ethane vertical profiles in the 2005–2006 epoch. Figure 4 shows that the correlation of the spacecraft results with our 2006 observations is extremely good, if the Cassini CIRS results are appropriately blurred to match the average spatial resolution of the ground-based data. This implies that the amplitude of the actual variations is much higher in the atmosphere than is indicated in Fig. 3. Figure 4 also shows that the observed variability in brightness temperature difference between 2003 and 2007, which far exceeds the errors associated with the individual measurements, may arise simply from variations of atmospheric seeing conditions.

The ~15-yr period we derive is considerably longer than the two- to four-year cycles associated with the Earth's quasi-biennial

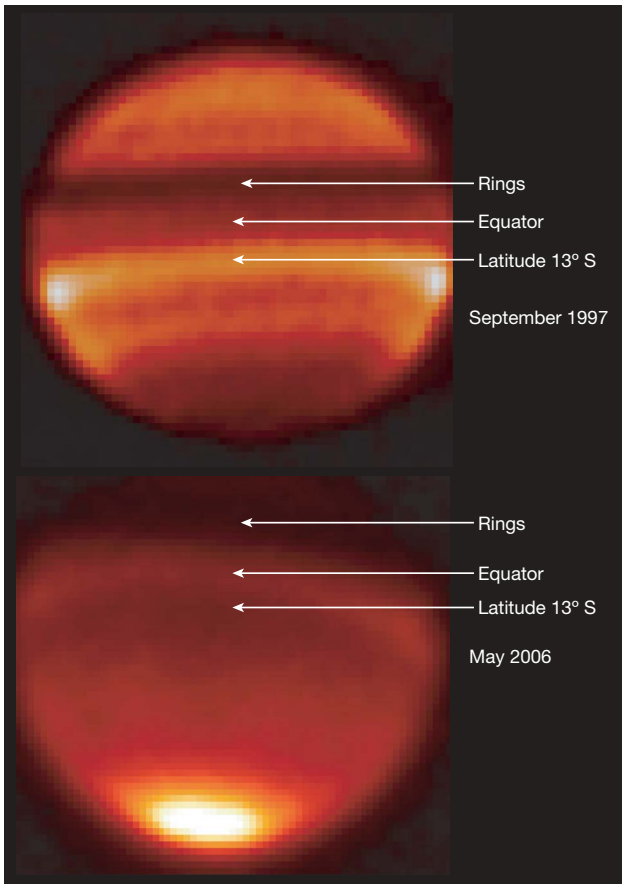


Figure 1 | Representative thermal images of Saturn showing two phases of the oscillation. Images showing 7.8-μm methane emission from Saturn's stratosphere taken at the IRTF, illustrating the alternating amplitudes of low-latitude temperatures in the 20-mbar altitude range between the equator and latitude 15° S. At this wavelength, emission from Saturn's rings is lower than that from the planet, and their primary effect is to obscure thermal emission from the atmosphere. These images demonstrate Saturn's appearance during a relative minimum of equatorial temperatures (1997) and during a relative maximum of equatorial temperatures (2006). A similar equatorial brightening was noted in comparing IRTF images from 1989 with low equatorial temperatures derived from the Voyager-1 Infrared Radiometer Interferometer Spectrometer experiment[8]. We note that the warm south pole in May 2006 is driven by seasonal variability[10].

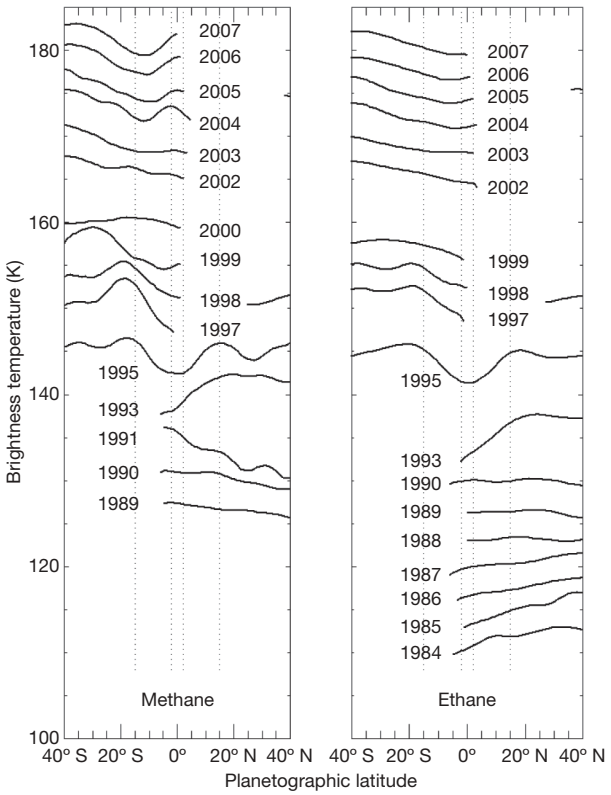


Figure 2 | Representative plots of zonally averaged brightness temperatures versus time. Methane emission at wavelengths of around 7.8 μm is shown in the left panel and ethane emission at wavelengths of around 12.2 μm is shown in the right panel. The methane brightness temperature plot is correct for 1989 and offset by +3 K per year thereafter. The ethane plot is correct for 1984 and offset by +3 K per year thereafter. The dotted lines denote latitudes of 3.6° and 15.5° north and south, for reference; these are the two reference latitudes used (per hemisphere) to plot the brightness temperature differences shown in Fig. 3. Most of the observations were made at the IRTF. Single-element IRTF facility photometers were used to create raster-scanned maps of Saturn between 1985 and 1990 in a way similar to stratospheric maps of Jupiter[6]. The first two-dimensional thermal infrared imaging camera used at the IRTF observed both methane and ethane emission from Saturn[8] in 1989; we included additional data from this camera taken in 1988, 1990 and 1991. Other two-dimensional cameras used to collect data in subsequent years include MIRAC[11], MIRLIN[12], and MIRSI[13]; these data include 1993 MIRAC observations from the 2.3-m Steward Observatory Bok Telescope. Until 2003, Saturn was observed only as a target of opportunity during observation runs primarily dedicated to providing ground-based support for the Galileo mission at Jupiter. Beginning in 2003, Saturn became a primary target, with a concomitant increase in observation time and a higher signal-to-noise ratio characterizing each image. We also included data from larger telescopes with improved diffraction-limited angular resolution: images of both methane and ethane emission taken using MIRLIN on the 10-m Keck I Telescope, an image of methane emission taken on the Keck II Telescope using the same facility's long-wavelength spectrometer[10], and an image of ethane emission taken on the 8.2-m Subaru Telescope.

©2008 Nature Publishing Group

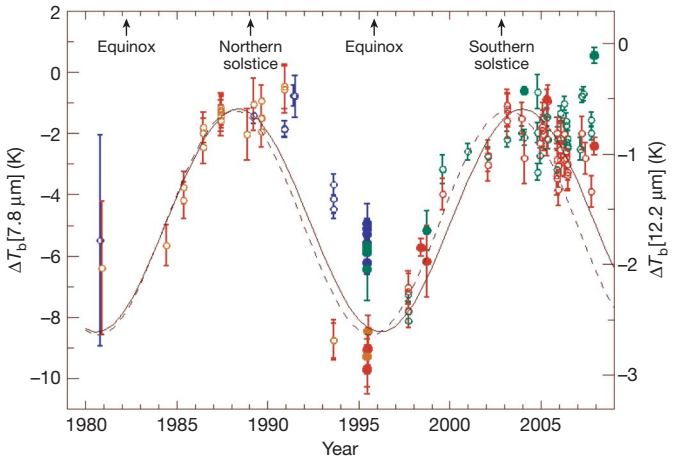


Figure 3 | Difference of zonally averaged brightness temperatures at two latitudes versus time. The planetographic latitudes are 3.6° and 15.5° north and south, and the plot is given for spectral regions representing methane emission (~7.8 μm) and ethane emission (~12.2 μm), showing one-s.d. uncertainties. Peak values represent brightness temperatures higher at latitude 3.6° than at latitude 15.5°. Open circles represent data taken at the IRTF or the Bok Telescope and filled circles represent data taken at Palomar Observatory (1995), at the W. M. Keck Observatory (1998, 2004) or using the Subaru Telescope (2005, 2007). Methane brightness temperature differences between 3.6° N and 15.5° N are denoted by blue symbols, and those between 3.6° S and 15.5° S, by green symbols. Ethane brightness temperature differences between 3.6° N and 15.5° N are denoted by orange symbols, and those between 3.6° S and 15.5° S, by red symbols. Despite evidence for the non-uniform distribution of ethane with latitude[7], a high correlation between the variations in brightness temperature differences at 7.8 μm and 12.2 μm exists for all years (except 1993 and 1995); we find that ΔT_b[12.2 μm] = 2.5ΔT_b[7.8 μm] + 2.71 K, allowing us to plot the two sets of data as shown. The difference in ΔT_b values can be attributed to the different vertical regions sampled by the two different wavelengths. A sinusoidal function that best fits the individual data (solid line) yields a period of 15.6 yr; a similar sinusoidal function that best fits an annual average of the data (dashed line) yields a period of 15.0 yr. The true time dependence is clearly not a simple sinusoid, and a correlation between the 1985–1992 12.2-μm data and 1997–2007 data yields a period of 13.9 yr. The mean of the three calculated periods is 14.8 ± 1.2 yr, which we take as our fitted value. The 1980 points, taken from convolutions of appropriate filter functions over Voyager-1 Infrared Radiometer Interferometer Spectrometer spectra from a north–south mapping sequence, were not considered in the fit of the sinusoidal function, although they are consistent with it.

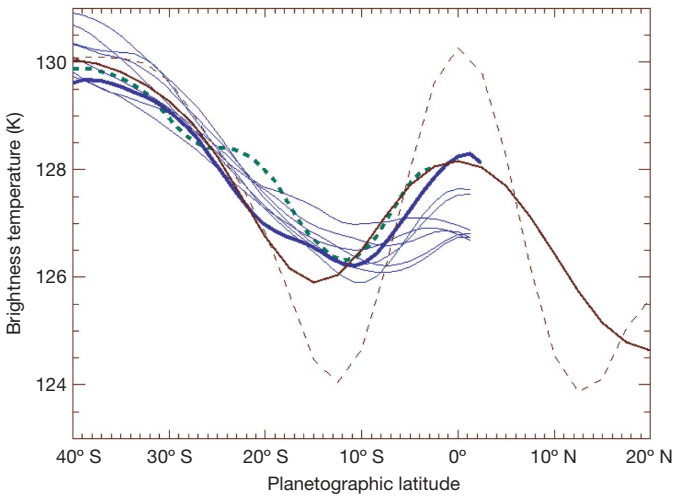


Figure 4 | Comparison of zonally averaged measurements and simulations of the meridional variability. Measurements are from 2006 (blue lines) and the simulated brightness temperatures are derived from convolving the upwelling radiance over the 7.8-μm filter used in the measurements. The thick blue line represents the observation with the best seeing, and the thick dashed green line represents higher resolution 2004 observations[10], for comparison. The upwelling radiance simulations (dashed brown line) are based on the zonal-mean temperature structures derived from Cassini CIRS spectra of Saturn's limb[7]. They have been convolved with a typical seeing to simulate observing conditions more accurately (solid brown line). It is most probable that the apparent variability in the observations arises from changes of seeing conditions from observation to observation; the best image of 2006 is plotted (thick blue line) for comparison. These changes are also likely to contribute to the ostensible short-term variability of the brightness temperature differences in Fig. 3.

oscillation[1,2] or Jupiter's quasi-quadrennial oscillation[3–5]. In fact, it is roughly one-half of Saturn's 29.4-yr orbital period, which raises the possibility that, like the Earth's semi-annual oscillation[1], these oscillations are phase locked by seasonal changes. In the Earth's semi-annual oscillation, westerly acceleration of the zonal wind is due, at least in part, to absorption of upwardly propagating, high-speed Kelvin waves, whereas easterly acceleration is associated with a combination of forcing by the seasonally reversing mean meridional circulation and seasonal momentum transfer from planetary waves in the winter hemisphere. Similar processes may be occurring on Saturn. Continued observations from the ground in the next few years and by the Cassini CIRS experiment during the operations extended beyond the July 2008 end of the primary mission should confirm both their periodicity and how they affect the spatial wave structure during a time when a rapid return to cooler equatorial brightness temperatures is expected.

Theoretical models similar to those for Jupiter's quasi-quadrennial oscillation[4,9] must be applied to the spatial and temporal variabilities we have observed in Saturn to determine the influence on the stratosphere of perturbations of the troposphere and radiative heating, including the influence of shadowing by the rings. Such models will allow us to determine the extent to which phenomena similar to Earth's quasi-biennial oscillation may be responsible for latitudinal transport[7]. Identifying analogous phenomena would have wide implications for transport in the stratospheres of all outer planets.

Received 11 December 2007; accepted 7 March 2008.

1. Andrews, D. G., Holton, F. R. & Leovy, C. B. *Middle Atmosphere Dynamics* (Academic, New York, 1987).
2. Baldwin, M. *et al.* The quasi-biennial oscillation. *Rev. Geophys.* **39,** 179–230 (2001).
3. Leovy, C. B., Friedson, A. J. & Orton, G. S. The quasiquadrennial oscillation of Jupiter's equatorial stratosphere. *Nature* **354,** 380–382 (1991).
4. Friedson, A. J. New observations and modeling of a QBO-like oscillation in Jupiter's stratosphere. *Icarus* **137,** 34–55 (1999).
5. Flasar, F. M. *et al.* An intense stratospheric jet on Jupiter. *Nature* **427,** 132–135 (2004).
6. Orton, G. S. *et al.* Thermal maps of Jupiter: Spatial organization and time dependence of stratospheric temperatures, 1980–1991. *Science* **252,** 537–542 (1991).
7. Fouchet, T. *et al.* An equatorial oscillation in Saturn's middle atmosphere. *Nature* doi:10.1038/nature06912 (this issue).
8. Gezari, D. Y. *et al.* New features in Saturn's atmosphere revealed by high-resolution thermal infrared images. *Nature* **342,** 777–780 (1989).
9. Li, X. & Read, P. L. A mechanistic model of the quasi-quadrennial oscillation in Jupiter's stratosphere. *Planet. Space Sci.* **48,** 637–669 (2000).
10. Orton, G. S. & Yanamandra-Fisher, P. A. Saturn's temperature field from high-resolution middle-infrared imaging. *Science* **307,** 696–701 (2005).
11. Hoffman, W. F., Fazio, G. G., Shivanandan, K., Hora, J. L. & Deutsch, L. K. in *Infrared Detectors and Instrumentation (Proc. SPIE 1946)* (ed. Fowler, A. M.) 449–460 (SPIE, 1993).
12. Ressler, M. E., Werner, M. E., Van Cleve, J. & Chou, H. A. The JPL deep-well mid-infrared array camera. *Exp. Astron.* **3,** 277–280 (1994).
13. Deutsch, L. K., Hora, J. L., Adams, J. D. & Kassis, M. in *Instrument Design and Performance for Optical/Infrared Ground-based Telescopes (Proc. SPIE 4841)* (eds Iye, M. & Moorwood, A. F. M.) 106–116 (SPIE, 2003).

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

Acknowledgements We acknowledge help from the staffs of the NASA IRTF, the Hale Observatories, the W. M. Keck Observatory, the Subaru Telescope and the

©2008 Nature Publishing Group

Steward Observatory for making these observations possible, as well as support for JPL from the Cassini project and NASA research and analysis programs. P.D.P. was a NASA Postdoctoral Fellow; L.F. was supported by the NASA Postdoctoral Program; J.F.N., A.S.B. and E.N. were Caltech Summer Undergraduate Research Fellows; and J.F.N. was supported by the NASA Space Grant program at JPL during the course of this work. The radiative-transfer calculations were performed on JPL supercomputer facilities. Full acknowledgments are given in the Supplementary Information section.

Author Contributions G.S.O., P.A.Y.-F., B.M.F., A.J.F., P.D.P., L.F., D.Y.G., F.V., A.T.T., J.C., K.H.B., J.L.H., M.E.R., J.L.H., Ta.F., Te.F., T.Z.M., J.T.B., W.F.H., L.K.D., J.E.V.C. and C.H. made the observations. J.L.H., J.D.A., M.K. and E.T. built and adapted the MIRSI instrument to routine operations at the IRTF. G.S.O., P.A.Y.-F., B.M.F., P.D.P., J.F.N., A.S.B., E.N., D.Y.G. and F.V. reduced the data. G.S.O., J.F.N. and A.S.B. analysed the data for time dependence and J.F.N. provided initial sinusoidal fits to the data. G.S.O. performed the radiative-transfer calculations, made the final data selection, and derived the final fitted parameters. G.S.O. and A.J.F. wrote the initial draft. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.S.O. (glenn.orton@jpl.nasa.gov).

©2008 Nature Publishing Group

LETTERS

An equatorial oscillation in Saturn's middle atmosphere

T. Fouchet[1], S. Guerlet[1], D. F. Strobel[2], A. A. Simon-Miller[3], B. Bézard[1] & F. M. Flasar[3]

The middle atmospheres of planets are driven by a combination of radiative heating and cooling, mean meridional motions, and vertically propagating waves (which originate in the deep troposphere). It is very difficult to model these effects and, therefore, observations are essential to advancing our understanding of atmospheres. The equatorial stratospheres of Earth and Jupiter oscillate quasi-periodically on timescales of about two and four years, respectively, driven by wave-induced momentum transport[1–5]. On Venus and Titan, waves originating from surface–atmosphere interaction and inertial instability are thought to drive the atmosphere to rotate more rapidly than the surface (superrotation). However, the relevant wave modes have not yet been precisely identified. Here we report infrared observations showing that Saturn has an equatorial oscillation like those found on Earth and Jupiter, as well as a mid-latitude subsidence that may be associated with the equatorial motion. The latitudinal extent of Saturn's oscillation shows that it obeys the same basic physics as do those on Earth and Jupiter. Future highly resolved observations of the temperature profile together with modelling of these three different atmospheres will allow us determine the wave mode, the wavelength and the wave amplitude that lead to middle atmosphere oscillation.

The stratosphere of Saturn is a region where temperatures increase with altitude, implying a large static stability and important wave activity. The upper stratosphere has a rich hydrocarbon chemistry that is initiated by methane photolysis and produces several species, from acetylene to benzene, which affect temperatures through their radiative properties. Unfortunately, observations performed before the arrival of the NASA Cassini spacecraft were generally limited in vertical coverage to the lower stratosphere (pressures greater than 1 hPa), where the dynamical and chemical timescales were longer than or comparable to a saturnian year[6–9]. Observations were also restricted to the summer hemisphere because the winter hemisphere was always hidden behind the optically thick rings. Cassini now gives us access to the full stratospheric range over the whole planet, in particular the upper stratosphere, where the seasonal forcing triggers rapid observable changes.

We acquired limb thermal spectra of Saturn for 2005–2006 using the Composite Infrared Spectrometer (CIRS) aboard Cassini. Limb viewing provides us with a vertical resolution of ~1.5 times the scale height (~75 km) and probes higher altitudes than nadir sounding[10]. Figure 1a displays the pressure–latitude cross-section of the zonal-mean temperature inverted from these CIRS limb spectra. Of particular note is the sharp difference in vertical structure between the northern and southern mid-latitudes and the equatorial region. At extratropical latitudes, the temperature steadily increases with altitude, whereas in the equatorial region it oscillates, with maximum amplitude centred on the equator. Although our temperature retrievals agree with previous measurements[11,12] wherever the latter are available, only the vertical resolution and the horizontal coverage on both sides of the equator provided by CIRS limb observations allow us to detect and map the equatorial vertical oscillation. Using the thermal-wind equation (see Supplementary Information), we derived a zonal thermal-wind pressure–latitude cross-section for the stratosphere (Fig. 1b). The vertical structure of the zonal thermal wind at the equator clearly shows alternate bands of eastward and westward vertical shear from the 0.2-hPa altitude down to the 5-hPa altitude.

This temperature and zonal wind structure resembles those of Earth's quasi-biennial oscillation (QBO) and Jupiter's quasi-quadrennial oscillation (QQO), in which temperature anomalies and eastward/westward winds alternate in altitude[1–5]. As the QBO and QQO vertical structures of alternating temperature anomalies and winds propagate down, the equatorial wind regime at a given height cyclically changes from easterly to westerly. On Earth, the alternating wind regimes repeat at intervals that vary from 22 to 34 months, with an average period of about 28 months[1,2]. On Jupiter, the equatorial stratospheric temperature exhibits a 4.4-year period and the equatorial zonal winds in the upper troposphere oscillate with a 4.5-year period[3,13,14]. Long-term ground-based monitoring reveals a period of 14.7 ± 0.9 terrestrial years on Saturn[15]. The observational similarities between Saturn's oscillation and the QBO and QQO are the strong equatorial confinement of temperature minima and maxima and associated shear layers, a stronger eastward than westward shear layer, and the bounding of the equatorial oscillation at latitudes 15–20° north and south. Temperatures near these latitudes are relatively high when equatorial temperatures are relatively low, and vice versa.

On Earth and Jupiter, the quasi-periodic oscillations are triggered by the interaction between upwardly propagating waves and the mean zonal flow[2,4,16–20]. Planetary-scale Kelvin waves contribute to eastward stresses, whereas mixed Rossby-gravity waves contribute to westward stresses. Over time, the zonal-mean wind anomalies descend and are damped by viscous processes at the tropopause, leading to the observed quasi-periodic cycle. Observations of Saturn (either cloud tracking or thermal measurements[21–24]) show abundant evidence for the presence of waves of sizes ranging from planetary scale to mesoscale. These waves may be generated by the deep convection that carries the internal heat flux and waves generated by shear instabilities. Hence, Saturn exhibits conditions similar to the terrestrial and jovian environments that are favourable to the establishment of an equatorial oscillation driven by vertically propagating waves. A simple model that considers the equations for the evolution of a longitudinally symmetric atmosphere subject to a mechanical forcing shows that the equatorial oscillation should be confined within a distance L of the equator (see Supplementary

[1]LESIA, Observatoire de Paris, Université Pierre et Marie Curie, CNRS, Paris 7, Meudon F-92195, France. [2]Departments of Earth and Planetary Sciences & Physics and Astronomy, Johns Hopkins University, Baltimore, Maryland 21218, USA. [3]NASA/Goddard Space Flight Center, Code 693, Greenbelt, Maryland 20771, USA.

©2008 Nature Publishing Group

Information). A scale analysis of these equations yields the relation $L < (aDN/(2\Omega))^{1/2}(\omega/\alpha)^{1/4}$, where a is Saturn's radius, D is a height scale, N is the buoyancy frequency, Ω is Saturn's spin rate, α is a constant radiative timescale and ω is the mechanical forcing frequency. This scale analysis for Saturn yields $L = 7{,}400$ km (seven degrees of latitude), in satisfactory agreement with our observations.

At extratropical latitudes, the response of a rapidly rotating atmosphere to the forcing due to planetary-scale waves is quite distinct from that in the tropics: the forcing is balanced by the Coriolis force produced by a mean meridional circulation. Indeed, at mid-latitudes, the stratospheric thermal profile that we retrieved differs completely from that in the equatorial region, in that it steadily increases with altitude. Surprisingly, the temperature difference between 45° S and 45° N at pressures between 10 and 1 hPa is larger than that measured for pressures between 0.1 and 0.01 hPa. At 1–10 hPa the temperature difference peaks at 15 K, whereas at 0.01–0.1 hPa it reaches only 6–7 K. The season observed on Saturn was southern midsummer (the solstice took place in October 2002). The observed temperature gradient between the southern and northern hemispheres should be close to its greatest seasonal amplitude[8]. Because the thermal inertia increases with depth, owing to larger optical depths, the latitudinal gradient was expected to be higher in the upper stratosphere than it is in the lower stratosphere.

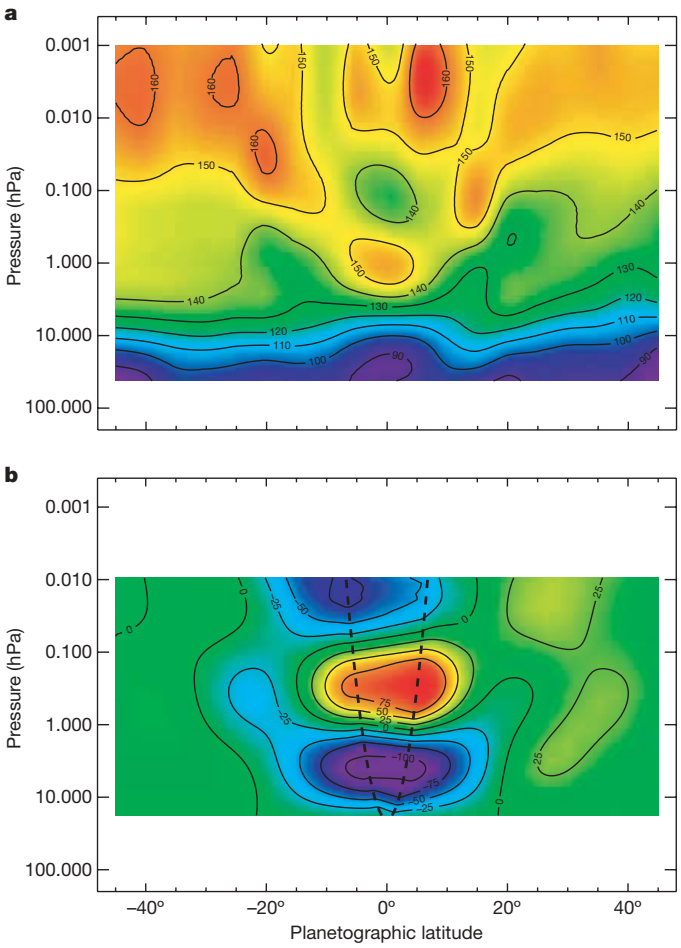


Figure 1 | Saturn's equatorial oscillation. **a**, This temperature map was retrieved using the hydrogen collision-induced continuum (600 cm^{-1}) between the 20 and 2 hPa pressure levels, and the ν_4-band of methane (1,300 cm^{-1}) between the 3 and 0.003 hPa pressure levels (see Supplementary Information). The contours indicate temperature in Kelvin. The CIRS observed the planet's limb each 5° of latitude between 45° S and 45° N (here and in **b**, negative latitudes lie in the southern hemisphere). The ten detectors were aligned vertically, perpendicular to the planet's limb. Each detector independently probed a given altitude with a vertical resolution of 50–100 km. The temperatures presented in this figure were retrieved from 15.5-cm^{-1} spectra. Vertical profiles of atmospheric temperature were retrieved by means of a constrained inversion with strong low-pass filtering at each limb position. The number of independent longitudes sampled for a given latitude varies between two and four. At the equator the temperature near 10 hPa is 17 K colder than it is at 20° S and 20° N, whereas the equatorial temperature is 20 K warmer near 1 hPa; near 0.1 hPa the equator is again 20 K colder, and at 0.01 hPa the equatorial region is 10 K warmer than it is at adjacent latitudes. **b**, This map presents the thermal wind field obtained by upward integration of the thermal wind equation on cylinders (see Supplementary Information). The contours indicate measured wind speed minus wind speed at 20 hPa, in metres per second. The dashed parabola delineates the exclusion zone within which the winds cannot be calculated from the thermal wind equation, but are instead interpolated. The error in temperature due to instrument noise is 2 K, and this propagates to an error of 5/(sine of latitude) m s^{-1} per scale height in the zonal winds.

To further explore this issue, we calculated the radiative cooling rates using the hydrocarbon vertical profiles inverted from CIRS limb data (Fig. 2). At 0.01 hPa we obtained a similar cooling rate of ~0.2 K per day for latitudes 30° S and 30° N (Fig. 3). This similarity in upper stratosphere cooling rates between the summer and winter hemispheres is in sharp contrast with the differences in the daily mean insolation, which combines the effects of Saturn's obliquity and screening by the thick rings. At the time of our observations, the ratio of the solar insolation at 30° S to that at 30° N is greater than three (daily mean of 5 W m^{-2} in the southern hemisphere and 1.5 W m^{-2} in the northern hemisphere[6,9]). Hence, the observed state of the upper stratosphere requires either that the heating rate is dominated by a season-independent source or that heat is meridionally redistributed.

The meridional–vertical cross-section of the ethane abundance supports the hypothesis that heat is meridionally redistributed (Fig. 2). In the southern hemisphere, the observed equator-to-mid-latitude abundance gradient agrees with the prediction of photochemical models that it should resemble the yearly average insolation rather than the current seasonal insolation[9]. However, in the northern hemisphere, the local maximum of the abundance in the upper atmosphere centred at 30° N matches neither the yearly average insolation nor the current insolation. This local maximum constitutes a second strong piece of evidence for transport of hydrocarbon-rich air masses from the southern upper stratosphere to the northern hemisphere. We estimated the vertical velocity needed to explain the

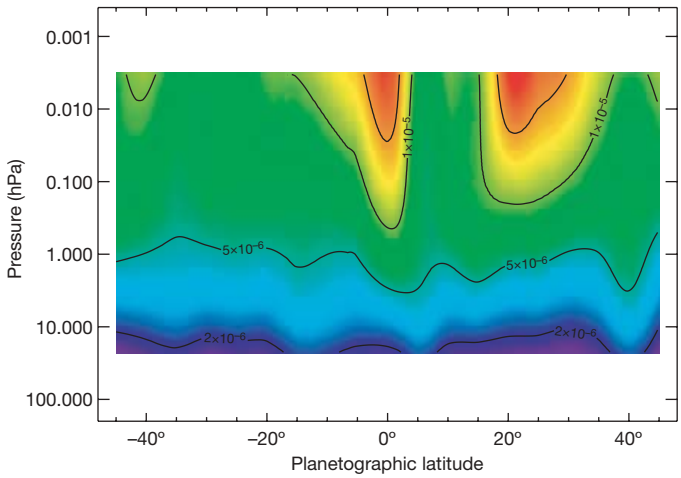


Figure 2 | Vertical–zonal map of ethane abundance. This pressure–latitude cross-section of the ethane volume mixing ratio (indicated by the contours) is inverted by fitting the ν_9-band of ethane (820 cm^{-1}) observed by the CIRS between 1 and 0.01 hPa. The temperature field used in the calculation is presented in Fig. 1. CIRS spectra measuring the temperature and the composition were obtained simultaneously at the same vertical and spectral resolution. Vertical profiles of the ethane volume mixing ratio were retrieved by means of a constrained inversion with strong low-pass filtering (see Supplementary Information).

©2008 Nature Publishing Group

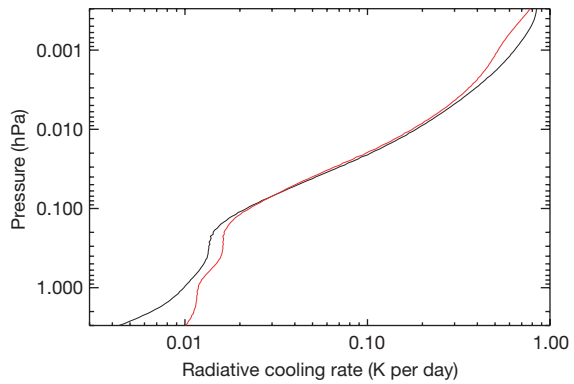


Figure 3 | Radiative cooling rates in the northern and southern hemispheres. The radiative cooling rates were calculated at 30° N (black) and 30° S (red) between 2 and 0.003 hPa, by differentiating the net thermal flux integrated over the ν_4-band of methane, the ν_9-band of ethane and the ν_5-band of acetylene. The hydrocarbon abundance was set to the retrieved value presented in Fig. 2.

observed temperature, assuming adiabatic heating ($N^2HR^{-1}w = J/c_p$, where H is the atmospheric scale height, w is the vertical wind speed, J is the heating rate, R is the gas constant and c is the specific heat capacity at constant pressure). We did not calculate the solar heating rate, but we assumed that the atmosphere is globally in radiative balance, that is, that the total radiative cooling at 30° south and north is equal to the total solar heating at these two latitudes. With a heating rate three times larger at 30° S than it is at 30° N, this gives heating rates of 0.30 and 0.10 K per day, respectively, leading to a vertical velocity of 1.5 mm s^{-1}.

This extratropical meridional circulation might be influenced by the equatorial oscillation. On Earth, the QBO does modulate the mean meridional circulation. Equatorial temperature minima are associated with local ascent and, by continuity, descent at extratropical latitudes. One interesting aspect of the global QBO is that the extratropical temperature anomalies are seasonally synchronized, occurring primarily during winter in each hemisphere. The reason for this is not fully known, but one possibility is that during northern winter the meridional flow from south to north, with subsidence in the northern hemisphere, will be reinforced by the QBO-induced sinking motion at extratropical latitudes. On Saturn, the latitude-15° temperature anomalies at the 0.1-Pa and 20-Pa pressure levels are indeed warmer in the northern hemisphere than they are in the southern hemisphere. This seasonal asymmetry is another strong similarity in the behaviours of the stratospheres of Saturn and Earth.

Received 10 December 2007; accepted 10 March 2008.

1. Andrews, D. G., Holton, J. R. & Leovy, C. B. *Middle Atmosphere Dynamics* (Academic, New York, 1987).
2. Baldwin, M. P. *et al.* The quasi-biennial oscillation. *Rev. Geophys.* **39,** 179–230 (2001).
3. Leovy, C. B., Friedson, A. J. & Orton, G. S. The quasiquadrennial oscillation of Jupiter's equatorial stratosphere. *Nature* **354,** 380–382 (1991).
4. Friedson, A. J. New observations and modelling of a QBO-like oscillation in Jupiter's stratosphere. *Icarus* **137,** 34–55 (1999).
5. Flasar, F. M. *et al.* An intense stratospheric jet on Jupiter. *Nature* **427,** 132–135 (2004).
6. Barnet, C. D., Beebe, R. F. & Conrath, B. J. A seasonal radiative-dynamic model of Saturn's troposphere. *Icarus* **98,** 94–107 (1992).
7. Bézard, B. & Gautier, D. A seasonal climate model of the atmospheres of the giant planets at the Voyager encounter time. I. Saturn's stratosphere. *Icarus* **60,** 296–310 (1985).
8. Conrath, B. J., Gierasch, P. J. & Leroy, S. S. Temperature and circulation in the stratosphere of the outer planets. *Icarus* **83,** 255–281 (1990).
9. Moses, J. I. & Greathouse, T. K. Latitudinal and seasonal models of stratospheric photochemistry on Saturn: Comparision with infrared data from IRTF/TEXES. *J. Geophys. Res.* **110,** 09007 (2005).
10. Flasar, F. M. *et al.* Exploring the Saturn system in the thermal infrared: The Composite Infrared Spectrometer. *Space Sci. Rev.* **115,** 169–297 (2004).
11. Flasar, F. M. *et al.* Temperatures, winds, and composition in the Saturnian system. *Science* **307,** 1247–1251 (2005).
12. Greathouse, T. K. *et al.* Meridional variations of temperature, C_2H_2 and C_2H_6 abundances in Saturn's stratosphere at southern summer solstice. *Icarus* **177,** 18–31 (2005).
13. Orton, G. S. *et al.* Thermal maps of Jupiter: Spatial organization and time dependence of stratospheric temperatures, 1980 to 1990. *Science* **252,** 537–542 (1991).
14. Simon-Miller, A. A., Poston, B. W., Orton, G. S. & Fisher, B. Wind variations in Jupiter's equatorial atmosphere: A QQO counterpart? *Icarus* **186,** 192–203 (2007).
15. Orton, G. S. *et al.* Semi-annual oscillations in Saturn's low-latitude stratospheric temperatures. *Nature* doi:10.1038/nature06897 (this issue).
16. Lindzen, R. S. & Holton, J. R. A theory of the quasi-biennial oscillation. *J. Atmos. Sci.* **25,** 1095–1107 (1968).
17. Holton, J. R. & Lindzen, R. S. An updated theory for the quasi-biennial cycle of the tropical troposphere. *J. Atmos. Sci.* **29,** 1076–1080 (1972).
18. Gray, L. J. & Pyle, J. A. A two-dimensional model of the quasi-biennial oscillation in ozone. *J. Atmos. Sci.* **46,** 203–220 (1989).
19. Dunkerton, T. J. Nonlinear propagation of zonal winds in an atmosphere with Newtonian cooling and equatorial wavedriving. *J. Atmos. Sci.* **48,** 236–263 (1991).
20. Li, X. & Read, P. L. A mechanical model of the quasi-quadrennial oscillation in Jupiter's stratosphere. *Planet. Space Sci.* **48,** 637–669 (2000).
21. Beebe, R. F., Barnet, C., Sada, P. V. & Murrell, A. S. The onset and growth of the 1990 equatorial disturbance on Saturn. *Icarus* **95,** 163–172 (1992).
22. Sanchez-Lavega, A. Observations of Saturn's Ribbon Wave 14 years after Its discovery. *Icarus* **158,** 272–275 (2002).
23. Achterberg, R. K. & Flasar, F. M. Planetary-scale thermal waves in Saturn's upper troposphere. *Icarus* **119,** 350–369 (1996).
24. Schinder, P. J. *et al.* Vertically propagating waves in the upper atmosphere of Saturn from Cassini radio occultations. (AGU Fall Meeting Abstract No. P23D-06, 2005).

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

Author Contributions T.F. and S.G. inverted the temperature and composition from CIRS limb observations. T.F., F.M.F. and D.F.S. integrated the thermal wind equation. T.F., S.G. and B.B. calculated the radiative cooling rates. A.A.S.-M. designed the observations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.F. (thierry.fouchet@obspm.fr).

©2008 Nature Publishing Group

LETTERS

Quantum oscillations in a molecular magnet

S. Bertaina[1]†, S. Gambarelli[2], T. Mitra[3], B. Tsukerblat[4], A. Müller[3] & B. Barbara[1,2]

The term 'molecular magnet' generally refers to a molecular entity containing several magnetic ions whose coupled spins generate a collective spin, S (ref. 1). Such complex multi-spin systems provide attractive targets for the study of quantum effects at the mesoscopic scale. In these molecules, the large energy barriers between collective spin states can be crossed by thermal activation or quantum tunnelling, depending on the temperature or an applied magnetic field[2–4]. There is the hope that these mesoscopic spin states can be harnessed for the realization of quantum bits—'qubits', the basic building blocks of a quantum computer—based on molecular magnets[5–8]. But strong decoherence[9] must be overcome if the envisaged applications are to become practical. Here we report the observation and analysis of Rabi oscillations (quantum oscillations resulting from the coherent absorption and emission of photons driven by an electromagnetic wave[10]) of a molecular magnet in a hybrid system, in which discrete and well-separated magnetic V_{15}^{IV} clusters are embedded in a self-organized non-magnetic environment. Each cluster contains 15 antiferromagnetically coupled $S = 1/2$ spins, leading to an $S = 1/2$ collective ground state[11–13]. When this system is placed into a resonant cavity, the microwave field induces oscillatory transitions between the ground and excited collective spin states, indicative of long-lived quantum coherence. The present observation of quantum oscillations suggests that low-dimension self-organized qubit networks having coherence times of the order of 100 μs (at liquid helium temperatures) are a realistic prospect.

In the context of quantum computing, it was recently discussed how the decoherence of molecular magnet spin quantum bits could be suppressed, with reference to the discrete low spin clusters V_{15} and Cr_7Ni (ref. 7; see also refs 8 and 14). In both systems, their low spin states cause weak environmental coupling[7], making them candidates for the realization of a long-lived quantum memory. Measurement of the spin relaxation time τ_2 in Cr_7Ni was subsequently reported and found to be interestingly large[15,16]; however, the important Rabi quantum oscillations were not observed, probably because electronic and nuclear degrees of freedom were too strongly linked to each other. As these oscillations have until now only been observed in non-molecular spin systems (see, for example, refs 17–20), it has remained an open question whether quantum oscillations could in principle be realized in molecular magnets[7,8]. This question is now answered by our observation of quantum oscillations of the Rabi type in V_{15}. The main reason for this success lies in the fact that the important pairwise decoherence mechanism[7,8] associated with dipolar interactions could be strongly reduced.

Before discussing the observed quantum oscillations, we first briefly describe the magnetic/electronic structure of the V_{15}^{IV} species as determined experimentally. Following the synthesis of the quasi-spherical mesoscopic cluster anion $\left[V_{15}^{IV}As_6^{III}O_{42}(H_2O)\right]^{6-}$ ($\equiv V_{15}$) nearly two decades ago (ref. 11), the properties of this molecule have received considerable attention (see, for example, refs 1, 11, 14, 21–25). The V_{15} cluster with an ~1.3 nm diameter exhibits an unique structure with layers of different magnetizations: a large central V_3^{IV} triangle is sandwiched by two smaller V_6^{IV} hexagons[11] (Fig. 1). The 15 $S = 1/2$ spins are coupled by antiferromagnetic super-exchange and Dzyaloshinsky–Moriya (DM) interaction[13,21–25] (see also refs 26, 27) through different pathways, which results in a collective low spin ground state with $S = 1/2$ (refs 12, 13, 24, 25).

Energy spectrum calculations for the full cluster spin space give two $S = 1/2$ (spin doublet) ground states slightly shifted from each other by DM interactions, and an $S = 3/2$ (spin quartet) excited state; these states are 'isolated' from a quasi-continuum of states lying at energy $E/k_B \approx 250$ K above the $S = 3/2$ excited state. These low-lying energy states can be obtained with a good accuracy using the generally accepted three-spin approximation (valid below 100 K), in which the spins of the inner triangle are coupled by an effective interaction $|J_0| \ll |J'|$ mediated by the spins of the hexagons[12,13,21–25] (Fig. 2 and Methods; J_0 and J' are shown in Fig. 1b).

The spin hamiltonian of V_{15} can be written as:

$$H = -J_0 \sum_{\substack{i,j=1 \\ (i<j)}}^{3} \mathbf{S}_i \mathbf{S}_j + \sum_{ij=12,13,31} \mathbf{D}_{ij}\left(\mathbf{S}_i \times \mathbf{S}_j\right) + A \sum_{i=1}^{3} \mathbf{I}_i \mathbf{S}_j + g\mu_B \mathbf{H} \sum_{i=1}^{3} \mathbf{S}_i \quad (1)$$

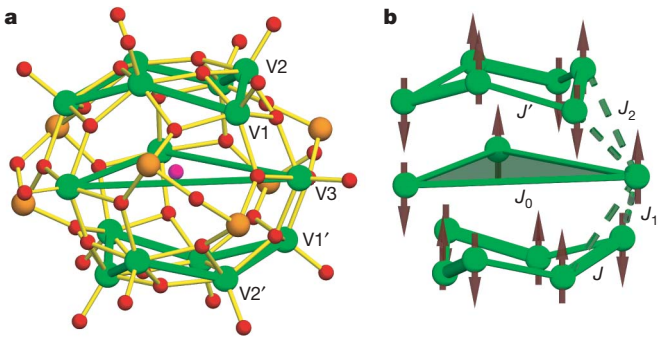


Figure 1 | Structure and exchange interaction pathways of the cluster anion $[V_{15}^{IV}As_6^{III}O_{42}(H_2O)]^{6-}$. a, The cluster is shown in ball-and-stick representation (green, V; orange, As; red, O). The outer V_6 hexagons are highlighted by thick green lines. A weak deviation from trigonal symmetry can be attributed to the water molecule located in the centre of the cavity (O of the encapsulated water molecule in purple) or/and to the presence of water in the lattice between molecules. The different types of V ions, namely V1, V2, V3, V1′ and V2′, are shown for the definition of different exchange pathways. **b**, Sketch showing the spin arrangement at low temperatures (three-spin approximation), emphasizing some of the exchange interaction pathways (J, J_1, J_2, J'). The coupling J_0 between the spins of the inner triangle is not direct but results from different exchange pathways through the hexagons. The magnetic layer system is defined by one V_3 triangle sandwiched by two V_6 hexagons (for further details see text).

[1]Institut Néel, CNRS, 25 Ave. des Martyrs, BP166, 38042 Grenoble Cedex 9, France. [2]Laboratoire de Chimie Inorganique et Biologique (UMR-E3 CEA-UJF), INAC, CEA-Grenoble, 17 Ave. des Martyrs, 38054 Grenoble Cedex 9, France. [3]Fakultät für Chemie, Universität Bielefeld, Postfach 100131, D-33501 Bielefeld, Germany. [4]Department of Chemistry, Ben-Gurion University of the Negev, PO Box 653, 84105 Beer-Sheva, Israel. †Present address: National High Magnetic Field Laboratory, Florida State University, 1800 East Paul Dirac Drive, Tallahassee, Florida 32310, USA.

©2008 Nature Publishing Group

where $\mathbf{D}_{ij}$ is the antisymmetric vector of the DM interaction associated with the pair ij, and A is the hyperfine coupling constant of the ^{51}V isotope (see below). The six components of $\mathbf{D}_{ij}$ can be expressed in terms of two parameters, namely D_Z (perpendicular to the plane) and D_{XY} (in-plane). The DM interaction removes the degeneracy of the two low-lying doublets and produces a first order zero-field splitting $\Delta_{DM} \approx \sqrt{3} D_Z$ (plus small second order corrections)[22–25]. The excited (quartet) state shows only a second order splitting caused by a small inter-multiplet mixing through the in-plane component of DM coupling, that is, $\Delta'_{DM} = -D_{XY}^2/8J_0$ (refs 24, 25). The energy separation between the doublet states and quartet state is given by $3J_0/2 \approx -3.67$K (refs 13, 21–25). Figure 2 shows the level scheme calculated by diagonalization of the hamiltonian (equation (1)), with only one free parameter $D_Z \approx 43$ mK adjusted to fit the positions of the measured resonances (a value close to that obtained from magnetization data[13,21,24]), and $D_{XY} = 0$, a choice conditioned by the fact that the transverse DM component has a negligible effect on resonance fields below 0.5 T (this is important in the calculation of transition probabilities only). To ensure legibility, hyperfine interactions are not included in Fig. 2 (they simply broaden the levels).

A new hybrid material, based on the use of a cationic surfactant—DODA ($[Me_2N\{(CH_2)_{17}Me\}_2]^+$)—as an embedding material for the anionic clusters, was developed for the present work (see Methods). The related frozen system contains V_{15} clusters integrated into the self-organized environment of the surfactant. The clusters—prepared according to ref. 11—were extracted from aqueous solution into chloroform by the surfactant DODA present in large excess. The surfactants, which wrap up the cluster anions, are amphiphilic cations, with their long hydrophobic tails pointing away from the cluster anions, enabling solubility in chloroform. The procedure ensures that the cluster anions cannot get into direct contact with one another; they are clearly separated by the surfactants (mean distance ~13 nm).

Electron paramagnetic resonance (EPR) experiments were performed on this hybrid material at ~4 K using a Bruker E-580 X-band continuous-wave (CW) and pulsed spectrometer operating at 9.7 GHz. The CW-EPR spectrum, recorded at 16 K on a frozen sample, corresponds precisely to that obtained in the solid state in a previous study[12]. In particular, the resonance field shows the same profile and line-width (~30 mT), compatible with the g-tensor values of a single crystal ($g_{//} = 1.98$ and $g_{\perp} = 1.95$). The measured transition width $W \approx 35$ mT is directly connected with the energy E occurring in the expression of decoherence calculated for a multi-spin molecule[7,8] (see below). Note that this transition width W should be associated with $S = 3/2$, the EPR spectrum being dominated by the excited quartet.

Rabi oscillations were recorded using a nutation pulse of length t, followed (after a delay greater than τ_2) by a $\pi/2-\pi$ sequence. Experimental results showed two different types of Rabi oscillations, corresponding to the resonant transitions 1, 2 and 3 for $S = 3/2$ spins, and 4, 5, 6 and 7 for $S = 1/2$ spins, here called '3/2' and '1/2', respectively (Fig. 3b and a, respectively). Although both types of oscillation are associated with the same collective degrees of freedom of the clusters, they show very different behaviour. In particular, the first type of Rabi frequency compares well with that of a single spin-3/2 system, whereas the Rabi frequency of the second type is much smaller than that of a single spin-1/2. This is a consequence of selection rules: the transition type '3/2' is always allowed, whereas the transitions 5 and 7 of the '1/2' type occur only due to transverse DM interactions or/and breaking of the C_3 symmetry[25] (Methods). Therefore we obtained Rabi oscillations with quite different frequencies, $\Omega_{R\,3/2} \approx 18.5 \pm 0.2$MHz and $\Omega_{R1/2} \approx 4.5 \pm 0.2$MHz, and a small ratio of transition probabilities (or intensities) $R < 6 \times 10^{-2}$ (Fig. 3, Methods). When the transition '1/2' is excited (by a single excitation pulse), a whole spectrum of Rabi oscillations is generated. The frequency of the detected oscillation depends on the characteristics of the detection pulse, such as its length or its amplitude (Fig. 3). This

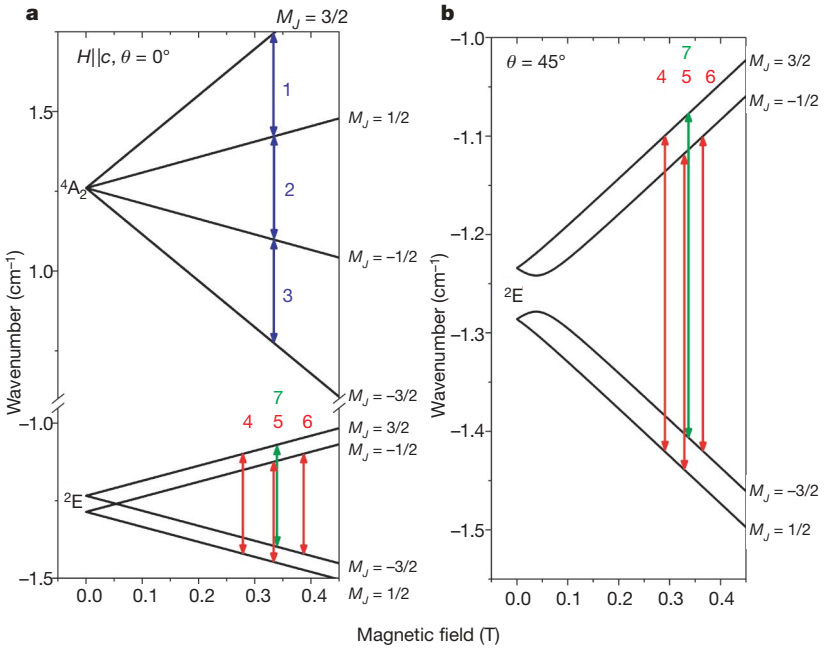


Figure 2 | Low-energy EPR transitions. These calculated diagrams used parameter values $g \approx 1.96$, $J_0 \approx -2.45$ K, $D_Z \approx 43$ mK and $D_{XY} = 0$ (see text for details). **a**, The magnetic field is parallel to the c axis. Whereas the orbital singlet $^4A_2$ ($S = 3/2$) gives the superposition of the three transitions 1, 2, and 3, the orbital doublet ^2E ($S = 1/2$) gives two inter-doublet transitions 4 and 6 which are basically allowed, as well as two intra-doublet transitions 5 and 7 (which are respectively allowed by transverse DM interactions and non-symmetrical exchange interactions due to a small deviation from the trigonal symmetry[24] (Methods)). Second order zero-field splitting of $^4A_2$ and small splitting of the lines 1, 2, and 3 is not shown. The M_J labels correspond to the quantization axis along the DM anisotropy field. **b**, Shown are the transitions 4, 5, 6, and 7 for the angle $\theta = 45^\circ$ between the field and cluster C_3-axis. The M_J labels correspond to the quantization axis along the field in the strong field limit. The boundaries of the measured resonance fields of Fig. 4 correspond to the field distribution given by the positions of labels 4, 5, 6 and 7 above the curves. The blue transitions correspond to $S = 3/2$; the red and the green transitions correspond to $S = 1/2$.

©2008 Nature Publishing Group

spectrum is due to the presence of an avoided level crossing and the special selection rules; these are caused by the uniaxial anisotropy introduced by the DM interactions in the spin-frustrated (orbitally degenerate) ground state giving the overlapping transitions 4–7 (Fig. 2). The glassy character of the investigated frozen material is also relevant here; this material contains different cluster orientations, leading to a distribution of transverse field components, which gives a scattering of the coefficients of the states entering in the two-level wavefunctions $|\varphi_1\rangle$ and $|\varphi_2\rangle$ and therefore a distribution of the Rabi frequencies $\Omega_{R\,1/2} \propto |\langle\varphi_1|S_+|\varphi_2\rangle|$ (Fig. 2 and Methods). Whereas the splitting of the excited quartet state in a magnetic field is almost isotropic, the distribution function of the associated Rabi frequency is very narrow.

An extension of the experiments shown in Fig. 3 to other values of the applied field showed that Rabi oscillations could be detected for each value of the applied field below 500 mT, while the transitions are inhomogeneously broadened. Figure 4 gives the result of a systematic investigation, consisting of the measurement of the spin-echo intensity at time $t = 0$ in a sweeping magnetic field. Two broad resonance distributions are observed, which correspond to the Rabi oscillations '3/2' and '1/2' of Fig. 3b and a, respectively, which were measured near the maxima $H_{3/2} \approx 357$ mT and $H_{1/2} \approx 335$ mT of the curves of Fig. 4. Whereas the nearly symmetrical type '3/2' distribution shows resonances which are optimally excited by pulse durations and powers similar to those generally used for isolated 3/2 spins, the asymmetrical type '1/2' distribution shows resonances requiring larger power and pulse length, confirming much smaller transition probabilities. The observed inhomogeneous widths ($\sim 50 \pm 10$ mT) result from the existence of different transitions—that is 1 to 3 and 4 to 7 shifted by the longitudinal field components associated with the glassy character of the frozen solution. The width of the resonance of type '1/2' (Fig. 4) fits the transition fields calculated from the hamiltonian (equation (1)) for the resonances 4 to 7 with limiting angles $\vartheta = 0$ and $\pi/2$ (Fig. 2), whereas the width of the resonance of type '3/2' is simply given by the unique resonance field of transitions 1 to 3 (Fig. 2 a). In both cases, the ^{51}V hyperfine interactions contribute equally to the resonance widths.

To conclude, it was possible to entangle the 15 spins of a molecular magnet—a complex system which, formally speaking, entails a Hilbert space of dimension $D_H = 2^{15}$ (Methods)—with photons by performing pulse EPR experiments on a frozen solution of randomly oriented and well separated clusters. Despite the complexity of the system[11–14,21–25] (involving in a formal consideration dozens of cluster electrons and nuclear spins of ^{51}V, ^{75}As and ^{1}H), long-lived Rabi oscillations[10] were generated and selectively detected. An analysis, based on the widely used three-spin approximation of V_{15} (refs 12, 13, 21–25; the related interactions are mediated by the 12 other spins) gives a global interpretation of the results.

The observed coherence on the microsecond timescale seems to be mainly limited by the bath of nuclear spins. Each V_{15} cluster is correspondingly weakly coupled to 36 first-neighbour protons of the six DODA methyl groups distributed around the cluster, and to two water protons at the cluster centre. According to the charge (6−) of V_{15}, six cationic DODA surfactants are relevant, with their positively charged parts (six dimethyl groups) attached to the O atoms of the cluster surface (see also ref. 28); the corresponding neutral hybrid just leads to the solubility in the organic solvent. The distance from the H atoms of a methyl group to a V^{IV} is ~0.45 nm. For this typical spin–proton distance, the half-width of the gaussian distribution of the coupling energy of a cluster/surfactant unit is $E \approx 3.5$ mK, giving, for the level separation $\Delta \approx 0.4$ K (Fig. 2), the coherence time[7,8]

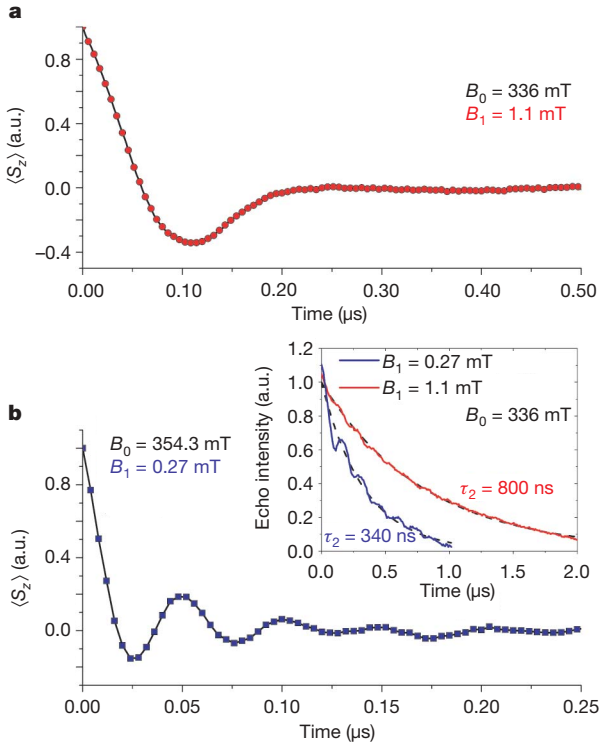


Figure 3 | Generation and detection of Rabi oscillations. **a**, Time evolution of the average spin $\langle S_z\rangle$ after a spin-echo sequence. The '1/2' type transition observed near the maximum of the corresponding resonance of Fig. 4 (B_0, 336.0 mT) requires unusually large excitation power $B_1 = 1.1$ mT and pulse length $T_{\pi/2} = 64$ ns. It corresponds to the transitions 4–7 of Fig. 2. The Rabi frequency $\Omega_R = 4.5$ MHz was selected by a detection pulse with characteristics $B_1 = 0.3$ mT and $T_{\pi/2} = 200$ ns. **b**, The '3/2' type transition with Rabi frequency $\Omega_R = 18.5$ MHz was excited near the maximum of the corresponding resonance of Fig. 4 ($B_0 = 354.3$ mT). It requires excitation and detection pulses similar to those usually used for a single spin of 3/2 ($B_1 = 0.27$ mT, $T_{\pi/2} = 16$ ns) and corresponds to the transitions 1–3 of Fig. 2. Inset, spin-echo intensity measured versus time for both oscillations. The coherence times τ_2 obtained from exponential fits are inverse functions of the spin values: 800 ns for $S = 1/2$ (red) and 340 ns for $S = 3/2$ (blue). Superimposed oscillations, mainly observed on the '3/2' type curve come from the precession of proton spins[19]. These oscillations correspond to only a weak perturbation of the Rabi coherence. Temperature, 4 K for all results shown.

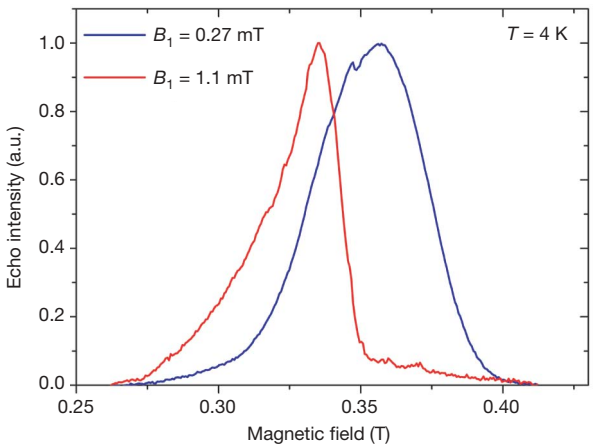


Figure 4 | Distribution of spin-echo intensities. The measurements were done in field sweep experiments for two excitation pulse configurations. The blue curve, corresponding to '3/2' type transitions (obtained with the excitation pulse $B_1 = 0.27$ mT and $T_{\pi/2} = 16$ ns), is nearly symmetrical and has a high transition probability. The red curve, corresponding to '1/2' type transitions (obtained with the excitation pulse $B_1 = 1.1$ mT and $T_{\pi/2} = 64$ ns), is asymmetrical and has a low probability (involving collective orbital degrees of freedom). The resonance fields form a 'band' due to random cluster orientations, while the corresponding distribution widths can be well explained by the dispersions of the resonance transitions 1–3 and 4–7.

©2008 Nature Publishing Group

$\tau_2^{\mathrm{H}} = 4\pi\Delta/E^2 \approx 18\,\mu s$. The contribution of more distant neighbouring protons should reduce this value to a few microseconds. Regarding the decoherence effect from ^{51}V, the transition width $W \approx 35$ mK gives $E = W/2 \approx 17$ mK and $\tau_2^{\mathrm{V}} = 0.75\,\mu s$, suggesting that the observed decoherence of the $S = 3/2$ resonances is almost entirely caused by the ^{51}V nuclear spins. The observed larger coherence time of the $S = 1/2$ transitions is presumably due to their smaller hyperfine coupling. In spite of the relatively high temperature of the measurement, the phonons' decoherence[7,8] $\tau_2^{\mathrm{ph}} \propto S^{-4}$ is strongly lowered due to the low spin and anisotropy values involved in the electron–phonon[29,30] coupling $\propto \left| < i \left| S_y S_z + S_z S_y \right| f > \right|^2$, giving $\tau_2^{\mathrm{ph}} \approx 100\,\mu s$, that is, $\tau_2^{\mathrm{ph}} \gg \tau_2^{\mathrm{H}} > \tau_2^{\mathrm{V}}$. Finally, the pairwise decoherence mechanism originating from electronic dipolar interaction[7], which is usually considered as the most destructive, is nearly negligible, owing to the strong dilution of the clusters that results from the surfactant environment. This allows weak dipolar interactions only ($\sim$0.5 μK) and very large coherence times ($\tau_2^{\mathrm{pw}} \approx 100\,\mu s$). A comparison of the different decoherence mechanisms suggests that coherence times greater than 100 μs should be obtained in molecular magnets at liquid-helium temperatures if nuclear-spin-free molecules and deuterated surfactants are used.

The control of complex coherent spin states of molecular magnets—in which exchange interactions can be tuned by well defined chemical changes of the metal cluster ligand spheres—could finally lead to a way to avoid the 'roadblock' of decoherence. This would be particularly important in the case of self-organized one- or two-dimensional supramolecular networks, where well separated magnetic species could be addressed selectively, following different schemes already proposed for the molecular magnet option.

METHODS SUMMARY

When we refer to the three-spin approximation of V_{15} (refs 12, 13, 21–25), we consider the three spins located on each corner of the inner triangle (Fig. 1b). However these spins do not interact directly but via the other spins of the cluster. Strictly speaking, each hexagon contains three pairs of spins strongly coupled with $J \approx -800$ K ('dimers') and each spin of the inner triangle is coupled to two of those pairs, one belonging to the upper hexagon and one belonging to the lower hexagon ($J_1 \approx -150$ K and $J_2 \approx -300$ K). This gives three groups of five spins with resultant spin $S = 1/2$ (superposition of 'entangled' states, coupled through inter-dimer hexagon superexchange $J' \approx -150$ K), showing that, in fact, the three-spin approximation involves all of the 15 spins of the cluster and therefore the Hilbert space has the dimension $D_{\mathrm{H}} = 2^{15}$ (D_{H} for the three-spin system is 2^3). This approximation simplifies the evaluation of the low-lying energy levels of the 15 'entangled' states of the V_{15} cluster. For $D_Z \neq 0$ the $S = 1/2$ orbital doublet ^{2}E, whose basis functions can be labelled by the quantum number of the total pseudo-angular momentum $M_J = M_L + M_S$, is associated with the pseudo-orbital momentum $M_L = +1$ or $M_L = -1$ (refs 24, 25). The allowed EPR transitions satisfy the subsequent selection rules: $\Delta M_L = 0$, $\Delta M_S = \pm 1$, that is $\Delta M_J = \pm 1$ for the inter-doublet transitions 4 and 6, and $\Delta M_L = \mp 1$, $\Delta M_S = \pm 2$, that is $\Delta M_J = \pm 1$ for the weak intra-doublet transition 5 whose transition probability is caused by a small intermultiplet mixing through the in-plane component of the DM coupling. The intensity of this transition is significantly increased when transition 7 becomes allowed due to a weak deviation from the C_3 symmetry (Fig. 1). This also leads to an increased zero-field gap $\left[\Delta_{DM}^2 + \delta^2\right]^{1/2}$ where δ is the parameter in the exchange shift $\delta \mathbf{S}_1 \mathbf{S}_2$.

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

Received 28 January; accepted 1 April 2008.

1. Gatteschi, D., Sessoli, R. & Villain, J. *Molecular Nanomagnets* (Oxford Univ. Press, Oxford, UK, 2006).
2. Barbara, B. *et al.* Mesoscopic quantum tunneling of the magnetization. *J. Magn. Magn. Mater.* **140–144,** 1825–1828 (1995).
3. Thomas, L. *et al.* Macroscopic quantum tunneling of magnetization in a single crystal of nanomagnets. *Nature* **383,** 145–147 (1996).
4. Friedman, J. R. *et al.* Macroscopic measurements of resonant magnetization tunneling in high spin molecules. *Phys. Rev. Lett.* **76,** 3830–3833 (1996).
5. Leuenberger, M. N. & Loss, D. Quantum computing in molecular magnets. *Nature* **410,** 789–793 (2001).
6. Aharonov, D., Kitaev, A. & Preskill, J. Fault-tolerant quantum computation with long-range correlated noise. *Phys. Rev. Lett.* **96,** 050504 (2006).
7. Stamp, P. C. E. & Tupitsyn, I. S. Coherence window in the dynamics of quantum nanomagnets. *Phys. Rev. B* **69,** 014401 (2004).
8. Morello, A., Stamp, P. C. E. & Tupitsyn, I. S. Pairwise decoherence in coupled spin qubit networks. *Phys. Rev. Lett.* **97,** 207206 (2006).
9. Prokof'ev, N. V. & Stamp, P. C. E. Theory of the spin bath. *Rep. Prog. Phys.* **63,** 669–726 (2000).
10. Rabi, I. I. Space quantization in a gyrating magnetic field. *Phys. Rev.* **51,** 652–654 (1937).
11. Müller, A. & Döring, J. A novel heterocluster with D_3-symmetry containing twenty-one core atoms: $[As^{III}{}_6V^{IV}{}_{15}O_{42}(H_2O)]^{6-}$. *Angew. Chem. Int. Edn Engl.* **27,** 1721 (1988).
12. Gatteschi, D., Pardi, L., Barra, A. L., Müller, A. & Döring, J. Layered magnetic structure of a metal cluster ion. *Nature* **354,** 463–465 (1991).
13. Barbara, B. On the richness of supra-molecular chemistry and its openings in physics. *J. Mol. Struct.* **656,** 135–140 (2003).
14. Wernsdorfer, W., Müller, A., Mailly, D. & Barbara, B. Resonant photon absorption in the low spin molecule V_{15}. *Europhys. Lett.* **66,** 861–867 (2004).
15. Ardavan, A. *et al.* Will spin-relaxation times in molecular magnets permit quantum information processing? *Phys. Rev. Lett.* **98,** 057201 (2007).
16. Wernsdorfer, W. A long-lasting phase. *Nature Mater.* **6,** 174–176 (2007).
17. Mehring, M., Mende, J. & Scherer, W. Entanglement between an electron and a nuclear spin ½. *Phys. Rev. Lett.* **90,** 153001 (2003).
18. Morton, J. J. L. *et al.* Bang–bang control of fullerene qubits using ultrafast phase gates. *Nature Phys.* **2,** 40–43 (2006).
19. Bertaina, S. *et al.* Rare earth solid state qubits. *Nature Nanotechnol.* **2,** 39–42 (2007).
20. Nellutla, S. *et al.* Coherent manipulation of electron spins up to ambient temperatures in Cr^{5+} (S=1/2) doped K_3NbO_8. *Phys. Rev. Lett.* **99,** 137601 (2007).
21. Chiorescu, I., Wernsdorfer, W., Müller, A., Bögge, H. & Barbara, B. Butterfly hysteresis loop and dissipative spin reversal in the $S = 1/2$, V_{15} molecular complex. *Phys. Rev. Lett.* **84,** 3454–3457 (2000).
22. De Raedt, H. D., Miyashita, S., Michielsen, K. & Machida, M. Dzyaloshinskii-Moriya interactions and adiabatic magnetization dynamics in molecular magnets. *Phys. Rev. B* **70,** 064401 (2004).
23. Chaboussant, G. *et al.* Mechanism of ground-state selection in the frustrated molecular spin cluster V_{15}. *Europhys. Lett.* **66,** 423–429 (2004).
24. Tarantul, A., Tsukerblat, B. & Müller, A. Static magnetization of V_{15} cluster at ultra-low temperatures: Precise estimation of antisymmetric exchange. *Inorg. Chem.* **46,** 161–169 (2007).
25. Tsukerblat, B., Tarantul, A. & Müller, A. Low temperature EPR spectra of the mesoscopic cluster V_{15}: The role of antisymmetric exchange. *J. Chem. Phys.* **125,** 054714 (2006).
26. Dzyaloshinsky, I. A thermodynamic theory of "weak" ferromagnetism of antiferromagnetics. *J. Phys. Chem. Solids* **4,** 241–255 (1958).
27. Moriya, T. Anisotropic superexchange interactions and weak ferromagnetism. *Phys. Rev.* **120,** 91–98 (1960).
28. Volkmer, D. *et al.* Towards nanodevices: Synthesis and characterization of the nanoporous surfactant-encapsulated keplerate $(DODA)_{40}(NH_4)_2[(H_2O)_n \subset Mo_{132}O_{372}(CH_3COO)_{30}(H_2O)_{72}]$. *J. Am. Chem. Soc.* **122,** 1995–1998 (2000).
29. Prokof'ev, N. V. & Stamp, P. C. E. Quantum relaxation of magnetisation in magnetic particles. *J. Low Temp. Phys.* **104,** 143–210 (1996).
30. Hartmann-Boutron, F., Politi, P. & Villain, J. Tunneling and magnetic relaxation in mesoscopic molecules. *Int. J. Mod. Phys. B* **10,** 2577–2637 (1996).

Acknowledgements We acknowledge I. Chiorescu from NHMFL-FSU, Tallahassee, USA, for discussions. We thank M.-N. Collomb for help in processing samples for EPR measurements, and G. Desfonds for technical support. B.B. and A.M. thank the European Research Council for support through network projects MAGMANet, MolNanoMag, QueMolNa and INTAS; A.M. thanks the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie for support; and B.T. and A.M. thank the German–Israeli Foundation for Scientific Research and Development for support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.B. (bernard.barbara@grenoble.cnrs.fr) or A.M. (a.mueller@uni-bielefeld.de).

©2008 Nature Publishing Group

METHODS

Sample synthesis. 0.04 g (0.0175 mmol) of freshly prepared brown $K_6\left[V^{IV}_{15}As^{III}_6O_{42}(H_2O)\right]\cdot 8H_2O$ obtained as reported[10] was dissolved in 20 ml of degassed water. After addition of 25 ml of a (degassed) trichloromethane solution of [DODA]Br (1.10 g/1.75 mmol) the reaction medium was stirred under inert atmosphere. The stirring was continued until the olive-brown coloured aqueous layer turned colourless and the corresponding colour appeared in the organic phase. The organic layer was then quickly separated, put into an EPR tube and frozen to liquid nitrogen temperature. All operations were done in an inert atmosphere.

Comparing Rabi frequencies. The frequency of the Rabi oscillations between two states 1 and 2 is given by[6–8,19]:

$$\Omega_R = \Omega_{R0}\,|\langle\varphi_1|S_+|\varphi_2\rangle| \tag{2}$$

Here $\Omega_{R0} = 2g\mu_B B_1/h_{Planck} = 55.96B_1$ (MHz, mT) is the Rabi frequency of a spin 1/2, B_1 is the amplitude of the a.c. microwave fields, $g \approx 2$ the Landé factor, S_+ the ladder operator and $|\varphi_1\rangle$, $|\varphi_2\rangle$ the wavefunctions associated with these states. The probability of a transition, defined as $P = |\langle\varphi_1|S_+|\varphi_2\rangle|^2$, is directly connected with its Rabi frequency:

$$P = (\Omega_R/\Omega_{R0})^2 \tag{3}$$

This allows one to evaluate the ratio (R) of the probabilities associated with two transitions (here the '3/2' and '1/2' types) from the measurement of their Rabi frequencies without the knowledge of their wavefunctions:

$$R = P_{3/2}/P_{1/2} = \left(\Omega_{R\,3/2}/\Omega_{R\,1/2}\right)^2 \tag{4}$$

Using the values of the Rabi frequencies given in Fig. 3, one gets $R \approx (4.5/18.5)^2 \approx 5.9 \times 10^{-2}$. The time $T_{\pi/2}$, during which the excitation pulse is applied to induce a $\pi/2$ rotation, is by definition equal to $1/4\Omega_R$ (refs 6, 19), showing that equation (4) is equivalent to:

$$R = P_{3/2}/P_{1/2} = \left(T_{\pi/2,1/2}/T_{\pi/2,3/2}\right)^2 \tag{5}$$

This gives another way to determine R. Using the $T_{\pi/2}$ values given in Fig. 4 legend, one gets $R \approx (16/64)^2 \approx 6.2 \times 10^{-2}$, which is very close to the first one and shows that the probability associated with the '1/2' type transition is much smaller than the one associated with '3/2'.

©2008 Nature Publishing Group

CORRIGENDUM

doi:10.1038/nature09362

Quantum oscillations in a molecular magnet

S. Bertaina, S. Gambarelli, T. Mitra, B. Tsukerblat, A. Müller & B. Barbara

Nature **453**, 203–206 (2008)

The authors report that in a recent extension of this Letter, considering different concentrations, they were not able to reproduce the very weak signal of the ground-state EPR transition (the red curves termed '1/2' in Figs 3 and 4). The authors report that the signal is hardly observable (with intensity lower than that of the cavity background); but the more important $S = 3/2$ transitions (blue curves in Figs 3 and 4), which give rise to well-defined oscillations with several periods and high signal intensity, are reproducible. The extension of the work — about the concentration dependence of decoherence — will be published[1] but leads to the same major conclusion as in the original manuscript. The authors thank J. Du and colleagues for independently bringing the mentioned problem to their attention.

1. Shim, J. H. *et al.* Driven spin-bath decoherence in the molecular magnet V15. Preprint at ⟨http://arxiv.org/abs/1006.4960⟩ (2010).

©2010 Macmillan Publishers Limited. All rights reserved

LETTERS

Colossal cages in zeolitic imidazolate frameworks as selective carbon dioxide reservoirs

Bo Wang[1], Adrien P. Côté[1], Hiroyasu Furukawa[1], Michael O'Keeffe[2] & Omar M. Yaghi[1]

Zeolitic imidazolate frameworks (ZIFs) are porous crystalline materials with tetrahedral networks that resemble those of zeolites: transition metals (Zn, Co) replace tetrahedrally coordinated atoms (for example, Si), and imidazolate links replace oxygen bridges[1]. A striking feature of these materials is that the structure adopted by a given ZIF is determined by link–link interactions, rather than by the structure directing agents used in zeolite synthesis[2]. As a result, systematic variations of linker substituents have yielded many different ZIFs that exhibit known or predicted zeolite topologies. The materials are chemically and thermally stable, yet have the long-sought-after design flexibility offered by functionalized organic links and a high density of transition metal ions[1–7]. Here we report the synthesis and characterization of two porous ZIFs—ZIF-95 and ZIF-100—with structures of a scale and complexity previously unknown in zeolites[8–10]. The materials have complex cages that contain up to 264 vertices, and are constructed from as many as 7,524 atoms. As expected from the adsorption selectivity recently documented for other members of this materials family[3], both ZIFs selectively capture carbon dioxide from several different gas mixtures at room temperature, with ZIF-100 capable of storing 28 litres per litre of material at standard temperature and pressure. These characteristics, combined with their high thermal and chemical stability and ease of fabrication, make ZIFs promising candidate materials for strategies aimed at ameliorating increasing atmospheric carbon dioxide levels.

Since the initial observation[4] that crystalline imidazolate frameworks with structures reminiscent of those exhibited by zeolites can be readily synthesized, a number of porous ZIFs have been developed[1–3,5–7] by realizing that the 145° Si-O-Si bond angle in zeolites is close to the angle formed by imidazolates (IMs) when bridging transition metals (Fig. 1a). ZIFs reported so far—for example, those with RCSR (Reticular Chemistry Structure Resource) structure symbols **sod**, **cag**, **mer**, **crb**, **dft**, **gis**, **rho**, **gme** and **lta**—are predominantly simple uninodal (one type of tetrahedral centre) structures[11]. Whereas the structure of silicate-based zeolites is templated by structure directing agents (SDAs, typically organic amines) that remain as extra-framework entities, the structure adopted by a given ZIF predominantly depends on the type of IM and solvent used. Introduction of greater structural diversity to ZIF chemistry thus naturally begins with functionalization of IM linkers. In fact, it has been shown that link–link interactions can direct the formation of specific ZIF structures[2]. In the present case, an examination of 2-methylimidazolate (mIM) and benzimidazolate (bIM), which form ZIF-8 (structure symbol **sod**; Fig. 1b) and ZIF-11 (**rho**; Fig. 1c), respectively[1], and modelling studies of **sod** and **rho** type structures, indicated that functionalization of the 6-membered ring (6MR) at the 4 or 5 position of bIM overcrowds links and inhibits the formation of **rho** (Fig. 1d)[12]. Thus we reasoned that proceeding with readily available 5-chlorobenzimidazolate (cbIM) should produce new ZIF topologies by increasing the IM girth (Fig. 1).

ZIF-95 was synthesized and structurally characterized by X-ray diffraction, and was found to be tetragonal (see Methods and Supplementary Information Section S1). It has a neutral framework with all Zn nodes tetrahedrally coordinated by cbIM. The 4-coordinated net of linked Zn atoms has a topology we name '**poz**'. Salient features of the structure are two large cages (Fig. 2a–c): a $[3^{16}4^{28}8^{2}12^{4}]$ **poz** A cage with 8MR and 12MR faces, and a $[3^{32}4^{36}8^{2}10^{8}12^{4}]$ **poz** B cage, with 8MR, 10MR and 12MR faces (in the symbols $[...m^{n}...]$ means that there are n faces that are m-rings). There are also smaller $[4^{6}10^{4}]$ **poz** C cages and $[4^{6}]$ **poz** D cages located at the interstices between A and B cages (Figs 2a–c, 3a).

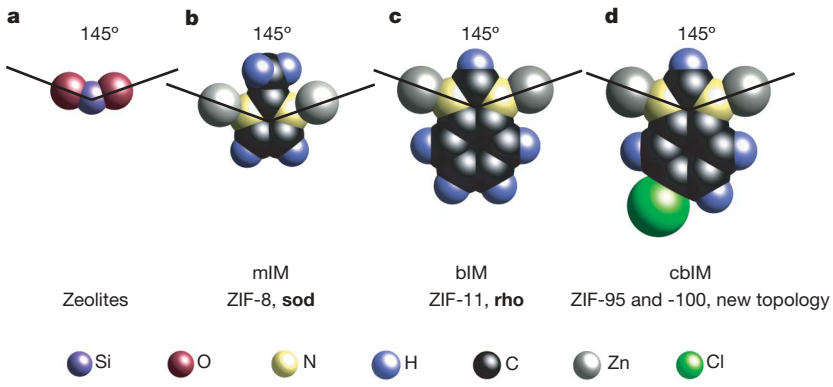


Figure 1 | Bridging angles and girths in zeolites and IMs. a, Si_2O in zeolites. **b**, Zn_2(2-methylimidazolate), mIM, in ZIF-8 (structure symbol **sod**). **c**, Zn_2(benzimidazolate), bIM, in ZIF-11 (**rho**). **d**, Zn_2(5-chlorobenzimidazolate), cbIM, in ZIF-95 and ZIF-100 (previously unknown topologies).

[1]Center for Reticular Chemistry, Department of Chemistry and Biochemistry, University of California-Los Angeles, 607 East Charles E. Young Drive, Los Angeles, California 90095, USA. [2]Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287, USA.

©2008 Nature Publishing Group

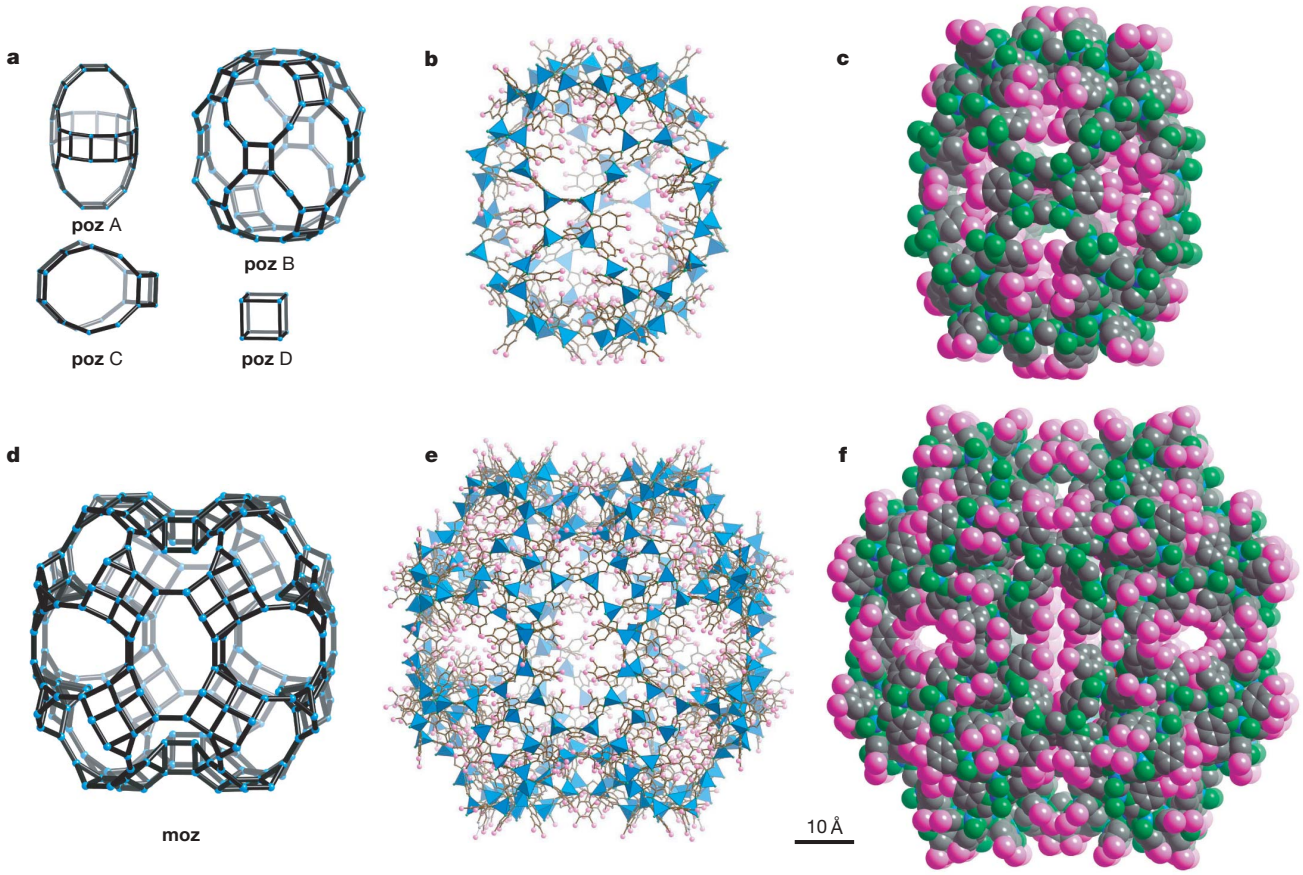


Figure 2 | The cages in ZIF-95 and ZIF-100. a, Ball and stick diagrams of the cages in ZIF-95; Zn in blue. **b**, Structure of **poz** B cage in ZIF-95; ZnN_4 tetrahedra are shown in blue, N in green, C in grey and Cl in pink. **c**, Space filling diagram of **poz** B cage in ZIF-95; Zn in blue, N in green, C in grey and Cl in pink. **d**, Ball and stick diagram of **moz** cage in ZIF-100. **e**, Structure of **moz** cage in ZIF-100; O in red. **f**, Space filling diagram of **moz** cage in ZIF-100; O in red. All H atoms are omitted for clarity. Note that Cl atoms fill one half of the positions, however the disorder is illustrated here.

The packing of these cages can be seen from the illustration of them as a tiling in Fig. 3a. In Table 1 we list the numbers of faces, edges and vertices in these cages. It may be seen that the A cage has the same size, by this measure, as the faujasite 'super cage', which serves as a benchmark for large cage zeolites; the B cage is significantly larger. The pores of ZIF-95 are of ellipsoidal shape: from the van der Waals surfaces, the A cage measures 25.1 × 14.3 Å and comprises 1,056 atoms (48 Zn, 504 C, 288 H, 144 N and 72 Cl). The B cage measures 30.1 × 20.0 Å and comprises 1,648 atoms (80 Zn, 784 C, 448 H, 224 N and 112 Cl) (Fig. 2b, c).

Restricting the water content of the reaction medium used for ZIF-95 synthesis promotes the crystallization of a new phase,

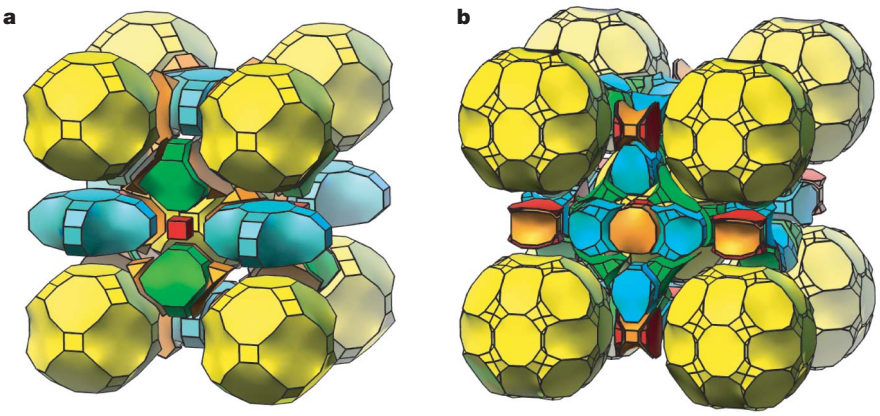


Figure 3 | The frameworks of ZIF-95 and ZIF-100 shown as natural tilings[27]. **a**, Natural tiling of ZIF-95; the A and B cages are shown in blue and yellow, respectively. **b**, Natural tiling of ZIF-100; the giant cage is in yellow.

©2008 Nature Publishing Group

Table 1 | Complexity of cages in FAU, MIL-101, mesoporous MOFs, and ZIFs

Material	Cages	Cage size (Å)*		Vertices	Faces	Edges	Metal atoms per cage
		Inner diameter	Outer diameter				
FAU	Super cage	14.1	18.1	48	26	72	48
ZIF-95	**poz** A	25.1 × 14.3	33.6 × 23.9	48	26	72	48
ZIF-95	**poz** B	30.1 × 20.0	41.1 × 33.9	80	34	112	80
MIL-101[13]	Large cage	33.8	46.7	28	16	42	126
Mes. MOF-1[14]	L cage	47.1	59.5	28	16	42	84
ZIF-100	**moz**	35.6	67.2	264	182	444	264

Mes., mesoporous. FAU, faujasite.
* Inner diameter of each cage is measured by the diameter of the largest sphere or ellipsoid that fits into the cavity; outer diameter is calculated from the centre-to-centre distance of the outmost atoms of each cage.

ZIF-100. Less hydrated conditions are achieved by using anhydrous $Zn(O_3SCF_3)_2$ in place of $Zn(NO_3)\cdot 4H_2O$. Control experiments rule out anions as SDAs (see Methods). Full understanding of the crystallization processes for ZIF-100 and ZIF-95 is difficult to attain, but the fact that cbIM leads to new ZIF structures, and not **rho**, supports our thesis that strategic link functionalization can access unrealized structural diversity. ZIF-100 was found to have the composition $Zn_{20}(cbIM)_{39}(OH)$, where one Zn atom (labelled Zn8, see Supplementary Information Section S1) is only tri-coordinated by cbIM.

X-ray diffraction techniques were used to determine the structure of ZIF-100. Initial X-ray photographs of the crystals clearly revealed a dense diffraction pattern indicative of a very large unit cell. ZIF-100 has $Im\bar{3}$ symmetry, with a unit cell parameter $a = 71.9797(4)$ Å and unit cell volume of 372,932(4) Å^3. Its primitive cell volume of 1.86×10^5 Å^3 makes ZIF-100 one of the largest inorganic structures ever determined: it exceeds the primitive cell volume (1.75×10^5 Å^3) of the previous record holder, a metal-organic framework (MOF) designated MIL-101 (ref. 13), although a structurally related and more recently reported MOF has an even larger cell[14].

Application of conventional small molecule chemical crystallographic methods was inadequate in achieving a structural solution for ZIF-100. Thus a new strategy was developed, which borrows from techniques and computational methods used for biological macromolecule crystallography (see Supplementary Information Section S1). Here, dual-space direct-methods procedures (Shake-and-Bake v2.0) produced initial phases from the observed structure factors (F_{obs}) and electron density maps identifying the positions of heavy atoms (Zn and disordered Cl)[15]. Positions of C, H and N were not revealed at this stage. From our previous ZIF studies[1] we knew that the nearest neighbour Zn–Zn distances measure 6.0(4) Å, which is close to the Zn–Cl distance of 6.0–6.5 Å expected from molecular modelling. This created a challenge for initial placement of cbIM C and N atoms, which was reconciled through systematic positioning of a rigid five-membered C_3N_2 IM ring and testing for all possible connections between heavy atom positions. Upon refinement, correct IM placement was identified by: (1) obtaining chemically logical bond lengths and angles, and (2) the emergence of peaks from subsequent Fourier difference maps corresponding to atoms from 6MRs of cbIM which are geometrically aligned to the positions of disordered Cl atoms. Given the magnitude of the ZIF-95 and ZIF-100 structures, their high porosity, and the population of their pores with highly disordered guest molecules, the final refined structures have expectedly elevated reliability factors compared to those from small molecule crystallography. However, bond lengths and angles in the final structures are statistically and chemically reasonable values. Indeed, the reliability factors are better than those generally obtained for biological crystal structures, indicating that our structural refinements accurately describe the composition and atomic connectivity of ZIF-95 and ZIF-100.

ZIF-100 has a beautifully complex structure constructed from simple building blocks (Figs 2d–f and 3b). Of the ten Zn atoms that form the (3,4)-coordinated net with the topology symbolized as **moz**, nine are connected to four cbIM linkers and the remaining one (Zn8) is connected to three linkers. The salient feature is a giant cage $[3^{48}4^{108}12^{26}]$ with 264 vertices (see Figs 2d–f and 3b, and Table 1), which is one of the most complex polyhedra ever found in an inorganic material. This **moz** cage (95.6 kDa) is built from 7,524 atoms—264 Zn, 3,604 C, 2,085 H, 26 O, 1,030 N, 515 Cl—and has a 67.2 Å outer sphere diameter and a 35.6 Å inner sphere diameter. The inner sphere diameter is determined by fitting a sphere from the centroid of the cage to the van der Waals surface of its wall. In comparison, the corresponding distances in the faujasite super cage are 18.1 Å and 14.1 Å (Table 1).

The remarkably large pore sizes prompted us to examine the structural and thermal stability of these ZIFs and their gas uptake behaviour. Thermal gravimetric analysis and powder X-ray diffraction studies revealed a thermal stability range up to 500 °C for both ZIF-95 and ZIF-100 (Supplementary Information Sections S2 and S4). Full activation of ZIF-95 and ZIF-100 was achieved by soaking crystals in acetone or methanol, respectively, for 72 h and heating at 50 °C for 10 h under reduced pressure (10^{-3} torr) and then at 100 °C for 12 h; their permanent porosity subsequently was proven using a N_2 or Ar adsorption measurement (Supplementary Information Section S5). The Langmuir surface areas were 1,240 m^2 g^{-1} and 780 m^2 g^{-1} for ZIF-95 and ZIF-100, respectively (Fig. 4a, b), which is more than double those of the most porous zeolites. The BET method[16] yielded surface areas of 1,050 m^2 g^{-1} and 595 m^2 g^{-1}, respectively.

Slow diffusion of gas was evident for both ZIFs during adsorption analysis. The slow diffusion can be ascribed to pore apertures that are not much larger than the kinetic diameter of many small gases and therefore restrict migration of the gases into the pores. The two ZIFs thus uniquely combine huge cavities (24.0 Å and 35.6 Å for ZIF-95 and ZIF-100, respectively) and highly constricted windows (largest apertures, 3.65 Å and 3.35 Å, respectively), which points to their possible utility in carbon dioxide capture and storage. It is now commonly believed that carbon dioxide emissions from burning of fossil fuels in power plants and automobiles are altering the temperature of the atmosphere and the acidity of the oceans with undesirable consequences for the Earth's environment[17]. It has been shown[3,18–22] that MOFs and ZIFs can hold large amounts of gases, including carbon dioxide, and that gas uptake can be selective. But it remains challenging to efficiently and selectively capture CO_2 from industrial emission streams (which invariably contain other gases such as CH_4, N_2 and so on); such a capability could allow for an efficient and economic reduction in CO_2 emissions[23,24].

Carbon dioxide, methane, carbon monoxide and nitrogen adsorption isotherms were measured for both ZIFs (Fig. 4b), which clearly show a disproportionately high affinity and capacity for CO_2 (Fig. 4b). ZIF-100 is seen to outperform ZIF-95 and the prototypical adsorbent BPL carbon, which is widely used in industry owing to its ease of desorption and regeneration[25]. We note that some zeolite materials outperform not only BPL carbon but also our ZIFs in terms of selective CO_2 adsorption[26], but regeneration difficulties mean that their suitability for practical use is uncertain[25] and they were therefore not included in the comparative adsorption measurements. Figure 4b also illustrates that both ZIFs show complete reversibility

©2008 Nature Publishing Group

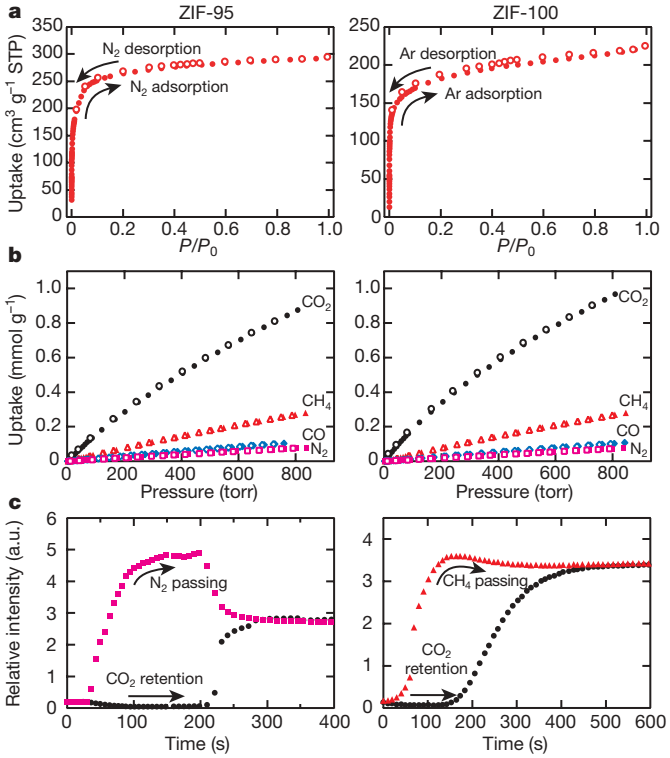


Figure 4 | Gas adsorption isotherms of ZIF-95 (left) and ZIF-100 (right). **a**, **b**, N_2 at 77 K for ZIF-95 (**a**, left); Ar at 87 K for ZIF-100 (**a**, right); CO_2 (black, circles), CH_4 (red, triangles), CO (blue, diamonds) and N_2 (pink, rectangles) at 298 K for ZIF-95 (**b**, left) and ZIF-100 (**b**, right); the filled and open shapes represent adsorption and desorption, respectively. P/P_0, relative pressure at the saturation vapour pressure of the adsorbate gas. **c**, Breakthrough curves, N_2 (pink, rectangles) and CO_2 (black, circles) for ZIF-95 using a CO_2/N_2 gas mixture (left); CH_4 (red, triangles) and CO_2 (black, circles) for ZIF-100 using a CO_2/CH_4 gas mixture (right). The relative intensities of each gas passing through the ZIF-95- and ZIF-100-packed column were obtained using a mass spectrometer to detect ion peaks at $m/z^+ = 44$ (CO_2), 16 (CH_4), 28 (N_2) and 12 (CO). Mean value from multiple parallel experiments ($n \geq 3$, standard deviation $\sigma \leq 5\%$ for all isotherms and $\leq 6\%$ for breakthrough curves) was adopted for each data point. Owing to the slow diffusion, ZIF-100 does not exhibit a stepped isotherm as expected for a mesoporous material.

of adsorption, indicating that they might serve as selective carbon dioxide reservoirs before the gas is processed further (for example, injected into an underground geologic formation). More detailed analysis of the data in Fig. 4b indicates that one litre of ZIF-100 can hold up to 28.2 l (55.4 g, or 1.7 mmol per g of ZIF-100) of CO_2 at 273 K and 15.9 l (31.2 g) at 298 K, with a single **moz** cage capturing as many as 121 CO_2 molecules at 273 K and 68 molecules at 298 K and ambient pressure. These adsorption characteristics are in line with the recent report[3] of 1 l of ZIF-69 storing 82.6 l of CO_2 at 273 K. The selectivity of ZIFs for carbon dioxide thus clearly exceeds that of BPL carbon (Table 2 and Supplementary Information Section S5; see also ref. 3). Some MOFs[20] have a higher overall capacity for carbon dioxide, but they do not show the level of selectivity observed with ZIFs.

Preliminary breakthrough experiments (see Methods and Supplementary Information Section S6) were also carried out to show that carbon dioxide can be held in the pores of ZIF-95 and ZIF-100 when exposed to streams containing binary mixtures of CO_2/CH_4, CO_2/CO or CO_2/N_2 (50:50 v/v). Figure 4c shows the results for CO_2/N_2 passing through 1.2 g of ZIF-95 and CO_2/CH_4 passing through 1.1 g of ZIF-100. It is very clear that in both cases only carbon dioxide is retained in the pores while the other gas passes through without hindrance; this observation lends further support to the idea of using ZIFs as selective carbon dioxide reservoirs. We attribute the high selectivity for carbon dioxide seen here to the combined effects of the slit width of the pore apertures being similar in size to carbon dioxide (kinetic diameter ~3 Å), and the strong quadrupolar interactions of carbon with nitrogen atoms present in the links.

Table 2 | Gas separation selectivity of ZIFs and BPL carbon

Material	Gas pairs	ZIF selectivity*[28]	BPL carbon selectivity[25]	Ratio ZIF/BPL carbon*
ZIF-100	CO_2/CH_4	5.9 ± 0.4	2.5	2.4 ± 0.2
ZIF-100	CO_2/CO	17.3 ± 1.5	7.5	2.3 ± 0.2
ZIF-100	CO_2/N_2	25.0 ± 2.4	11.1	2.3 ± 0.2
ZIF-95	CO_2/CH_4	4.3 ± 0.4	2.5	1.7 ± 0.2
ZIF-95	CO_2/CO	11.4 ± 1.1	7.5	1.5 ± 0.1
ZIF-95	CO_2/N_2	18.0 ± 1.7	11.1	1.6 ± 0.2

* See Supplementary Information Section S5.

METHODS SUMMARY

Synthesis of compounds. Light yellow crystalline plates of ZIF-95 (framework composition: $Zn(cbIM)_2$) were isolated from the reaction of a mixture containing $Zn(NO_3)_2 \cdot 4H_2O$, 5-chlorobenzimidazole (cbIM), *N*,*N*-dimethylformamide (DMF) and water at 120 °C (75.0% yield based on zinc). Precise control of water added and use of $Zn(O_3SCF_3)_2$ under the same conditions yields dark yellow cubic crystals of ZIF-100 (framework composition: $Zn_{20}(cbIM)_{39}(OH)$) (70.5% yield). These crystals were used for single-crystal X-ray diffraction studies. Both compounds were prepared as microcrystalline powder on the gram scale for examination of their porosity and selectivity to gases.

X-ray single-crystal diffraction studies. ZIF-95 is tetragonal: $I4/mmm$, $a = b = 38.7657(4)$ Å, $c = 56.454(1)$ Å; $V = 84{,}837(2)$ Å^3. Total reflections, 77,043; independent reflections, 3,666 (R(int) = 0.1660); parameters, 296; R1 = 0.1699, wR2 = 0.2750. ZIF-100 is cubic: $Im\bar{3}$, $a = 71.9797(4)$ Å, $V = 372{,}932(4)$ Å^3. Total reflections, 783,540; independent reflections, 25,973 (R(int) = 0.2642); parameters, 786; R1 = 0.1798, wR2 = 0.4315 (Supplementary Information Section S1).

Gas adsorption and separation measurements. The N_2 (for ZIF-95, 77 K) and Ar (for ZIF-100, 87 K) isotherms were measured for ZIF-95 and ZIF-100. ZIF-95 (ZIF-100): an apparent surface area of 1,240 (780) $m^2 g^{-1}$ and a micropore volume of 0.43 (0.37) $cm^3 g^{-1}$ were found (Supplementary Information Section S5). The gas-separation properties of ZIF-95 and ZIF-100 were derived from gas adsorption isotherms and by breakthrough experiments using CO_2/CH_4, CO_2/N_2 or CO_2/CO.

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

Received 11 October 2007; accepted 29 February 2008.

1. Park, K.-S. *et al.* Exceptional chemical and thermal stability of zeolitic imidazolate frameworks. *Proc. Natl Acad. Sci. USA* **103**, 10186–10191 (2006).
2. Hayashi, H., Côté, A. P., Furukawa, H., O'Keeffe, M. & Yaghi, O. M. Zeolite A imidazolate frameworks. *Nature Mater.* **6**, 501–506 (2007).
3. Banerjee, R. *et al.* High-throughput synthesis of zeolitic imidazolate frameworks and application to CO_2 capture. *Science* **319**, 939–943 (2008).
4. Tian, Y.-Q. *et al.* Two polymorphs of cobalt(II) imidazolate polymers synthesized solvothermally by using one organic template *N*,*N*-dimethylacetamide. *Inorg. Chem.* **43**, 4631–4635 (2004).
5. Tian, Y.-Q. *et al.* Design and generation of extended zeolitic metal-organic frameworks (ZMOFs): Synthesis and crystal structures of zinc(II) imidazolate polymers with zeolitic topologies. *Chem. Eur. J.* **13**, 4146–4154 (2007).
6. Huang, X.-C., Lin, Y.-Y., Zhang, J.-P. & Chen, X.-M. Ligand-directed strategy for zeolite-type metal-organic frameworks: Zinc(II) imidazolates with unusual zeolitic topologies. *Angew. Chem. Int. Edn* **45**, 1557–1559 (2006).
7. Liu, Y., Kravtsov, V. Ch., Larsena, R. & Eddaoudi, M. Molecular building blocks approach to the assembly of zeolite-like metal-organic frameworks (ZMOFs) with extra-large cavities. *Chem. Commun.* **14**, 1488–1490 (2006).
8. Baerlocher, C. & McCusker, L. B. *Database of Zeolite Structures* ⟨http://www.iza-structure.org/databases/⟩ (accessed, 11 October 2007).
9. Davis, M. E. Ordered porous materials for emerging applications. *Nature* **417**, 813–821 (2002).
10. Thomas, J. M. Design, synthesis and in situ characterization of new solid catalysts. *Angew. Chem. Int. Edn* **38**, 3588–3628 (1999).
11. Ockwig, N. W., Delgado-Friedrichs, O., O'Keeffe, M. & Yaghi, O. M. Reticular chemistry: Occurrence and taxonomy of nets and grammar for the design of frameworks. *Acc. Chem. Res.* **38**, 176–182 (2005).

©2008 Nature Publishing Group

12. Cerius2. *Modeling Environment* (Molecular Simulations, San Diego, California, 1999).
13. Férey, G. *et al.* A chromium terephthalate-based solid with unusually large pore volumes and surface area. *Science* **309,** 2040–2042 (2005).
14. Park, Y. K. *et al.* Crystal structure and guest uptake of a mesoporous metal-organic framework containing cages of 3.9 and 4.7 nm in diameter. *Angew. Chem. Int. Edn* **46,** 8230–8233 (2007).
15. Weeks, C. M. & Miller, R. The design and implementation of SnB version 2.0. *J. Appl. Crystallogr.* **32,** 120–124 (1999).
16. Walton, K. S. & Snurr, R. Q. Applicability of the BET method for determining surface areas of microporous metal-organic frameworks. *J. Am. Chem. Soc.* **129,** 8552–8556 (2007).
17. Houghton, J. T., Meira Filho, L. G., Griggs, D. J. & Noguer, M. *Implications of Proposed CO_2 Emissions Limitations* (Technical Paper IV, IPCC, Geneva, Switzerland, 1997).
18. Li, H., Eddaoudi, M., Groy, T. L. & Yaghi, O. M. Establishing microporosity in open metal-organic frameworks: Gas sorption isotherms for Zn(BDC) (BDC = 1,4-benzenedicarboxylate). *J. Am. Chem. Soc.* **120,** 8571–8572 (1998).
19. Millward, A. R. & Yaghi, O. M. Metal-organic frameworks with exceptionally high capacity for storage of carbon dioxide at room temperature. *J. Am. Chem. Soc.* **127,** 17998–17999 (2005).
20. Surblé, S. *et al.* Synthesis of MIL-102, a chromium carboxylate metal-organic framework, with gas sorption analysis. *J. Am. Chem. Soc.* **128,** 14889–14896 (2006).
21. Maji, T. K., Mostafa, G., Changa, H.-C. & Kitagawa, S. Porous lanthanide-organic framework with zeolite-like topology. *Chem. Commun.* **19,** 2436–2438 (2005).
22. Horike, S., Tanaka, D., Nakagawa, K. & Kitagawa, S. Selective guest sorption in an interdigitated porous framework with hydrophobic pore surfaces. *Chem. Commun.* **32,** 3395–3397 (2007).
23. Sircar, S. Basic research needs for design of adsorptive gas separation processes. *Ind. Eng. Chem. Res.* **45,** 5435–5448 (2006).
24. Sasaki, A., Matsumoto, S., Fujitsuka, M., Shinoki, T. & Tanaka, T. CO_2 recovery in molten carbonate fuel cell system by pressure swing adsorption. *IEEE Trans. Energy Convers.* **8,** 26–32 (1993).
25. Sircar, S., Golden, T. C. & Rao, M. B. Activated carbon for gas separation and storage. *Carbon* **34,** 1–12 (1996).
26. Sircar, S. & Golden, T. C. Isothermal and isobaric desorption of carbon dioxide by purge. *Ind. Eng. Chem. Res.* **34,** 2881–2888 (1995).
27. Blatov, V. A., Delgado-Friedrichs, O., O'Keeffe, M. & Proserpio, D. M. Three-periodic nets and tilings: natural tilings for nets. *Acta Crystallogr. A* **63,** 418–423 (2007).
28. Malek, A. & Farooq, S. Comparison of isotherm models for hydrocarbon adsorption on activated carbon. *Am. Inst. Chem. Eng. J.* **42,** 3191–3201 (1996).

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

Acknowledgements Funding was provided by BASF Ludwigshafen-Germany (synthesis), US Department of Defense (DTRA, separation instruments) and US Department of Energy (BES-Separation Program, gas separations). We acknowledge J. Eckert, D. Tranchemontagne and D. Britt (Yaghi group) for discussions.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Crystallographic data for the reported crystal structures have been deposited at the Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk with codes 668214 (ZIF-95) and 668215 (ZIF-100). Correspondence and requests for materials should be addressed to O.M.Y. (yaghi@chem.ucla.edu).

©2008 Nature Publishing Group

METHODS

General. 5-chlorobenzimidazole (cbIM), zinc(II) trifluoromethanesulphonate, $Zn(O_3SCF_3)_2$, anhydrous *N,N*-dimethylformamide (DMF), acetone and methanol were purchased from Aldrich Chemical Co.; imidazole and DMF were purchased from Fisher Scientific International, Inc. Zinc(II) nitrate tetrahydrate, $Zn(NO_3)_2{\cdot}4H_2O$, was purchased from EM Science. All starting materials were used without further purifications. Weighing and transferring zinc(II) trifluoromethanesulphonate was performed in a glove box. All the other experimental operations, unless otherwise noted, were performed in air. Single-crystal X-ray diffraction data were collected using a Bruker SMART APEX-II three-circle diffractometer equipped with a CCD (charge-coupled device) area detector operated at 1,200 W (40 kV, 30 mA) to generate Cu Kα radiation ($\lambda = 1.5417$ Å) (Supplementary Information Section S1). Powder X-ray diffraction data were collected using a Bruker D8 Discover θ–2θ diffractometer in reflectance Bragg–Brentano geometry at 40 kV, 40 mA (1,600 W) for Cu Kα_1 radiation ($\lambda = 1.5406$ Å) (Supplementary Information Section S2). Fourier-transform infrared spectra (FT-IR) of samples prepared as KBr pellets were measured using a Nicolet FT-IR Impact 400 system (Supplementary Information Section S3). Thermal gravimetric analysis was carried out using a TA Q500 thermal analysis system (Supplementary Information Section S4).

Synthesis of ZIF-95. A solid mixture of zinc(II) nitrate tetrahydrate ($Zn(NO_3)_2{\cdot}4H_2O$, 1.044 g, 4.0 mmol) and cbIM ($C_7H_5N_2Cl$, 6.106 g, 40.0 mmol) was dissolved in 295 ml DMF and 5 ml H_2O in a 500 ml container. The container was capped and heated at 120 °C for 72 h to give light yellow plate crystals. The reaction mixture was allowed to cool gradually to room temperature and crystals were washed with DMF and dried in air (yield: 1.103 g, 75% based on $Zn(NO_3)_2{\cdot}4H_2O$). Elemental microanalysis for $Zn(cbIM)_2 \equiv C_{14}H_8N_4Cl_2Zn_1$, calculated based on framework only (%): C, 45.63; H, 2.19; N, 15.20; Cl, 19.24; Zn, 17.74. Found (%): C, 46.06; H, 2.40; N, 14.38; Cl, 19.13; Zn, 18.17. FT-IR (KBr, 4,000–400 cm^{-1}): 3,426(br), 3,080(w), 2,935(w), 2,860(w), 1,671(vs), 1,612(m), 1,566(w), 1,475(vs), 1,393(m), 1,347(s), 1,291(s), 1,261(m), 1,240(m), 1,194(m), 1,130(w), 1,104(m), 1,067(m), 1,011(w), 930(m), 853(m), 807(s), 726(m), 655(m), 599(m), 488(m), 431(m).

Synthesis of ZIF-100. Under a dry nitrogen atmosphere a solid mixture of zinc(II) trifluoromethanesulphonate ($Zn(O_3SCF_3)_2$, 1.453 g, 4.0 mmol) and cbIM (6.103 g, 40.0 mmol) was dissolved in anhydrous DMF (270 ml) in a 500 ml container. Subsequently, H_2O (5.8 ml) was added, the container capped and heated in an isothermal oven at 120 °C for 48 h to give dark yellow cubic crystals. The reaction mixture was allowed to cool naturally to room temperature and the crystals were washed with DMF and dried in air. Yield: 1.012 g, 70.5% based on $Zn(O_3SCF_3)_2$. Elemental microanalysis for $Zn_{20}(cbIM)_{39}(OH) \equiv C_{273}H_{157}N_{78}O_1Cl_{39}Zn_{20}$, calculated based on framework only (%): C, 45.31; H, 2.19; N, 15.10; Cl, 19.11; Zn, 18.07. Found (%): C, 45.11; H, 2.18; N, 14.79; Cl, 19.19; Zn, 18.93. FT-IR (KBr, 4,000–400 cm^{-1}): 3,430(br), 3,080(m), 2,931(m), 2,860(w), 1,668(vs), 1,612(s), 1,576(m), 1,573(w), 1,475(vs), 1,437(m), 1,415(w), 1,342(s), 1,286(s), 1,256(m), 1,240(s), 1,195(s), 1,134(m), 1,098(m), 1,072(s), 1,014(w), 925(s), 859(m), 803(s), 762(w), 726(s), 659(m), 604(s), 487(m), 431(m).

Under these reaction conditions (metal, ligand, solvents, temperature, time, and so on), use of different counter anions (such as zinc(II) trifluoromethanesulphonate, nitrate, tetrafluoroborate, chloride) yield similar quantities of ZIF-100. However, tuning the water content of the reaction ultimately controls the yield of ZIF-100. Therefore anions are not SDAs.

Activation of ZIF-95 and ZIF-100. As-synthesized samples of ZIF-95 were immersed in anhydrous acetone in a glove box for 3 days; during the exchange the acetone was refreshed three times. The resulting acetone-exchanged sample of ZIF-95 was transferred as a suspension to a quartz cell in a glove box and the solvent was decanted. The wet sample was then evacuated (10^{-3} torr) at 50 °C for 10 h then at 100 °C for 12 h. ZIF-100 was identically activated, except that anhydrous methanol was used as the exchange solvent.

Gas adsorption and separation measurements. Low-pressure gas adsorption experiments (up to 850 torr) were carried out on a Quantachrome AUTOSORB-1 automatic volumetric instrument. Ultrahigh-purity-grade gases were used in all adsorption measurements. The N_2 (for ZIF-95, 77 K) and Ar (for ZIF-100, 87 K) isotherms were measured using a liquid nitrogen and an argon bath, respectively. CO_2, CH_4, CO and N_2 isotherms were also tested for both ZIF-95 and ZIF-100 at 273 and 298 K, respectively.

ZIF-95. An apparent surface area of 1,240 $m^2\,g^{-1}$ (Langmuir, the linearity of fitting, $R = 0.9999$) was obtained by using the data points on the adsorption branch of the N_2 isotherm in the range $P/P_0 = 0.1$–0.35 (Supplementary Information Section S5). A BET surface area of 1,050 $m^2\,g^{-1}$ was obtained by using $P/P_0 = 0.03$–0.08. A micropore volume of 0.43 $cm^3\,g^{-1}$ was obtained by applying non-local density functional theory (NLDFT) with a N_2-zeolites and silica model on the adsorption branch (the fitting error was 0.155%) (Supplementary Information Section S5).

ZIF-100. An apparent surface area of 780 $m^2\,g^{-1}$ ($R = 0.9998$) was obtained by using the data points on the adsorption branch of the Ar isotherm in the range $P/P_0 = 0.1$–0.3 (Supplementary Information Section S5). A BET surface area of 595 $m^2\,g^{-1}$ was obtained by using $P/P_0 = 0.03$–0.09. A micropore volume of 0.37 $cm^3\,g^{-1}$ was obtained by applying NLDFT with an Ar-zeolites and silica model on the adsorption branch (the fitting error was 0.095%) (Supplementary Information Section S5).

The gas-separation properties of ZIF-95 and ZIF-100 were tested by breakthrough experiments using a CO_2/CH_4, CO_2/N_2 or CO_2/CO (about 50:50 v/v) gas mixture. 1.2 g of activated ZIF-95 and 1.1 g of activated ZIF-100 were packed into a stainless-steel column (0.46 cm inner diameter and 8 cm length) in a glove box. The columns were then attached to gas-separation apparatus built as shown in Supplementary Fig. 32. Helium gas was used to initially purge the sample columns. All the experiments were carried out at room temperature. The gases (20 p.s.i.) were dosed into the column at a flow rate of 20 ml min^{-1}. The relative amounts of the gases passing through the column were monitored on a Hiden Analytical HPR20 gas analysis system detecting ion peaks at $m/z^+ = 44$ (CO_2), 16 (CH_4), 28 (N_2), 12 (CO).

©2008 Nature Publishing Group

Vol 453|8 May 2008|doi:10.1038/nature06960

LETTERS

Increasing risk of Amazonian drought due to decreasing aerosol pollution

Peter M. Cox[1,2], Phil P. Harris[3], Chris Huntingford[3], Richard A. Betts[2], Matthew Collins[2], Chris D. Jones[2], Tim E. Jupp[1], José A. Marengo[4] & Carlos A. Nobre[4]

The Amazon rainforest plays a crucial role in the climate system, helping to drive atmospheric circulations in the tropics by absorbing energy and recycling about half of the rainfall that falls on it. This region (Amazonia) is also estimated to contain about one-tenth of the total carbon stored in land ecosystems, and to account for one-tenth of global, net primary productivity[1]. The resilience of the forest to the combined pressures of deforestation and global warming is therefore of great concern[2], especially as some general circulation models (GCMs) predict a severe drying of Amazonia in the twenty-first century[3–5]. Here we analyse these climate projections with reference to the 2005 drought in western Amazonia, which was associated[6] with unusually warm North Atlantic sea surface temperatures (SSTs). We show that reduction of dry-season (July–October) rainfall in western Amazonia correlates well with an index of the north–south SST gradient across the equatorial Atlantic (the 'Atlantic N–S gradient'). Our climate model is unusual among current GCMs in that it is able to reproduce this relationship and also the observed twentieth-century multidecadal variability in the Atlantic N–S gradient[7], provided that the effects of aerosols are included in the model[8]. Simulations for the twenty-first century using the same model[3,8] show a strong tendency for the SST conditions associated with the 2005 drought to become much more common, owing to continuing reductions in reflective aerosol pollution in the Northern Hemisphere[9].

In 2005 large areas of the Amazon river basin experienced one of the most intense drought episodes of the last 100 years[6]. The drought most directly affected western Amazonia, and especially the catchments of the Solimões and the Madeiras rivers. Navigation along these major tributaries had to be suspended because the water levels were so low. Unlike during the El Niño-related droughts in 1926, 1983 and 1997, central and eastern Amazonia were not directly affected, although river levels on the Rio Negro in central Amazonia did reach unusually low levels in October 2005, owing to reduced inflow from the tributaries to the west.

Rainfall in Amazonia is sensitive to seasonal, interannual and decadal variations in SSTs[7,10,11]. The warming of the tropical East Pacific during El Niño events suppresses wet-season rainfall through modification of the (east–west) Walker circulation and via the Northern Hemisphere extratropics[12]. El Niño-like climate change[13] has similarly been shown to influence the annual mean rainfall over South America in GCM climate-change projections[4,5].

However, variations in Amazonian precipitation are also known to be linked to SSTs in the tropical Atlantic[11]. A warming of the tropical Atlantic in the north relative to the south leads to a northwestwards shift in the intertropical convergence zone and compensating atmospheric descent over Amazonia[10]. For northeast Brazil the relationship between the north–south Atlantic SST gradient and rainfall is sufficiently strong to form the basis for a seasonal forecasting system[14]. Here we show that Atlantic SSTs exert a large influence on dry-season rainfall in western Amazonia by delaying onset of the South American monsoon[15,16].

July to October 2005 was associated with a persistent warm anomaly in the North Atlantic[17] centred on latitudes 10–15° N, and a weaker cold anomaly in the South Atlantic at around latitude 15° S (Fig. 1). The absence of a warming in the tropical East Pacific implies that El Niño was in a near-neutral state and therefore did not contribute to the 2005 drought. The black box over land in Fig. 1 denotes the western Amazonian area chosen for the purposes of this study (75° W–60° W, 12° S–0°). The black boxes over ocean show the northern (15° N–35° N, 75° W–30° W) and southern (25° S–5° S, 40° W–5° E) regions used to calculate an index of the Atlantic N–S gradient (ANSG). These areas were chosen for comparability with the GCM climate projections that we present below, but are also near-optimal on the basis of a statistical analysis using observations alone[18].

We analyse results from the HadCM3LC coupled climate–carbon-cycle model[4], which is based upon the Met Office Hadley Centre's third-generation ocean–atmosphere GCM, HadCM3[19]. This is one of only three current, global climate models that fit within the observational limits on the July–October western Amazonian rainfall and the ANSG (see Supplementary Information). The HadCM3 model has also performed well in GCM intercomparison exercises and was recently recognized to be one of two GCMs that simulate the Amazonian climate with reasonable accuracy[5]. HadCM3LC in addition includes dynamic vegetation and an interactive carbon cycle,

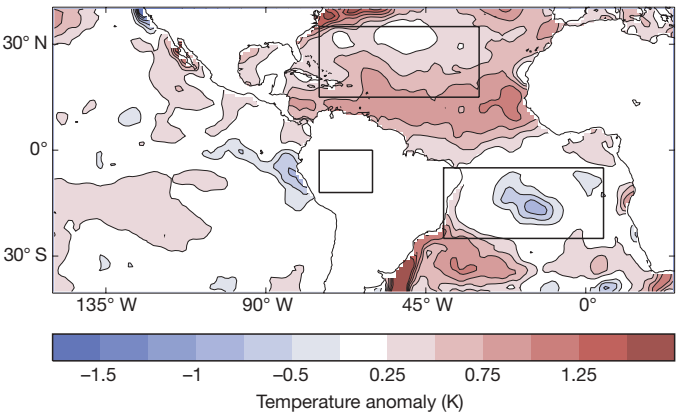


Figure 1 | Anomalies in SSTs for July–October 2005, relative to the July–October mean values over the standard climatological period 1961–1990. The black boxes show the regions used in this study.

[1]School of Engineering, Computing and Mathematics, University of Exeter, Exeter EX4 4QF, UK. [2]Met Office Hadley Centre, Exeter EX1 3PB, UK. [3]Centre for Ecology and Hydrology, Wallingford OX10 8BB, UK. [4]Brazilian Centre for Weather Forecasting and Climate Studies, CPTEC/INPE, Sao Paulo, Brazil.

©2008 Nature Publishing Group

meaning that atmospheric CO_2 concentrations can be updated on the basis of anthropogenic emissions, taking into account the effects of climate change on ocean and land CO_2 uptake.

We consider two separate simulations using HadCM3LC for the period from 1860 to 2100 (refs 4, 8). In both cases the model is driven with CO_2 emissions consistent with the IS92a scenario, which is approximately in the centre of the spread of future emissions represented by the results of the more recent Special Report on Emissions Scenarios[20]. Both model experiments also include prescribed time-varying concentrations of trace greenhouse gases (CH_4, N_2O) based on the IS92a scenario. The second run additionally includes changes in tropospheric and stratospheric ozone, solar variability and, most notably, forcing due to sulphate and volcanic aerosols. This model experiment was able to reproduce the observed warming and CO_2 increase over the twentieth century to good accuracy, especially when we used a revised estimate of the net CO_2 flux from land-use change (run 'ALL70' from ref. 8).

We compared the results of these simulations with observations of the ANSG index[17] and rainfall in western Amazonia[21] for the twentieth century, using 20-yr running means in each case (Fig. 2). The 2005 ANSG index value of 4.92 K was exceeded regularly during the 1930s, 1940s and 1950s, but has not reached such a high value since 1960. There is significant variability in the 20-yr-mean ANSG index (which is much larger than the standard deviation of the annual-mean values), with values declining from around 1960 to the mid-1980s before increasing in the past two decades. This variation is mirrored by the 20-yr mean western Amazonian rainfall measurements, which decreased from the 1920s to 1960 and then increased until the mid-1980s before decreasing again to the values observed today.

Both HadCM3LC simulations capture the observed inverse relationship between the ANSG index and July–October precipitation in the western Amazon. However, the simulation with greenhouse gases only (Fig. 2, red lines) fails to reproduce the observed decadal variability in the ANSG index (the correlation coefficient between the 20-yr running means of observed and modelled ANSGs is −0.75), and instead predicts a near-monotonic increase in the ANSG index and a corresponding near-monotonic decrease in western Amazonian rainfall from the mid-twentieth century onwards. On the other hand, the HadCM3LC run with aerosols (Fig. 2, green lines) produces a good fit to the multi-decadal variation in the ANSG index (the correlation coefficient between the 20-yr running means is in this case +0.82), and reproduces the observed decrease in 20-yr mean western Amazonian rainfall from the 1920s to 1960.

This suggests that longer-term variations in the ANSG, and associated effects on Amazonian rainfall, are as much a consequence of forced variability as they are of internal variability. Furthermore, the two model runs taken together indicate that non-greenhouse-gas forcing, primarily from anthropogenic and volcanic aerosols, has acted to suppress the development of a stronger ANSG, and thereby delayed the suppression of dry-season rainfall in the western Amazon. It seems that reflective sulphate aerosol pollution produced in the Northern Hemisphere may have helped to maintain rainfall in South America, just as it may have contributed to the Sahelian droughts of the 1970s and 1980s[22].

The HadCM3LC model also produces a realistic correlation between interannual variations in the western Amazonian rainfall and the ANSG index (Fig. 3). The best-fit straight lines linking these variables for the model and the observations are very nearly parallel, having respective gradients of -0.65 ± 0.29 and -0.58 ± 0.46 mm d^{-1} K^{-1} (throughout the paper we quote error bars as ±2 s.d. giving approximately 95% confidence limits). The HadCM3LC model validates best among current GCMs in this respect (see Supplementary Information). For the twenty-first century, the model predicts a strengthening of the relationship between the (decreasing) western Amazonian rainfall for the July–October period and the (increasing) ANSG index, as indicated by the quadratic best-fit shown in Fig. 3.

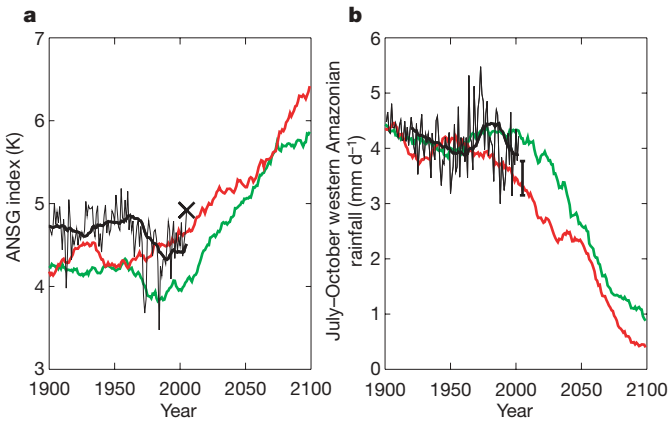


Figure 2 | Comparison of observed and modelled climate variables relevant to the Amazonian drought of 2005. Evolution of July–October mean values of **a**, the ANSG index; **b**, rainfall in western Amazonia. Observations are shown in black[17,20], with the thin lines corresponding to annual values and the thick lines showing 20-yr running means. The other lines show 20-yr running means from the HadCM3LC GCM. The red line corresponds to a simulation of climate change driven by greenhouse gas increases only[3], whereas the green line additionally includes changes in aerosols, stratospheric and tropospheric ozone, and revised fluxes of CO_2 from land-use change[8]. The large cross in **a** and the bar in **b** show the estimated values for July–October 2005.

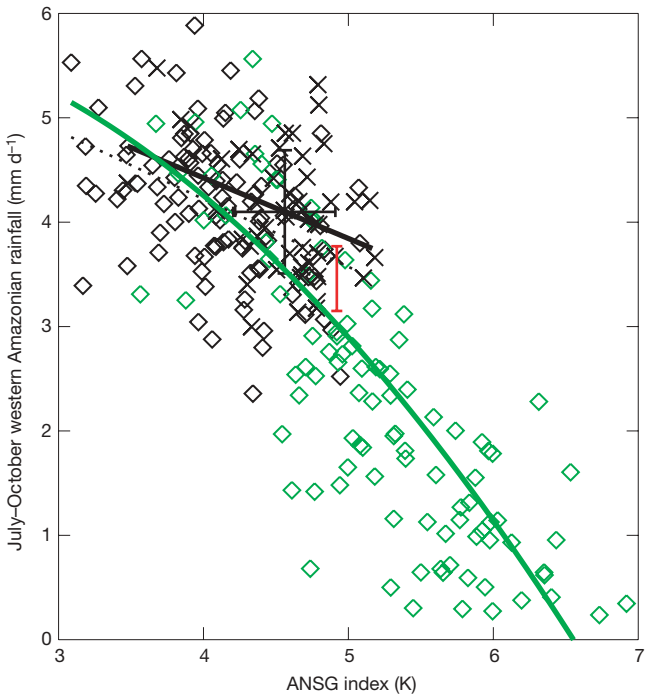


Figure 3 | Relationship between the July–October mean values of western Amazonian rainfall and the ANSG index. Observations for the period 1901–2002 are plotted with black crosses[17,20]. Model output from the HadCM3LC GCM run with aerosols[8] is plotted with black diamonds for the historical period (1901–2002) and with green diamonds for the simulation of the twenty-first century (2003–2100). The black lines are the best-fit straight lines to the observations (1950–1999; solid) and the twentieth-century simulation (1900–1999; dashed). The green line is the best quadratic fit to the entire GCM simulation (1860–2099). The large black cross shows the mean and standard deviation of the observations, and the red bar shows the range of estimated values for the 2005 Amazon drought.

©2008 Nature Publishing Group

Our simulations suggest that the North Atlantic region will warm more rapidly than the South Atlantic in the future, leading to a northwards movement of the intertropical convergence zone and suppression of July–October rainfall in western Amazonia. Aerosol pollution has occurred predominantly in the industrialised north, which tends to suppress the development of this north–south warming gradient. However, as air quality improves, aerosol cooling of the climate is expected to decrease[9], potentially revealing a larger ANSG. This is evident in the HadCM3LC run with aerosols (Fig. 2, green lines), which predicts a 2 K increase in the ANSG index by the end of the twenty-first century. As a consequence, this GCM projection suggests that the conditions of 2005 will be experienced more and more frequently as atmospheric greenhouse gas concentrations increase and sulphur dioxide emissions decrease in the Northern Hemisphere (see Supplementary Information).

Climate model projections are known to differ markedly with respect to regional rainfall changes over Amazonia, but a comparison of the results from 20 GCMs included in the Fourth Assessment Report of the Intergovernmental Panel on Climate Change reveals two robust features: there is a cross-model relationship between the twenty-first-century trends in western Amazonian rainfall and the twenty-first-century trends in the ANSG index, which is consistent with the observed interannual variability in these variables; and models that include aerosol forcing tend to predict greater increases in the ANSG index in the first few decades of the twenty-first century (see Supplementary Fig. S3). Taken together, these findings support our basic conclusion that aerosol forcing has delayed reductions in Amazonian rainfall but is unlikely to do so for much longer.

We estimated the probability of a '2005-like' year occurring in the HadCM3LC run with aerosols, based on the fraction of years in a centred 20-yr window that exceed the ANSG index for 2005 (Fig. 4). The model suggests that 2005 was an approximately 1-in-20-yr event, but will become a 1-in-2-yr event by 2025 and a 9-in-10-yr event by 2060. These thresholds obviously depend on the rate of increase of CO_2, which is itself dependent on the emissions scenario chosen. Figure 4b shows how the probability of a 2005-like event increases as a function of CO_2 concentration in this HadCM3LC model, with the 50% probability level exceeded at about 450 p.p.m.v. and the 90% probability level exceeded at around 610 p.p.m.v. These results suggest a rapidly increasing risk of 2005-like droughts in Amazonia under conditions of reduced aerosol loading and increased greenhouse gases.

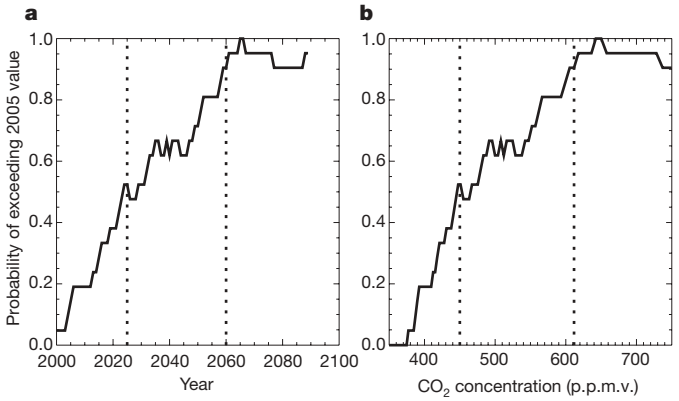


Figure 4 | Predicted change in the probability of a 2005-like drought in Amazonia, based on results from the HadCM3LC GCM run with aerosols[8]. Probabilities are calculated as the fraction of years that exceed the July–October 2005 anomaly in the ANSG index, using the 20-yr window centred on each year: **a**, probability versus year; **b**, probability versus simulated CO_2 concentration. The dotted lines respectively mark the points beyond which the 2005 anomaly is exceeded in 50% of the simulated years (2025, 450 p.p.m.v. of CO_2) and in 90% of the simulated years (2060, 610 p.p.m.v. of CO_2).

METHODS SUMMARY

The climate model used was HadCM3LC—a version of HadCM3[19] that includes an interactive carbon cycle[3]. Two model runs were carried out: one with greenhouse gases only[3], and one with greenhouse gases plus other key climate forcing factors including anthropogenic sulphate aerosols[8]. In both runs, atmospheric CO_2 concentration was modelled interactively by specifying emissions taken from the IS92a scenario and allowing the carbon-cycle model to partition these between oceanic, terrestrial and atmospheric carbon pools. Non-CO_2 greenhouse gases were prescribed to follow a standard IS92a concentration scenario[3]. The 'aerosol run' also included revised net CO_2 emissions from land use, solar variability, tropospheric and stratospheric ozone changes, and the climatic effects of volcanic and anthropogenic aerosols[8].

To investigate the relative importance of Atlantic and Pacific SST variabilities for Amazonian rainfall, we constructed SST indices from the Hadley Centre's HadISST data set[17]. The Pacific index was defined as the SST difference between the equatorial West Pacific (5° S–5° N, 120° E–180° E) and the Niño-3 region (5° S–5° N, 150° W–90° W) to remove the impacts of global warming. The Atlantic index used was the ANSG index, defined as the SST difference between the tropical North Atlantic (15° N– 35° N, 75° W–30° W) and the tropical South Atlantic (25° S–5° S, 40° W–15° E). Time series of precipitation from the Climatic Research Unit[21] (CRU TS 2.0) were calculated for western Amazonia (75° W–60° W, 12° S–0°). Linear regression analyses were carried out to determine the relative sensitivity of the western Amazonian rainfall to the Atlantic and Pacific SST indices. Rainfall in the dry season (July–October) was found to be correlated with variability in the ANSG index, with no significant dependence on the east–west Pacific SST gradient[16]. This analysis therefore allowed us to focus on the Atlantic SST anomalies when seeking to understand the Amazonian drought of July–October 2005.

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

Received 23 January; accepted 3 April 2008.

1. Melillo, J. M. *et al.* Global climate change and terrestrial net primary production. *Nature* **363,** 234–240 (1993).
2. Tian, H. *et al.* Climatic and biotic controls on annual carbon storage in Amazonian ecosystems. *Glob. Ecol. Biogeogr.* **9,** 315–335 (2000).
3. Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A. & Totterdell, I. J. Acceleration of global warming due to carbon cycle feedbacks in a coupled climate model. *Nature* **408,** 184–187 (2000).
4. Cox, P. M. *et al.* Amazon dieback under climate-carbon cycle projections for the 21st century. *Theor. Appl. Climatol.* **78,** 137–156 (2004).
5. Li, W., Fu, R. & Dickinson, R. E. Rainfall and its seasonality over the Amazon in the 21st century as assessed by the coupled models for the IPCC AR4. *J. Geophys. Res.* **111,** D02111 (2006).
6. Marengo, J. A. *et al.* The drought of Amazonia in 2005. *J. Clim.* **21,** 495–516 (2008).
7. Marengo, J. A. Interdecadal variability and trends of rainfall across the Amazon basin. *Theor. Appl. Climatol.* **78,** 79–96 (2004).
8. Jones, C. D., Cox, P. M., Essery, R. L. H., Roberts, D. L. & Woodage, M. Strong carbon cycle feedbacks in a climate model with interactive CO_2 and sulphate aerosols. *Geophys. Res. Lett.* **30,** doi:10.1029/2003GL016867 (2003).
9. Andreae, M. O., Jones, C. J. & Cox, P. M. Strong present-day aerosol cooling implies a hot future. *Nature* **435,** 1187–1190 (2005).
10. Fu, R., Dickinson, R. E., Chen, M. X. & Wang, H. How do tropical sea surface temperatures influence the seasonal distribution of precipitation in the equatorial Amazon? *J. Clim.* **14,** 4003–4026 (2001).
11. Liebmann, B. & Marengo, J. Interannual variability of the rainy season and rainfall in the Brazilian Amazon basin. *J. Clim.* **14,** 4308–4318 (2001).
12. Nobre, P. & Shukla, J. Variations of sea surface temperature, wind stress and rainfall over the tropical Atlantic and South America. *J. Clim.* **9,** 2464–2479 (1996).
13. Meehl, G. A. & Washington, W. M. El Nino-like climate change in a model with increased atmospheric CO_2 concentrations. *Nature* **382,** 56–60 (1996).
14. Folland, C. K., Colman, A. W., Rowell, D. P. & Davey, M. K. Predictability of Northeast Brazil rainfall and real-time forecast skill 1987–1998. *J. Clim.* **14,** 1937–1958 (2001).
15. Marengo, J. A., Liebmann, B., Kousky, V. E., Filizola, N. P. & Wainer, I. C. Onset and end of the rainy season in the Brazilian Amazon Basin. *J. Clim.* **14,** 833–852 (2001).
16. Harris, P., P, Huntingford C. & Cox, P. M. Amazon Basin climate under global warming: the role of the sea surface temperature. *Phil. Trans. R. Soc. B* **363,** 1753–1759 (2008).
17. Rayner, N. A. *et al.* Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophy. Res.* **108,** doi:10.1029/2002JD002670 (2003).
18. Good, P., Lowe, J., Collins, M. & Moufouma-Okia, W. An objective tropical Atlantic sea surface temperature gradient index for studies of South Amazon

©2008 Nature Publishing Group

dry-season climate variability and change. *Phil. Trans. R. Soc. B* **363,** 1761–1766 (2008).
19. Gordon, C. *et al.* Simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.* **16,** 147–168 (2000).
20. Nakicenovic, N. *et al. Special Report on Emissions Scenarios, Summary for Policy Makers* (Intergovernmental Panel on Climate Change, Geneva, Switzerland, 2000).
21. New, M., Hulme, M. & Jones, P. Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Clim.* **13,** 2217–2238 (2000).
22. Rotstayn, L. D. & Lohmann, U. Tropical rainfall trends and the indirect aerosol effect. *J. Clim.* **15,** 2103–2116 (2002).

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

Acknowledgements The authors acknowledge funding from the NERC CLASSIC programme and Great Western Research (P.M.C. and T.E.J.); the CEH Science Budget (C.H. and P.P.H.); the UK Department for Environment, Food and Rural Affairs and the UK Ministry of Defence (R.A.B, C.D.J. and M.C.); and the Brazilian Research Council and the Global Opportunities Fund from the UK Foreign and Commonwealth Office (C.A.N. and J.A.M.). We also acknowledge the modelling groups the Program for Climate Model Diagnosis and Intercomparison and the WCRP Working Group on Coupled Modelling for their roles in making available the WCRP CMIP3 multimodel data set. Support for this data set is provided by the Office of Science, US Department of Energy.

Author Contributions P.M.C. coordinated the work, identified the role of aerosols in delaying Amazonian drying in HadCM3LC, and drafted the paper; P.P.H. and C.H. defined the SST indices and analysed the relationships between these indices and western Amazonian rainfall, in both the observations and the model runs; C.D.J. and R.A.B. carried out the HadCM3LC runs and provided output data from them; J.A.M. and C.A.N. provided observational data and insights on the nature of the 2005 drought in western Amazonia; M.C. extracted and then intercompared data on predicted changes in Amazonia from the CMIP3 models; C.H. and T.E.J. provided guidance on statistical significance.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.M.C. (p.m.cox@exeter.ac.uk).

©2008 Nature Publishing Group

METHODS

Climate simulations. The climate model used was HadCM3LC—a version of HadCM3[19], but with a slightly lower ocean resolution (2.5° latitude by 3.75° longitude), which is necessary because of the extra computational expense of the oceanic carbon-cycle model. As a result, the model requires the use of flux corrections. The HadCM3LC model includes interactive terrestrial and oceanic carbon-cycle components. The terrestrial carbon-cycle model, TRIFFID, is a dynamic global vegetation model, which models carbon allocation between five competing plant functional types and a soil carbon reservoir[23]. The oceanic carbon-cycle model, HadOCC, includes a representation of oceanic chemistry and biology[24].

In both model runs, atmospheric CO_2 concentration was modelled interactively and non-CO_2 greenhouse gases were prescribed to follow a standard IS92a concentration scenario[3]. The aerosol run also included the following six features[8].

The direct effect of anthropogenic sulphate aerosols was calculated using an interactive sulphur-cycle scheme[25], driven by historic anthropogenic sulphur dioxide emissions and twenty-first-century anthropogenic emissions according to the A2 scenario from the Special Report on Emissions Scenarios[20]. Dry deposition was calculated separately for each of the five vegetation types in each grid box, using CO_2-dependent canopy resistance values calculated according to the MOSES-2 land-surface scheme[26].

The first indirect effect of anthropogenic sulphate (that is, the cloud albedo effect) was included using a non-interactive method in which cloud albedo perturbations are imposed on the basis of output from a set of preliminary sulphur-cycle runs[25].

Climate forcing due to volcanic eruptions was represented by specifying the stratospheric aerosol distribution up to present day[27], and assuming that it is zero thereafter.

Tropospheric and stratospheric ozone changes were prescribed on the basis of results from the atmospheric chemistry model STOCHEM[28].

Solar forcing up to present day was taken from a reconstruction of solar irradiances[29], and kept constant (equal to the mean of the last 11 years) after 2000.

Revised net CO_2 emissions from land-use change[8] were used.

23. Cox, P. M. *Description of the "TRIFFID" dynamic global vegetation model.* Technical Note 24 (Met Office Hadley Centre, UK, 2001); ⟨http://www.metoffice.gov.uk/research/hadleycentre/pubs/HCTN/HCTN_24.pdf⟩.
24. Palmer, J. R. & Totterdell, I. J. Production and export in a global ocean ecosystem model. *Deep-Sea Res. I* **48,** 1169–1198 (2001).
25. Johns, T. C. *et al.* Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios. *Clim. Dyn.* **20,** 583–612 (2003).
26. Essery, R. L. H., Best, M. J., Betts, R. A., Cox, P. M. & Taylor, C. M. Explicit representation of sub-grid heterogeneity in a GCM land-surface scheme. *J. Hydrometeorol.* **4,** 530–543 (2003).
27. Sato, M., Hansen, J. E., McCormick, M. P. & Pollack, J. B. Stratospheric aerosol optical depths (1850–1990). *J. Geophys. Res.* **98,** 22987–22994 (1993).
28. Stevenson, D. S., Johnson, C. E., Collins, W. J., Derwent, R. G. & Edwards, J. M. Future estimates of tropospheric ozone radiative forcing and methane turnover: The impact of climate change. *Geophys. Res. Lett.* **27,** 2073–2076 (2000).
29. Lean, J., Beer, J. & Bradley, R. Reconstruction of solar irradiance since 1610: Implications for climate change. *Geophys. Res. Lett.* **22,** 3195–3198 (1995).

©2008 Nature Publishing Group

LETTERS

Scale effects and human impact on the elevational species richness gradients

D. Nogués-Bravo[1,2,3], M. B. Araújo[1,2], T. Romdal[2] & C. Rahbek[2]

Despite two centuries of effort in characterizing environmental gradients of species richness in search of universal patterns, surprisingly few of these patterns have been widely acknowledged[1–3]. Species richness along altitudinal gradients was previously assumed to increase universally from cool highlands to warm lowlands, mirroring the latitudinal increase in species richness from cool to warm latitudes[1,4,5]. However, since the more recent general acceptance of altitudinal gradients as model templates for testing hypotheses behind large-scale patterns of diversity[5–9], these gradients have been used in support of all the main diversity hypotheses, although little consensus has been achieved. Here we show that when resampling a data set comprising 400,000 records for 3,046 Pyrenean floristic species at different scales of analysis (achieved by varying grain size and the extent of the gradients sampled), the derived species richness pattern changed progressively from hump-shaped to a monotonic pattern as the scale of extent diminished. Scale effects alone gave rise to as many conflicting patterns of species richness as had previously been reported in the literature, and scale effects lent significantly different statistical support to competing diversity hypotheses. Effects of scale on current studies may be affected by human activities, because montane ecosystems and human activities are intimately connected[10]. This interdependence has led to a global reduction in natural lowland habitats, hampering our ability to detect universal patterns and impeding the search for universal diversity gradients to discover the mechanisms determining the distribution of biological diversity on Earth.

Studies of altitudinal gradients in species richness have increasingly replaced the latitudinal gradient as a model template for large-scale gradient studies[9]. Altitudinal gradients encompass several gradients in climatic and environmental factors, such as area, net primary productivity and geometric constraints. These factors are expected to influence spatial variation in species richness (Supplementary Fig. 1) but are often correlated, making hypothesis testing problematic and controversial[3]. However, these very controversies make altitudinal gradients an illuminating field of study. A recent quantitative analysis of altitudinal species richness gradients including 204 data sets demonstrated that about 50% of the pattern distributions were hump-shaped, about 25% showed a monotonically decreasing pattern, and about 25% followed other distributions[9]. It has therefore been suggested that non-generality in altitudinal species richness patterns may be a result of differences in spatial design between studies[9]. These differences include the choice of grain size and the extent and proportion of gradients sampled. Nevertheless, statistical correlations between these diverse patterns and associated patterns of climate[11,12], area[8,13,14] and, more recently, geometric constrains[8,15] have been used as support for competing hypotheses[5,9,13,16,17].

In this study we used an extensive data set comprising 400,000 records covering 3,046 species of vascular plants, lichens and bryophytes from the Pyrenees to illustrate and evaluate the sensitivity of patterns to scale effects (see Methods). Scale effects were evaluated by re-sampling the data set and generating altitudinal species richness patterns after changes in grain size (that is, the resolution at which data are sampled) and the scale of extent (that is, the proportion of the complete altitudinal gradient sampled). In association with scale of extent, we also evaluated the effect of omitting segments from the lowest or highest ends of the gradient.

The relationship between species richness and altitude varied greatly with scale of extent (Fig. 1). When the entire elevational gradient was surveyed, the pattern was hump-shaped (top row in Fig. 1), changing progressively to a monotonically decreasing pattern as the scale of extent diminished. This trend was particularly apparent when the lower limit of the gradient was excluded from the analyses. When the upper limit of the gradient was excluded, the hump-shaped pattern was less sensitive to changes, although a monotonic increase in richness with altitude ultimately became apparent (Supplementary Fig. 2). This pattern has previously, although infrequently, been reported[5,9]. Regardless of which gradient segment was omitted, grain size did not markedly affect changes in species richness with elevation (Fig. 1 and Supplementary Fig. 2). This 'negative' result is noteworthy because variation in grain size has previously been shown to significantly influence the relative importance of factors determining large-scale continental patterns of species richness[18].

The implications of these scale effects for the assessment of competing diversity hypotheses were evaluated statistically. The empirical data on species richness were compared with predicted data generated by four well-documented diversity models developed to explain altitudinal and environmental species richness gradients[13]. Model 1 is a monotonic species-richness–productivity model in which productivity and, consequently, species richness are assumed to decrease with altitude; model 2 is a monotonic species-richness–area model in which area and, consequently, species richness are assumed to decrease with altitude; model 3 is a hump-shaped species richness-productivity model in which productivity is assumed to decrease with altitude and species richness is assumed to peak within the lower half of the gradient; and model 4 is a mid-domain-effect model with a peak in richness in the middle of the gradient as a consequence of geometric constraints and two hard boundaries. Because the four models are based on generalized functions, it is possible to choose the function that suits any specific pattern relevant to a given data set; for example, if most of the area occurs at mid-altitude regions, model 3 or 4 will be better suited to illustrate how scale effects may influence the interpretation of empirical analyses (see Methods for details, and Supplementary Fig. 1 for additional details on the four models).

[1]Department of Biodiversity and Evolutionary Biology, National Museum of Natural Sciences, CSIC, C/ José Gutiérrez Abascal, 2, 28006 Madrid, Spain. [2]Center for Macroecology, Institute of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark. [3]Pyrenean Institute of Ecology, CSIC, Avda. Montañana, 1005, 50080 Zaragoza (Zaragoza), Spain.

©2008 Nature Publishing Group

As expected, the correlation was best between the empirical hump-shaped pattern for the entire gradient and the predictions from models 3 and 4 (Fig. 2). However, when the extent of scale was reduced to cover a smaller segment of the gradient, models 1 and 2 provided a better correlation, especially when the lower limit of the gradient was omitted (Fig. 2B, a). Thus, statistical evidence supporting the hump-shaped models 3 and 4 increases when a larger proportion of the gradient is included. Goodness-of-fit values within each of the four diversity models also varied depending on whether gradient segments from the lower or upper limits were omitted (Fig. 2). In contrast, all correlation patterns were consistent across different grain sizes.

Scale effects have previously been quantified for the productivity–diversity gradient[19]. It has long been recognized that truncation of a gradient may affect species richness patterns[20], whereas tabulation of the shape of the pattern of altitudinal species richness has suggested that these may be sensitive to scale effects[9]. Until now, with the use of altitudinal gradient data to test hypotheses related to species diversity, the quantitative and qualitative impacts of scale effects and their consequences have never been explicitly assessed, and as a consequence of this the effects of scale have generally been underestimated. Previous studies acknowledging potential scale issues have attempted to circumvent these effects by, for example, considering only studies that have sampled in excess of 70% of the gradient[21]. However, as we show here, even the smallest truncation of the gradient can completely shift the statistical support for competing hypotheses. This degree of sensitivity to scale effects may well be universal[22], as we obtain the same results when repeating our analyses with a data set from Costa Rica[23], which is one of the very few complete single-transect, tropical elevational gradients remaining in the world (see Supplementary Fig. 3).

It is difficult to compare altitudinal studies or to use explicit meta-analysis statistics because studies are conducted on various organisms and in all parts of the world, with each evaluation requiring the use of case-specific study designs. In addition, almost all gradients have a unique history of human intervention in the environment. The variables characterizing the organisms, their environment and their perception of scale are intercorrelated[9], and the absence of suitable factorial techniques[24] makes the meta-analysis of potential scale effects difficult to interpret. Following the approach described in this paper—that is, resampling the same empirical data at various scales of analysis[18] and subsequently exploring the statistical relationships between empirical and predicted patterns conditional on competing diversity hypotheses—can circumvent some of these problems and seems to be a powerful technique.

On the basis of a few studies, the altitudinal species richness pattern was previously considered to be universal, with monotonic declines in richness with increasing altitude (and, it was believed, with decreasing temperature and resources)[1,4,5]. Today, with more than 1,000 studies[9], the altitudinal pattern is seen to be more complex. However, monotonic declines and hump-shaped patterns with peak richness at a wide range of altitudes are the most commonly reported patterns[5–9,11,13,16,17]. The perception of varying altitudinal patterns and the current lack of consensus on the mechanisms controlling altitudinal variation may be due largely to scale effects. Differences in sampling regimens, study quality and the sheer magnitude and diversity of studies may also contribute to the wide variability in patterns.

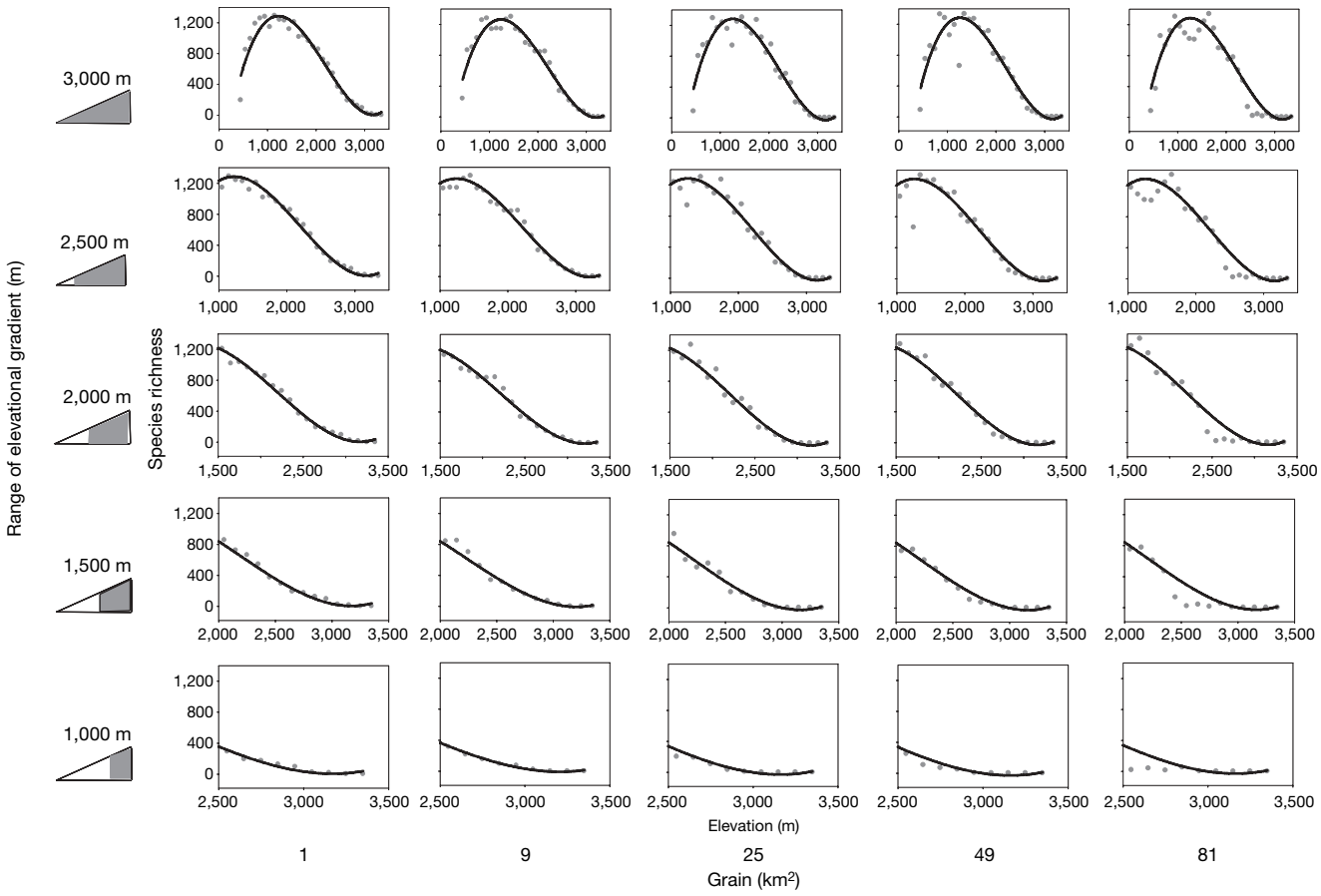


Figure 1 | Scale effects on altitudinal species richness patterns. These bivariate plots, generated by repeated sampling of the same data set, show the empirical species richness patterns based on 25 combinations of scale of extent (*y* axis) and grain size (*x* axis). The surveyed gradient was reduced by omitting segments from the lower limit. The sampled gradients are illustrated by the grey shaded areas in the triangles adjacent to the *y* axes. (Supplementary Fig. 2 shows the effects of scale as a result of omitting segments from the upper limit of the gradient.)

©2008 Nature Publishing Group

The results presented here do not provide direct evidence that a hump-shaped altitudinal species richness pattern describes the universal distribution better than a monotonically decreasing pattern. However, the results indicate that the extent of scale and omission of a part of the gradient tend to favour the monotonic pattern (see also Fig. 2 in ref. 9). In particular, the omission of the lowest part of the gradient produces a monotonic pattern (Fig. 2). A uniform pattern of human impact on altitudinal gradients worldwide (see below) may cause this scale effect to become a unidirectional bias.

In mountainous areas, lower regions are affected by settlements and exploitation of forest resources, and zones above the tree line are subject to grazing and anthropogenic fire practices intended to maintain grassland and to lower the tree line. Accordingly, deforestation is generally most extensive in the lowlands and at high altitudes, with most forest remaining at mid-altitude (Fig. 3a), while overall human impact is larger in the lowlands and decreases almost monotonically with increased elevation (Fig. 3b). That is, human activities have generally affected worldwide the lower and upper slopes more than the mid-altitudinal habitats (Fig. 3 and Methods). Today, it is increasingly rare to localize and work on complete, natural and untouched altitudinal gradients ranging from sea level to high-altitude mountaintops. Most of the existing 461 studies (Methods) have been conducted on gradients that include disturbed lowlands[5]. All regional studies include disturbed areas, and out of 203 single-transect altitudinal studies only 12 have been conducted on complete and natural gradients (Supplementary Tables 1 and 2). Paradoxically, the alternative solution of excluding lowland zones from analysis if the natural habitat has been destroyed—that is, the inclusion of lowland habitats even if disturbed—can also cause a bias towards a monotonic pattern. Disturbed habitats often have an elevated level of species richness as a result of the invasion of habitat generalists, which more than compensates for the potential loss of habitat specialists[25].

To manage biodiversity, today and in the future, it is crucial to understand the processes behind the observed natural patterns of biodiversity[26]. Unfortunately, because humans have destroyed many of the natural patterns it may be difficult to discover the mechanisms determining these patterns and to generate the knowledge required to manage biodiversity and natural systems efficiently and wisely. It is possible that human impact may already have permanently affected our ability to detect the processes that engender patterns of diversity.

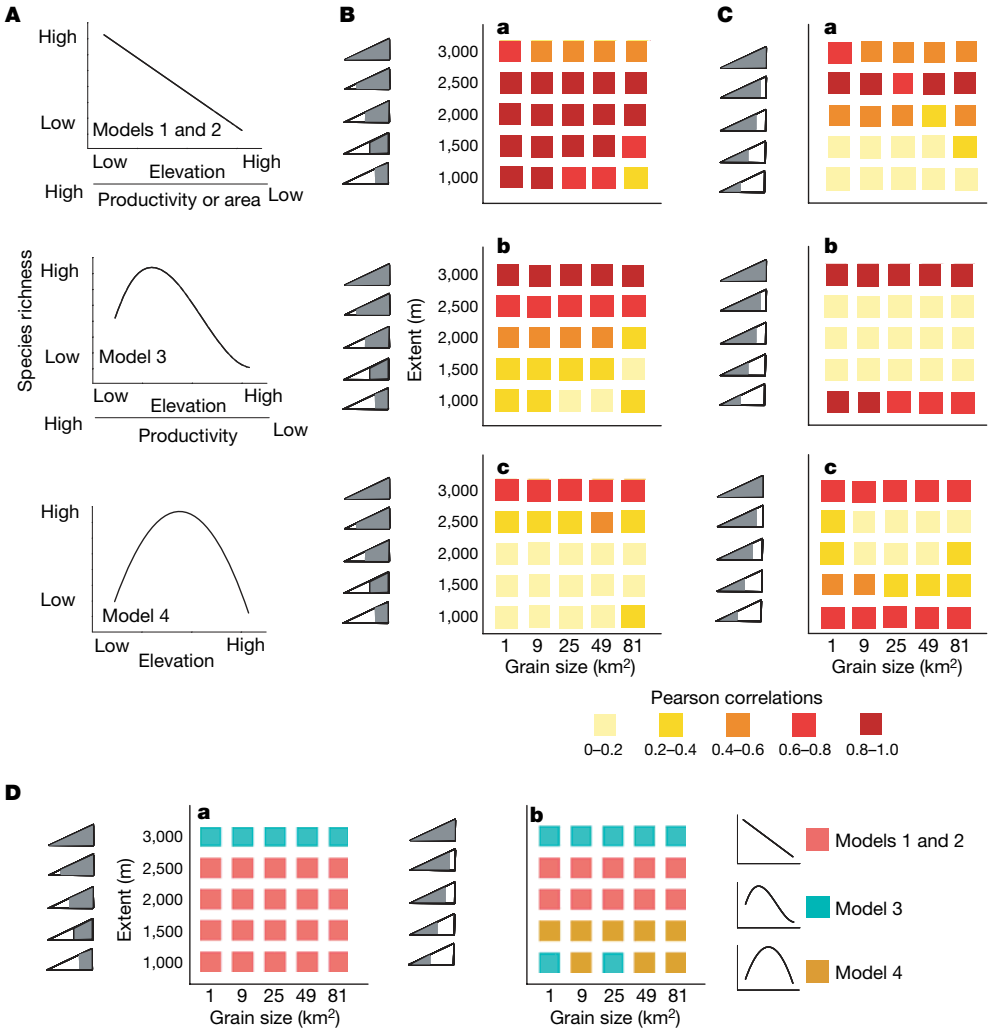


Figure 2 | Scale effects on patterns of altitudinal species richness and testing of four generalized diversity models. **A**, Schematic illustration of expected species richness patterns for four diversity models (see the text, Methods and Supplementary Fig. 1 for additional details on models). **B**, Degree of correlation (Pearson) between expected and empirical species richness values in 100-m altitudinal zones when sampling the same data with different combinations of grain size (1, 9, 26, 48 and 81 km^2) and scale of extent (omitting segments of 0, 500, 1,000, 1,500 and 2,000 m from the lower limit of the gradient as illustrated by the grey-shaded area of the small triangles next to each of the *y* axes). **a**, Models 1 and 2; **b**, model 3; **c**, model 4. **C**, As in **B**, but omitting segments from the upper limit of the gradient. **D**, Coloured squares indicate the model with the highest Pearson correlation (that is, the best fit) for 25 combinations of grain size and spatial extent when omitting segments from the lower (**B**) and upper (**C**) limits of the gradient, respectively, from the analysis.

©2008 Nature Publishing Group

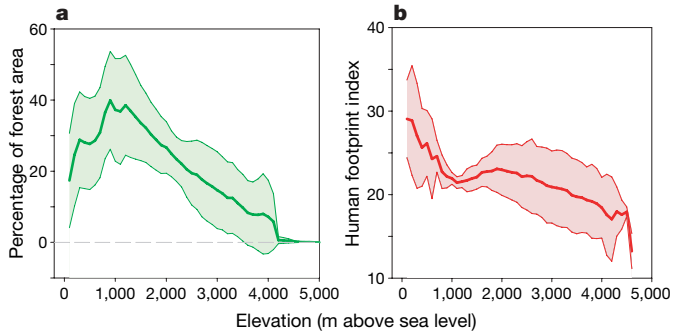


Figure 3 | Worldwide and generalized pattern of human impact along the altitudinal gradient. a, The bold green line indicates the percentage of area covered by forest (natural vegetation). **b**, The bold red line indicates the averaged human footprint index for each elevational band of 100 m above sea level. The lines are averaged on the basis of data from 13 of the largest mountain regions in the world. Shaded coloured areas indicate ±0.5 s.d. boundaries of the averaged values (for calculations see Methods, and see Supplementary Figs 4–6 for location and data for the 13 individual mountain regions).

METHODS SUMMARY

Scale effects were illustrated by using an empirical data set based on 400,000 site–species records of vascular plants, lichens and bryophytes[27] from the central Spanish Pyrenees (13,500 km^2) covering a complete regional altitudinal gradient, from the bottom of the valley at 400 m above sea level to 3,100 m above sea level. For the purpose of this paper it was assumed that the empirical data were without sampling errors or biases, and the derived altitudinal patterns of species richness were accepted at face value. Thus, no conclusions with regard to factors determining the Pyrenean altitudinal pattern of species richness should be derived from these analyses.

Species richness was calculated for each 100-m altitudinal band by using the Idrisi GIS software[28], varying grain size and scale of extent (Fig. 1 and Supplementary Fig. 2). Evaluations of scale effects on patterns of species richness and on the correlative fit between empirical and predicted data were done by resampling the distributions of 3,046 species. This was performed with five grain sizes (1, 9, 25, 49 and 81-km^2 cells) in combination with five scales of extent, for a total of 25 sampling combinations. Reduction in scale of extent was achieved through the omission of segments of 0, 500, 1,000, 1,500 and 2,000 m from the lower and upper limits of the original gradient.

The expected altitudinal pattern of species richness was calculated for four main diversity models of altitudinal variation in species richness (Methods and Supplementary Fig. 1). Predicted values of species richness were correlated with the empirical data for each altitudinal band by using the Pearson product moment correlation (see Methods for details). This was done for the 25 combinations of scale of extent and grain size for each of the four models (Figs 1 and 2, and Supplementary Fig. 2).

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

Received 26 December 2007; accepted 4 February 2008.

1. Rohde, K. Latitudinal gradients in species diversity: the search for the primary cause. *Oikos* **65,** 514–527 (1992).
2. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
3. Willig, M. R., Kaufman, D. M. & Stevens, R. D. Latitudinal gradients of biodiversity: Pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Evol. Syst.* **34,** 273–309 (2003).
4. MacArthur, R. *Geographical Ecology* (Princeton Univ. Press, Princeton, NJ, 1972).
5. Rahbek, C. The elevational gradient of species richness—a uniform pattern. *Ecography* **18,** 200–205 (1995).
6. Heaney, L. R. Small mammal diversity along elevational gradients in the Philippines: an assessment of patterns and hypotheses. *Global Ecol. Biogeogr.* **10,** 15–39 (2001).
7. Lomolino, M. Elevation gradients of species-density: historical and prospective views. *Global Ecol. Biogeogr.* **10,** 3–13 (2001).
8. Sanders, N. J. Elevational gradients in ant species richness: area, geometry, and Rapoport's rule. *Ecography* **25,** 25–32 (2002).
9. Rahbek, C. The role of spatial scale and the perception of large-scale species-richness patterns. *Ecol. Lett.* **8,** 224–239 (2005).
10. McNeill, J. R. *The Mountains of the Mediterranean World* (Cambridge Univ. Press, Cambridge, 1992).
11. Sanders, N. J., Moss, J. & Wagner, D. Patterns of ant species richness along elevational gradients in an arid ecosystem. *Global Ecol. Biogeogr.* **12,** 93–102 (2003).
12. Bhattarai, K. R., Vetaas, O. R. & Grytnes, J. A. Fern species richness along a central Himalayan elevational gradient, Nepal. *J. Biogeogr.* **31,** 389–400 (2004).
13. Rahbek, C. The relationship among area, elevation, and regional species richness in neotropical birds. *Am. Nat.* **149,** 875–902 (1997).
14. Romdal, T. S. & Grytnes, J. A. An indirect area effect on elevational species richness patterns. *Ecography* **30,** 440–448 (2007).
15. Colwell, R. K., Rahbek, C. & Gotelli, N. J. The mid-domain effect and species richness patterns: what have we learned so far? *Am. Nat.* **167,** E1–E23 (2004).
16. Dunn, R. R., McCain, C. M. & Sanders, N. J. When does diversity fit null model predictions? Scale and range size mediate the mid-domain effect. *Global Ecol. Biogeogr.* **16,** 305–312 (2007).
17. McCain, C. M. Could temperature and water availability drive elevational species richness patterns? A global case study for bats. *Global Ecol. Biogeogr.* **16,** 1–13 (2007).
18. Rahbek, C. & Graves, G. R. Multiscale assessment of patterns of avian species richness. *Proc. Natl Acad. Sci. USA* **98,** 4534–4539 (2001).
19. Chase, J. M. & Leibold, M. A. Spatial scale dictates the productivity–biodiversity relationship. *Nature* **416,** 427–430 (2002).
20. McCoy, E. D. The veiled gradients problem in ecology. *Oikos* **99,** 189–192 (2002).
21. McCain, C. M. Area and mammalian elevational diversity. *Ecology* **88,** 76–86 (2007).
22. Whittaker, R. J., Willis, K. J. & Field, R. Scale and species richness: towards a general, hierarchical theory of species diversity. *J. Biogeogr.* **28,** 453–470 (2001).
23. Brehm, G., Colwell, R. K. & Kluge, J. The role of environment and mid-domain effect on moth species richness along a tropical elevational gradient. *Global Ecol. Biogeogr.* **16,** 205–219 (2007).
24. Hillebrand, H. On the generality of the latitudinal diversity gradient. *Am. Nat.* **163,** 192–211 (2004).
25. Connell, J. H. Diversity in tropical rain forests and coral reefs—high diversity of trees and corals is maintained only in a non-equilibrium state. *Science* **199,** 1302–1310 (1978).
26. Mace, G. M. *et al.* It's time to work together and stop duplicating conservation efforts.... *Nature* **405,** 393 (2000).
27. Pyrenean Institute of Ecology. *Atlas de Flora de Aragón* ⟨http://www.ipe.csic.es/floragon/index3.php⟩ (2005).
28. Clark Labs. *Idrisi Kilimanjaro version 14.02. GIS software package* (Clark Labs, Clark University, Worcester, MA, 2004).

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

Acknowledgements We thank J. Hortal, F. W. Larsen, and D. Alagador for comments on the manuscript. The research of D.N.B. (I3P post-doc) was partly supported by the EC FP6 ALARM project and the DGA PM018/2006 project (Diputación General de Aragón). M.B.A., T.S.R. and C.R. acknowledge the Danish National Science Foundation for support of macroecological research; research by M.B.A. is also supported by the EC FP6 ECOCHANGE project. We thank the Pyrenean Institute of Ecology (CSIC) for providing the biological data set for the Spanish Pyrenees.

Author Contributions All authors designed the research. D.N.B. conducted all the analyses, except the quantitative review of the literature, which was conducted by T.R and C.R. The manuscript was written by D.N.B. and C.R. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.N.B. (davidnogues@mncn.csic.es).

©2008 Nature Publishing Group

METHODS

Calculation of predictive values for the four models. The two linear diversity models (models 1 and 2; see Supplementary Fig. 1) were calculated with a linearly decreasing function constrained by the maximum and minimum values from the empirical data set. The Digital Elevation Model (DEM) within the GIS software was used to simulate a monotonic decrease in richness with altitude for each grain size and extent combination (Supplementary Fig. 7). The same procedure was used for models 3 and 4, in which the hump-shaped function was defined within the FUZZY module in the GIS software[28].

Evaluating the sensitivity of patterns to scale effects in another elevational gradient. Here we assessed the effect of scale of extent in the Barva Transect (10° N, 84° W), a complete single-transect, tropical forested elevational gradient ranging from 40 to 2,730 m above sea level, located in the Braulio Carrillo National Park, Costa Rica, as well as adjacent areas. This transect is a unique gradient, being one of the very few complete elevational transects still existing; it has undisturbed habitats along the entire gradient while being probably the most thoroughly surveyed elevational gradients in the tropics (see http://viceroy.eeb.uconn.edu/alas/alas.html). The data analysed here were extracted from ref. 23. Because the data are from a single-transect gradient, we only evaluated scale effects associated with changes in the scale of extent (that is, the proportion of the complete altitudinal gradient sampled). The analyses of the correlative fit between empirical altitudinal patterns of species richness and predicted patterns of species richness were conducted for the grain size originally used in ref. 23; that is, elevational bands of 500 m (see Supplementary Fig. 3). The prediction of expected patterns followed the same method used for the Pyrenean data set (see Methods Summary); that is, for each elevational band an expected value was predicted by using the FUZZY module of the GIS software, following the functions that illustrate models 1, 2 and 3 (see Supplementary Fig. 1 for additional details on models).

Measuring the severity of human impact along elevational gradients in global mountain regions. The anthropogenic impacts along elevational gradients of 13 mountain ranges were evaluated (Supplementary Figs 4–6). Six of these ranges comprise tropical mountains (tropical Andes, Sierra Madre, Ethiopian highlands, Eastern Africa highlands, Mitumba mountains and Pegunungan Maoke, while seven are non-tropical (Rocky Mountains, non-tropical Andes, Pyrenees, Alps, Atlas, Caucasus and Himalayas). The Mountains of the World Geographical Information System (GIS) database was used to delimit the boundaries of the mountain ranges (http://www.mtnforum.org/mem/searchind.cfm?searchtype = atlas). For each mountain range we calculated the percentage area currently covered by forest for each 100-m elevational band (an estimator of anthropogenic disturbance suggested by the authors of ref. 29 in their 'human footprint' map), by using the US Geological Survey (USGS) Global Land Cover Database (Version 2.0) as well as the USGS GTOPO30 Global Digital Elevation Model (http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html), both of which have a horizontal grid spacing of 30 arcsec (about 1 km). Subsequently, the 'human footprint index'[29], a composite of human population and infrastructure data, was used as an estimator of human impact along the elevational gradients of the 13 mountain ranges analysed (1 km of horizontal grid spacing). We used the integrated GIS and RS (image processing) software solution, Idrisi Kilimanjaro[28] (Clark Labs) to measure changes in both estimators for each elevational band.

A quantitative review of the literature assembling the reported patterns of altitudinal gradients of species richness. The search for data sets follows the protocol of ref. 9 and is based on an ISI search performed on 12 October 2007 with the following search string: ('elevatio*' or 'altitud*') and ('richness' or 'diversit*') and ('gradien*' or 'patter*' or 'transec*' or variat*'). The search was conducted with the option 'all document types' for the period 1990–2007 and included title, abstract and keywords. A closer examination of the more than 1,000 data sets found provided 461 data sets that contained information on the variation of species richness with altitude. Of these only 78 data sets were gradients with data points from ≤500 to ≥2,000 m above sea level (Supplementary Table 1; see Supplementary Table 2 for details on the individual studies). Of the 78 data sets, 65 gradients were completely surveyed from the valley floor to the mountaintop, and most of these were based on regional compilations. All the regional studies include mountain areas along the altitudinal gradient that are in part affected by human activities (see Supplementary Figs 4 and 5). As judged from the description in the individual papers, only 12 of the 24 complete single-transect gradient data sets (of the 461 total number of altitudinal data sets) may be based on gradients with full natural habitat along the entire gradient.

29. Sanderson, E. W. *et al.* The human footprint and the last of the wild. *Bioscience* **52**, 891–904 (2002).

©2008 Nature Publishing Group

LETTERS

Neutral metacommunity models predict fish diversity patterns in Mississippi–Missouri basin

Rachata Muneepeerakul[1], Enrico Bertuzzo[1,2], Heather J. Lynch[3], William F. Fagan[3], Andrea Rinaldo[2,4] & Ignacio Rodriguez-Iturbe[1]

River networks, seen as ecological corridors featuring connected and hierarchical dendritic landscapes for animals and plants, present unique challenges and opportunities for testing biogeographical theories and macroecological laws[1]. Although local and basin-scale differences in riverine fish diversity have been analysed as functions of energy availability and habitat heterogeneity[2], scale-dependent environmental conditions[3] and river discharge[4,5], a model that predicts a comprehensive set of system-wide diversity patterns has been hard to find. Here we show that fish diversity patterns throughout the Mississippi–Missouri River System are well described by a neutral metacommunity model coupled with an appropriate habitat capacity distribution and dispersal kernel. River network structure acts as an effective template for characterizing spatial attributes of fish biodiversity. We show that estimates of average dispersal behaviour and habitat capacities, objectively calculated from average runoff production, yield reliable predictions of large-scale spatial biodiversity patterns in riverine systems. The success of the neutral theory in two-dimensional forest ecosystems[6–8] and here in dendritic riverine ecosystems suggests the possible application of neutral metacommunity models in a diverse suite of ecosystems. This framework offers direct linkage from large-scale forcing, such as global climate change, to biodiversity patterns.

The Mississippi–Missouri River System (MMRS) is an invaluable resource of great biotic diversity, including freshwater fish. Its vast extent spans diverse habitat types operating under varying environmental conditions (such as climate, hydrological regime, primary productivity and human disturbance); these diverse habitats are connected to each other by one river network. An analysis that adequately captures major spatial biodiversity patterns in such a system is therefore noteworthy.

In recent years, the neutral theory of biodiversity[6], with its minimal set of assumptions and parameters, has proven both influential[7–11] and controversial[12–14] as an explanation of biodiversity patterns. However, the theory has been tested mainly with ecosystems in two-dimensional landscapes or a mean-field context, to which spatial aspects contribute only weakly[6–8,14–17]. Only recently have the contributions of landscape spatial structure[18], for example, to biodiversity patterns in river networks[1], been investigated. Furthermore, implications of hydrological controls placed by river networks as ecological corridors have recently been explored[19,20]. Here we analyse a large database of fish diversity in the MMRS to compare empirical biodiversity patterns against those predicted by a neutral metacommunity model (see Methods). The data analysis provides significant insights in its own right, and the comparison with model results allows us to investigate the extent to which a neutral model captures observed patterns and extends inferences from the database.

In the following analysis, the 824 direct tributary areas comprising the MMRS are populated with occurrence data of 433 freshwater fish species from a database compiled by NatureServe[21] (see Methods). Here, a direct tributary area (DTA) is a geographical region directly draining to a group of streams (that is, not including areas upstream of it); the DTAs correspond to the United States Geological Survey (USGS) HUC8-scale sub-basins as defined in US National Hydrography Database Plus[22] (NHDPlus; see also Methods). Occurrence data and river network structure can be combined and analysed for several biodiversity patterns. We consider three patterns: first, the distribution of local species richness (LSR), or α diversity; second, species occupancies; and third, between-community (β) diversity. LSR is the number of species found in a randomly selected DTA. The occupancy of a species, in this case, is simply the number of DTAs in which that species is reported as present. To characterize β diversity, we consider the overall spatial decay of Jaccard's similarity index[7] (JSI). JSI of any pair of DTAs is defined as $S_{ij}/(S_i + S_j - S_{ij})$, where S_{ij} is the number of

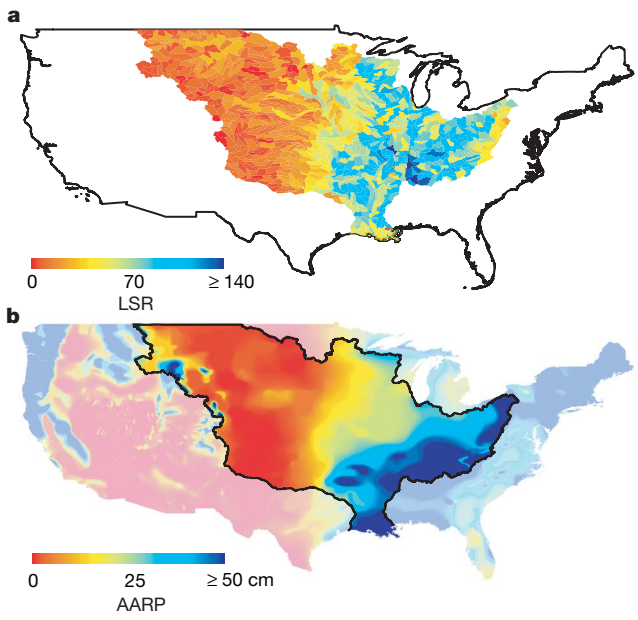


Figure 1 | Maps of freshwater fish diversity and AARP in the MMRS. **a**, Local species richness (LSR), or α diversity, of the freshwater fish in each DTA (that is, at the USGS HUC8 scale; see the text) of the MMRS. **b**, The AARP of the MMRS.

[1]Department of Civil and Environmental Engineering, E-Quad, Princeton University, Princeton, New Jersey 08544, USA. [2]Laboratory of Ecohydrology, Faculté ENAC, École Polytechnique Fédérale, CH-1015 Lausanne, Switzerland. [3]Department of Biology, University of Maryland, College Park, Maryland 20742, USA. [4]Dipartimento di Ingegneria Idraulica, Marittima, Ambientale e Geotecnica (IMAGE) and Centro Internazionale di Idrologia 'Dino Tonini', Università di Padova, via Loredan 20, I-35131, Padua, Italy.

©2008 Nature Publishing Group

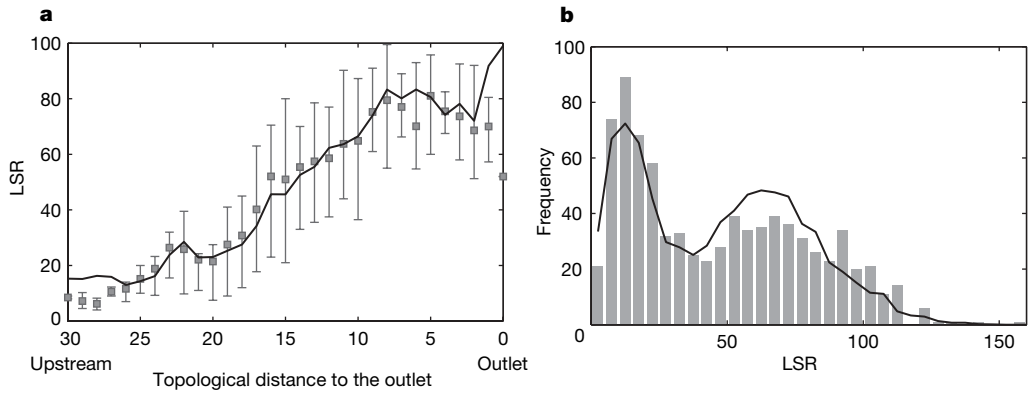


Figure 2 | Patterns of local species richness. **a**, LSR profile as a function of the topological distance (see Methods) from the outlet located in Atchafalaya, Louisiana (NHDPlus sub-basin number 08080101). **b**, Frequency distribution of LSR. The squares (average values) with error bars (ranging from the 25th to the 75th quantile) and bar plots represent the empirical data, and the lines represent the average values of the model results.

species present in both DTAs i and j, and S_i is the total number of species in DTA i. To achieve reliable statistics, we consider only topological distances (see Methods) for which more than 500 DTA pairs exist.

The map of LSR is shown in Fig. 1a. The DTA with the maximum LSR is Pickwick Lake (156 species) at the borders of the states of Alabama, Mississippi and Tennessee (NHDPlus sub-basin number 06030005). The map can be divided roughly into the western, species-poor half and the eastern, species-rich half. The sharp decrease in the species richness occurs around the 100° W meridian, which is also known to be the location of sharp gradients of annual precipitation[23] and runoff production[24,25] (Fig. 1b). Although these gradients partly explain the arid climate and low fish diversity in the western half[26,27], we argue that the western DTAs are low in fish diversity both because their climate is dry and because they are upstream portions of the river network (see Supplementary Information). If they were located downstream, they might receive enough water supply and have access to a larger species pool from their wetter upstream sub-networks to maintain high fish-habitat capacities and fish diversity.

Figure 2a shows the LSR as a function of the topological distance from the network outlet. The distance zero corresponds to Atchafalaya, Louisiana (NHDPlus sub-basin number 08080101). The LSR profile shows a significant increase in the downstream direction, except at the very end in Louisiana, where we suggest that the freshwater fish-habitat capacities are significantly reduced by salinity, co-occurrence/intrusion by some freshwater-tolerant estuarine or coastal fish species, human disturbance and pollution. The overall downstream increase in richness results from the converging character of the river network[28] and is steepened by the dry–wet climatic gradient mentioned above (see Supplementary Information). Figure 2b presents the frequency distribution of LSR, whose two peaks at low and high values reflect the difference between the western and eastern halves of the MMRS.

The species occupancies are presented in Fig. 3 as a rank–occupancy curve, in which the fish species are ranked by their occupancies. The rank–occupancy curve (akin to the familiar rank–abundance curves[6]) yields a straight line on a semilogarithmic scale, a pattern reminiscent of the rank–abundance curves predicted by the neutral theory[6,9,16]. Figure 4 shows that the JSI decreases as the topological distance between DTA pairs under consideration increases, an expected trend for β diversity. However, the JSI does not vanish even for DTA pairs that are very far apart. Such long-distance similarity in species composition is probably maintained by species with extremely large occupancies, for example *Ictalurus punctatus* (channel catfish), *Ameiurus melas* (black bullhead) and *Ameiurus natalis* (yellow bullhead).

As alluded to above, the neutral metacommunity model is a promising candidate for modelling the general spatial biodiversity patterns of the MMRS's freshwater fish. Here we show that by implementing the neutral model in the MMRS and incorporating the effect of average annual runoff production (AARP) on fish-habitat capacities, we can effectively reproduce a wide spectrum of observed biodiversity patterns. For instance, in addition to the general trend and magnitude, the model also captures fine-structured fluctuations of the LSR profile (Fig. 2a). The fits to the LSR frequency distribution and β diversity pattern are also very good (Figs 2b and 4). The straight-line character of the rank–occupancy curves is evident for both the data and the model result (Fig. 3). Simultaneous fits of these diverse patterns (and others, such as species–area relationship) are a very stringent test for a model[29], especially a model with only four parameters as in this case (see Supplementary Information). The model also permits additional inferences to be drawn. The parameters corresponding to the best fits imply that the spread of the average fish species is quite symmetrical; that is, significantly biased in neither the upstream nor the downstream direction ($w_u = 1$; see Methods). The model results also suggest that, on average, most fish disperse locally (that is, to nearby DTAs) but a non-negligible fraction travel very long distances (see Supplementary Information).

Given the diverse environmental conditions covered by the MMRS, our demonstration that a simple neutral metacommunity model coupled with an appropriate habitat capacity distribution and dispersal kernel can simultaneously reproduce several major observed biodiversity patterns has far-reaching implications. These results suggest that only parameters characterizing average fish behaviour—as opposed to those characterizing biological properties of all different fish species in the system—and habitat capacities and connected structure suffice for reasonably reliable predictions of large-scale biodiversity patterns to be obtained. The neutral metacommunity model also provides a null model against which more biologically realistic models may be compared, and further developments in our understanding

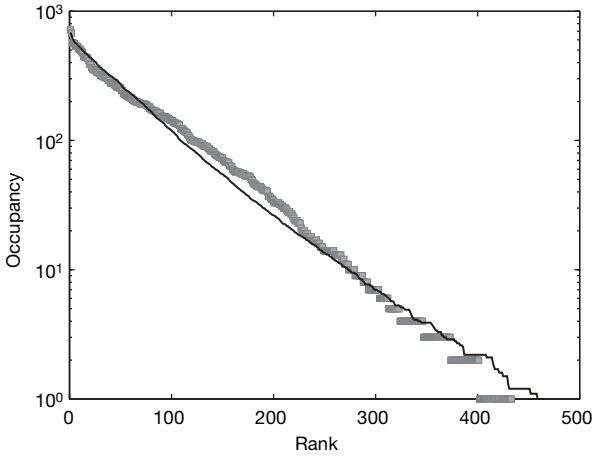


Figure 3 | Rank–occupancy curve. The squares (which are densely placed and appear as grey stripes) represent the data and the line represents the model result. Here, the occupancy of a freshwater fish species is simply the number of DTAs in which that species is reported as present. Note that the straight-line character in the semilogarithmic scale is shared by the familiar rank–abundance curves predicted by the neutral theory of biodiversity[6,9,16].

©2008 Nature Publishing Group

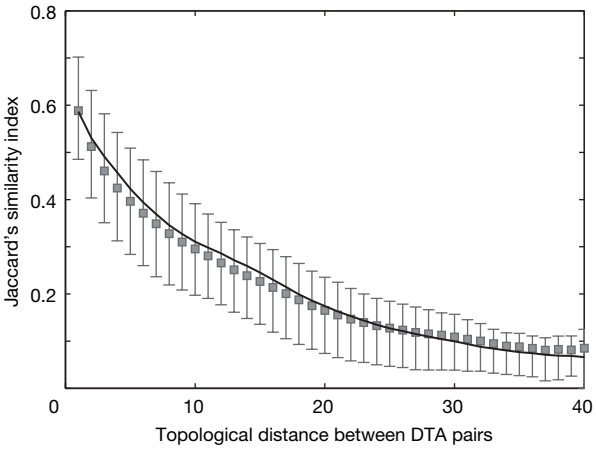


Figure 4 | The Jaccard's similarity index (JSI) as a function of topological distance between DTA pairs. The overall decay of JSI characterizes β diversity[7]; the squares with error bars represent the average values with the range between the 25th and 75th quantiles of the empirical data, and the line represents the average values of the model results. Note that the JSI does not vanish, even for widely separated DTA pairs.

of riverine networks and fish movement will permit a continued improvement between model and data. Indeed, although this modelling approach has been shown here to be useful for investigating key spatial patterns, it is crucial to recognize that "neutral pattern does not imply neutral process"[9]. Different approaches will therefore be necessary for predicting transient dynamics of the system or for understanding patterns and dynamics of specific species.

Finally, because mobile fish in a river network differ drastically from sessile trees in a forest, it is remarkable that the neutral theory can reproduce key biodiversity patterns of both sets of organisms quite well. This suggests that patterns predicted by the neutral metacommunity model—with appropriate habitat capacity distribution and dispersal kernel—may be broadly applicable across diverse ecosystems. It also offers a general, parsimonious modelling approach that acts as a coherent framework for studying several large-scale spatial biodiversity patterns simultaneously. This framework permits direct linkages to be made from various environmental changes to biodiversity patterns. For example, changes in precipitation patterns, perhaps as a result of global climate change, can now be mapped to changes in habitat capacities in the model; changes in connectivity among local communities, for example flow rerouting or damming in the case of fish, can be characterized by modifying the dispersal kernel. These linkages in turn enable us to make reliable predictions of a comprehensive set of altered biodiversity patterns, with significant implications for conservation campaigns and large-scale resource management.

METHODS SUMMARY

The biogeographical data on fish used in the analysis were obtained from the NatureServe[21] database of US freshwater fish distributions, which summarizes museum records, published literature and expert opinion about fish species distribution in the United States, except Alaska, and is tabulated at the USGS HUC8 scale[22]. Owing to the present lack of availability of data, the Canadian portions of the MMRS are not included in the analysis, but we do not expect this to affect the key results and conclusions reported here. The data were then analysed to produce spatial biodiversity patterns (see Methods).

Our model is of a structured metacommunity type. The neutral theory of biodiversity is implemented in the MMRS, using its network as the structure of the metacommunity. Each DTA is a local community in that metacommunity and has a different fish-habitat capacity, H, defined as the number of 'fish units' sustainable by resources in that particular DTA; a fish unit can be thought of as a subpopulation of fish of the same species. H is assumed to be proportional to the product of the DTA watershed area and AARP[25], an indicator of the quantity of resources available for fish[2,4]. The model uses the topological, rather than euclidean, distances between DTAs because they are representative of how far fish travel. The model captures basic ecological processes: birth, death, dispersal, colonization and diversification. The simulations are run until the system reaches a steady state; the biodiversity patterns of interest are then determined and compared with the empirical patterns.

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

Received 14 November 2007; accepted 1 February 2008.

1. Muneepeerakul, R., Weitz, J. S., Levin, S. A., Rinaldo, A. & Rodriguez-Iturbe, I. A neutral metapopulation model of biodiversity in river networks. *J. Theor. Biol.* **245,** 351–363 (2007).
2. Guégan, J.-F., Lek, S. & Oberdorff, T. Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature* **39,** 382–384 (1998).
3. Angermeier, P. L. & Winston, M. R. Local vs. regional influences on local diversity in stream fish communities of Virginia. *Ecology* **79,** 911–927 (1998).
4. Oberdorff, T., Guégan, J.-F. & Hugueny, B. Global scale patterns of fish species richness in rivers. *Ecography* **18,** 345–352 (1995).
5. Xenopoulos, M. A. & Lodge, D. M. Going with the flow: using species-discharge relationships to forecast losses in fish diversity. *Ecology* **87,** 1907–1914 (2006).
6. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, Princeton, NJ, 2001).
7. Condit, R. *et al.* Beta-diversity in tropical forest trees. *Science* **295,** 666–669 (2002).
8. Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Neutral theory and relative species abundance in ecology. *Nature* **424,** 1035–1037 (2003).
9. Purves, D. W. & Pacala, S. W. in *Biotic Interactions in the Tropics* (eds Burslem, D., Pinard, M. & Hartley, S.) 107–138 (Cambridge Univ. Press, Cambridge, 2005).
10. Holt, R. D. Emergent neutrality. *Trends Ecol. Evol.* **21,** 531–533 (2006).
11. Alonso, D., Etienne, R. S. & McKane, A. J. The merits of neutral theory. *Trends Ecol. Evol.* **21,** 451–457 (2006).
12. McGill, B. J. A test of the unified neutral theory of biodiversity. *Nature* **422,** 881–885 (2003).
13. Chave, J. Neutral theory and community ecology. *Ecol. Lett.* **7,** 241–253 (2004).
14. Dornelas, M., Connolly, S. R. & Hughes, T. P. Coral reef diversity refutes the neutral theory of biodiversity. *Nature* **440,** 80–82 (2006).
15. Etienne, R. S. & Olff, H. Confronting different models of community structure to species-abundance data: a Bayesian model comparison. *Ecol. Lett.* **8,** 493–504 (2005).
16. Walker, S. C. & Cyr, H. Testing the standard neutral model of biodiversity in lake communities. *Oikos* **116,** 143–155 (2007).
17. Volkov, I., Banavar, J. R., Hubbell, S. P. & Maritan, A. Patterns of relative species abundance in ecology. *Nature* **450,** 45–49 (2007).
18. Economo, E. P. & Keitt, T. H. Species diversity in neutral metacommunities: a network approach. *Ecol. Lett.* **10,** 52–62 (2007).
19. Campos, D., Fort, J. & Mendez, V. Transport on fractal river networks: Application to migration fronts. *Theor. Popul. Biol.* **69,** 88–93 (2006).
20. Bertuzzo, E., Maritan, A., Gatto, M., Rodriguez-Iturbe, I. & Rinaldo, A. River networks and ecological corridors: reactive transport on fractals, migration fronts, hydrochory. *Wat. Resour. Res.* **41,** W04419 (2007).
21. NatureServe. *Distribution of Native U.S. Fishes by Watershed* ⟨www.natureserve.org/getData/dataSets/watershedHucs/index.jsp⟩ (2004).
22. Seaber, P. R., Kapinos, F. P. & Knapp, G. L. U. S. *Geological Survey, Water Supply Paper 2294—Hydrologic Unit Maps* ⟨http://pubs.usgs.gov/wsp/wsp2294/html/pdf.html⟩ (1994).
23. Dingman, S. *Physical Hydrology* 2nd edn (Prentice Hall, Upper Saddle River, New Jersey, 2002).
24. Henning, D. *Atlas of the Surface Heat Balance of the Continents* (Gebrüder Borntraeger, Berlin, 1989).
25. Gebert, W. A., Graczyk, D. J. & Krug, W. *Average Annual Runoff in the United States, 1951–80* ⟨http://aa179.cr.usgs.gov/metadata/wrdmeta/runoff.htm⟩ (1987).
26. Master, L. L., Flack, S. R. & Stein, B. A. *Rivers of Life: Critical Watersheds for Protecting Freshwater Biodiversity* (Nature Conservancy, Arlington, VA, 1998).
27. Matthews, W. J. *Patterns in Freshwater Fish Ecology* (Kluwer Academic, Norwell, MA, 1998).
28. Fernandes, C. C., Podos, J. & Lundberg, J. G. Amazonian ecology: Tributaries enhance the diversity of electric fishes. *Science* **305,** 1960–1962 (2004).
29. Grimm, V. *et al.* Pattern-oriented modelling of agent-based complex systems: lessons from ecology. *Science* **310,** 987–991 (2005).

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

Acknowledgements We thank L. Master, who facilitated our access to the fish database, T. Mueller for help in data preparation, and S. Azaele for comments. We acknowledge NatureServe[21] as the author/compiler of the fish biogeography database on which IndicatorMatrix.txt (see Methods) is based and for authorizing us to include the matrix as part of the Supplementary Information. R.M., H.J.L., W.F.F. and I.R.-I. acknowledge the support of the James S. McDonnell Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.M. (rmuneepe@princeton.edu) or I.R.-I. (irodrigu@princeton.edu).

©2008 Nature Publishing Group

METHODS

Model simulations. Every DTA is assumed to always be saturated at its capacity; that is, no available resources are left unexploited. At each time step, a fish unit, randomly selected from all fish units in the system, dies and the resources that previously sustained the unit are freed and available for sustaining a new fish unit. With probability v, the diversification rate, the new unit will represent a new species (the diversification is a rate per birth and is due to speciation, to external introduction of non-native species, or to immigration (and reimmigration) of a new species from outside the MMRS); with probability $1 - v$, the new unit will belong to a species already existing in the system (the MMRS). In the latter case, the probability P_{ij} that an empty unit in DTA i will be colonized by a species from DTA j is determined as follows (including the probability $1 - v$):

$$P_{ij} = (1-v)\frac{K_{ij}H_j}{\sum_{k=1}^{N} K_{ik}H_k}$$

where K_{ij} is the dispersal kernel (see below), H_k is the habitat capacity of DTA k, and N is the total number of DTAs (here, $N = 824$). All the fish units in DTA j have the same probability of colonizing the empty unit in DTA i where the death took place. The reported model results are the average patterns after the system reaches a statistically steady state.

A dispersal kernel determines how the fish units move within the river network. Here, it is assumed to take the form of a combination of back-to-back exponential and Cauchy distributions; note that a combination of several theoretical dispersal kernels has been used to achieve a good representation of real dispersal kernels[30] (see also Supplementary Information). The dispersal kernel in this model can be expressed as

$$K_{ij} = C\left[a^{L_{ij}} + \frac{b^2}{L_{ij}^2 + b^2}\right]$$

where K_{ij} is the probability that a fish unit produced at DTA j arrives at DTA i after dispersal; C is the normalization constant; L_{ij} is the effective distance, defined as $ND_{ij} + w_u NU_{ij}$, where ND_{ij} and NU_{ij} are the numbers of downstream and upstream steps comprising the shortest path from DTA j to DTA i, and w_u is the weight factor modifying the upstream distance; $w_u > 1$ implies downstream-biased dispersal, thereby characterizing dispersal directionality; and a (less than 1) and b characterize the exponential and Cauchy decays, respectively. Here, C is determined numerically such that, for every DTA j, $\Sigma_i K_{ij} = 1$; that is, no fish can travel out of the network. At the upstream ends of the system this is obvious; it is also true at the downstream end of the system; namely the outlet to the Gulf of Mexico, a marine body that acts as a barrier to freshwater fish. Finally, the dispersal kernel of every species is assumed to be the same; this is perhaps a strong assumption because fish species obviously differ in their dispersal abilities. However, the 'functional equivalence' between species is a key way in which the neutral theory of biodiversity departs from classical ecological models. We assume the species equivalence to study just how good a fit the neutral metacommunity model can produce to our data in the absence of detailed, species-specific information.

Average annual runoff production (AARP). Runoff is the portion of precipitation that is drained by the river network. It depends on precipitation, evapotranspiration and infiltration. The map in Fig. 1b is estimated from the streamflow data of small tributaries collected from about 12,000 gauging stations averaged over the period 1951–80. For details see ref. 25.

Direct tributary area (DTA). The DTAs in the present analysis correspond to the HUC8-scale sub-basins designated by the USGS (available from www.horizon-systems.com/nhdplus/index.php). Details of how their boundaries are designated are given in ref. 22.

Habitat capacity, *H*. Habitat capacity of DTA i, H_i, is determined by

$$H_i = C_H N \frac{\mathrm{AARP}_i \times \mathrm{WA}_i}{\sum_{i=1}^{N} (\mathrm{AARP}_i \times \mathrm{WA}_i)}$$

rounded to the nearest integer. WA denotes watershed area, N (which here is 824) the total number of DTAs, and C_H the estimate (due to rounding) of average habitat capacity in a DTA.

Topological distance. The topological distance is a measure of distance along the network. An increment in the topological distance occurs when one travels along the network and crosses from one DTA to another. In the present case, one unit of topological distance corresponds to a distance in the range 100–200 km.

Notes on supplementary data. Two matrices summarizing the data used in the analysis are provided in the Supplementary Information. The first matrix, IndicatorMatrix.txt, reports the occurrence data[21] of each of the 433 fish species in each of the 824 DTAs included in the analysis. Its first column lists the identification numbers of the HUC8 sub-basins[22] (DTAs in the present analysis). The remaining 433 columns consist exclusively of zeros and ones, representing the absence and presence of each species, respectively. No species names are given; they are not necessary for the analysis. The sum along each row (433 elements) therefore gives the local species richness (LSR) of the corresponding DTA, and the sum along each column (824 elements) gives the occupancy of the corresponding species. The second matrix, TopologicalDistanceMatrix.txt, reports the topological distance between each pair of the 824 DTAs, which we derived from the data available from ref. 22. Its diagonal elements are zeros: each DTA is at zero distance from itself. The rows and columns of this matrix correspond to the DTA (that is, the HUC8 sub-basin) numbers in the first column of IndicatorMatrix.txt. The outlet corresponds to DTA number 08080101, which is the 364th row of IndicatorMatrix.txt. These two matrices can thus be combined to produce the profile of LSR as a function of topological distance from the outlet (Fig. 1a) and the pattern of Jaccard's similarity index (JSI) as a function of topological distance between DTA pairs (Fig. 4).

30. Levin, S. A., Muller-Landau, H. C., Nathan, R. & Chave, J. The ecology and evolution of seed dispersal: a theoretical perspective. *Annu. Rev. Ecol. Syst.* **34**, 575–604 (2003).

©2008 Nature Publishing Group

LETTERS

REST maintains self-renewal and pluripotency of embryonic stem cells

Sanjay K. Singh[1,5]*, Mohamedi N. Kagalwala[1,5]*†, Jan Parker-Thornburg[2], Henry Adams[1] & Sadhan Majumder[1,3,4,5,6]

The neuronal repressor REST (RE1-silencing transcription factor; also called NRSF) is expressed at high levels in mouse embryonic stem (ES) cells[1], but its role in these cells is unclear. Here we show that REST maintains self-renewal and pluripotency in mouse ES cells through suppression of the microRNA miR-21. We found that, as with known self-renewal markers, the level of REST expression is much higher in self-renewing mouse ES cells than in differentiating mouse ES (embryoid body, EB) cells. Heterozygous deletion of *Rest* (*Rest*$^{+/-}$) and its short-interfering-RNA-mediated knockdown in mouse ES cells cause a loss of self-renewal—even when these cells are grown under self-renewal conditions—and lead to the expression of markers specific for multiple lineages. Conversely, exogenously added REST maintains self-renewal in mouse EB cells. Furthermore, *Rest*$^{+/-}$ mouse ES cells cultured under self-renewal conditions express substantially reduced levels of several self-renewal regulators, including Oct4 (also called Pou5f1), Nanog, Sox2 and c-Myc, and exogenously added REST in mouse EB cells maintains the self-renewal phenotypes and expression of these self-renewal regulators. We also show that in mouse ES cells, REST is bound to the gene chromatin of a set of miRNAs that potentially target self-renewal genes. Whereas mouse ES cells and mouse EB cells containing exogenously added REST express lower levels of these miRNAs, EB cells, *Rest*$^{+/-}$ ES cells and ES cells treated with short interfering RNA targeting *Rest* express higher levels of these miRNAs. At least one of these REST-regulated miRNAs, miR-21, specifically suppresses the self-renewal of mouse ES cells, corresponding to the decreased expression of Oct4, Nanog, Sox2 and c-Myc. Thus, REST is a newly discovered element of the interconnected regulatory network that maintains the self-renewal and pluripotency of mouse ES cells.

REST is believed to be a major transcriptional repressor of neurogenesis[2–5], and activation of REST target genes was found to be sufficient to convert neural stem/progenitor cells to neuronal phenotypes[6,7]. However, REST activity seems to depend on the cellular context; for example, REST can show both an oncogenic[8–10] and tumour-suppressor function[5] as well as involvement in haematopoietic and cardiac differentiation[3–5]. Embryonic stem (ES) cells are pluripotent cells that have the potential for both indefinite self-renewal and differentiation into all three germ layers of the body[11]. Here we provide evidence that REST has a unique role as a protector of self-renewal and pluripotency in mouse ES cells, corresponding to the expression of critical regulators such as Oct4, Nanog, Sox2 and c-Myc.

We began by assessing the levels of REST protein in mouse ES cells growing under self-renewal conditions and differentiation conditions (Fig. 1a; ES and EB, respectively). As expected, western blotting showed that the ES cells had higher levels of REST expression and of the representative markers of self-renewal (proteins Oct4, Sox2 and c-Myc) than did the EB cells, indicating that REST expression is associated with self-renewal. To determine whether REST regulates the self-renewal of mouse ES cells, we took two approaches using two cell lines—YHC334 (YHC) and RRC160 (RRC) (see Supplementary Information). In the first approach, we performed polymerase chain reaction with reverse transcription (RT–PCR; Fig. 1b), quantitative RT–PCR (qRT–PCR; Fig. 1c) using different primer sets than those used for regular RT–PCR, and western blotting (Fig. 1d), all of which showed that both of the *Rest*$^{+/-}$ mouse ES cell lines (YHC and RRC) had substantially lower levels of *Rest* transcript and protein than the parental mouse ES cells. Alkaline phosphatase assays[12], performed to measure self-renewal within these cells, revealed that the levels of self-renewal in both of the *Rest*$^{+/-}$ mouse ES cell lines were less than half that in the parental mouse ES cells (Fig. 1e, f).

In the second approach, done to rule out any adaptive response accumulated by *Rest*$^{+/-}$ mouse ES cell lines, we determined whether the decreased self-renewal seen in the *Rest*$^{+/-}$ mouse ES cells could also be seen in mouse ES cells treated with short interfering RNA (siRNA) against *Rest* messenger RNA (si*Rest*). qRT–PCR assays, done to confirm that siRNA-treated cells actually had decreased *Rest* and *Oct4* levels, showed a 50% knockdown of *Rest* and *Oct4*, respectively, in both si*Rest*- and si*Oct4*-treated cells (Fig. 1g). This was further confirmed by western blotting (Fig. 1h). Self-renewal assays of these cells (Fig. 1i) then showed that whereas mouse ES cells treated with non-targeting control (NT) siRNA showed almost no change in self-renewal (94.5%), mouse ES cells treated with si*Rest* showed about a 50% decrease in self-renewal (see Supplementary Fig. 1 for the effect of four different si*Rest* constructs on the self-renewal capacity of mouse ES cells). Si*Oct4*-treated mouse ES cells also showed a decrease in self-renewal (47%), consistent with the role of Oct4 in mouse ES cell self-renewal[13]. Thus, si*Rest*-mediated knockdown of *Rest* transcripts, and not nonspecific siRNA activity, caused decreased self-renewal in mouse ES cells, a result similar to that seen in the *Rest*$^{+/-}$ mouse ES cell lines.

Gain-of-function experiments (Fig. 1j) in which self-renewing mouse ES cells were transfected with plasmids encoding a neomycin resistance gene and either green fluorescent protein (GFP) or REST[12] showed that cells transfected with GFP showed very little maintenance of self-renewal, whereas cells transfected with REST showed approximately 55% maintenance of self-renewal. Taken together, these results further confirmed the role of REST in maintaining self-renewal of mouse ES cells.

To ascertain the lineages into which *Rest*$^{+/-}$ haploinsufficient mouse ES cells differentiate on their exit from REST-mediated self-renewal, we performed RT–PCR (Fig. 2a) and qRT–PCR (Fig. 2b) assays of parental mouse ES cells, YHC and RRC cells growing under

[1]Department of Cancer Genetics, [2]Department of Biochemistry and Molecular Biology, [3]Department of Neuro-Oncology, [4]The Brain Tumor Center, [5]Center for Stem Cell and Developmental Biology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA. [6]Program in Genes and Development, The University of Texas Graduate School of Biomedical Sciences at Houston, Houston, Texas 77030, USA. †Present address: Laboratory of Genetics, Salk Institute for Biological Studies, La Jolla, California 92037, USA. *These authors contributed equally to this work.

©2008 Nature Publishing Group

self-renewal conditions in the presence of leukaemia inhibitory factor (LIF). *Rest*$^{+/-}$ cells showed activation of markers specific for mesoderm, endoderm, ectoderm and trophectoderm. Expression of all of these differentiation markers was absent or detectable at only very low levels in self-renewing parental mouse ES cells. The results in RRC cells (data not shown) and YHC cells were very similar. Notably, although the neuronal marker gene Calbindin is a direct target of REST[1], the other differentiation markers are not known to be direct REST targets, suggesting that REST can maintain the pluripotency of mouse ES cells partly by directly repressing the expression of some differentiation genes and partly by an indirect mechanism in which the non-REST-target differentiation genes are repressed as a consequence of the REST-mediated maintenance of self-renewal.

The differentiation of *Rest*$^{+/-}$ haploinsufficient cells into multiple lineages growing under self-renewal conditions prompted us to determine the differentiation characteristics of EB cells derived from wild-type as well as *Rest*$^{+/-}$ ES cells. RT–PCR of *Rest*$^{+/-}$ EB cells detected the expression of markers for all lineages, including ectoderm, mesoderm, endoderm and trophectoderm (Fig. 2c). Some of these markers were either absent or markedly reduced in wild-type ES-cell-derived EB cells. Thus, although both wild-type and *Rest*$^{+/-}$ cells started expressing multiple differentiation markers in EB cells, the *Rest*$^{+/-}$-derived cells expressed a greater number of lineage markers than did the EB cells derived from wild-type mouse ES cells, suggesting that REST deficiency predisposes mouse ES cells to differentiation. To determine whether si*Rest*-treated mouse ES cells also differentiate into various lineages, we performed a qRT–PCR assay with these cells growing under self-renewal conditions (Fig. 2d). The results indicated that si*Rest* indeed increased the expression of various lineage markers. Notably, the exact level of expression of each differentiation marker varied between si*Rest*-treated and *Rest*$^{+/-}$ mouse ES cells (Fig. 2b), presumably because of the differences in the methods used to reduce the REST levels in these cells (that is, whereas *Rest*$^{+/-}$ cells may show additional adaptive responses as a result of being cultured for longer than the si*Rest*-treated mouse ES cells, the latter show more immediate effects of REST knockdown). Nonetheless, the overall results indicated that on exit from REST-mediated self-renewal, both si*Rest*-treated wild type and *Rest*$^{+/-}$ mouse ES cells were able to differentiate into various lineages.

The inner cell mass of the blastocyst has been shown to harbour self-renewing cells that express proteins such as Nanog, Oct4 and Sox2. To determine whether REST is expressed in blastocysts, we subjected mouse blastocysts to double-immunofluorescence analysis using anti-REST antibodies and antibodies against Nanog, Oct4, or Sox2. Consistent with its role in the self-renewal of mouse ES cells, REST was co-expressed with these self-renewal regulators in the blastocyst inner cell mass (Fig. 2e) (see Supplementary Information for a discussion on the role of REST in cells of the inner cell mass).

When we subjected the parental mouse ES cells and *Rest*$^{+/-}$ (YHC and RRC) haploinsufficient cells growing under self-renewal

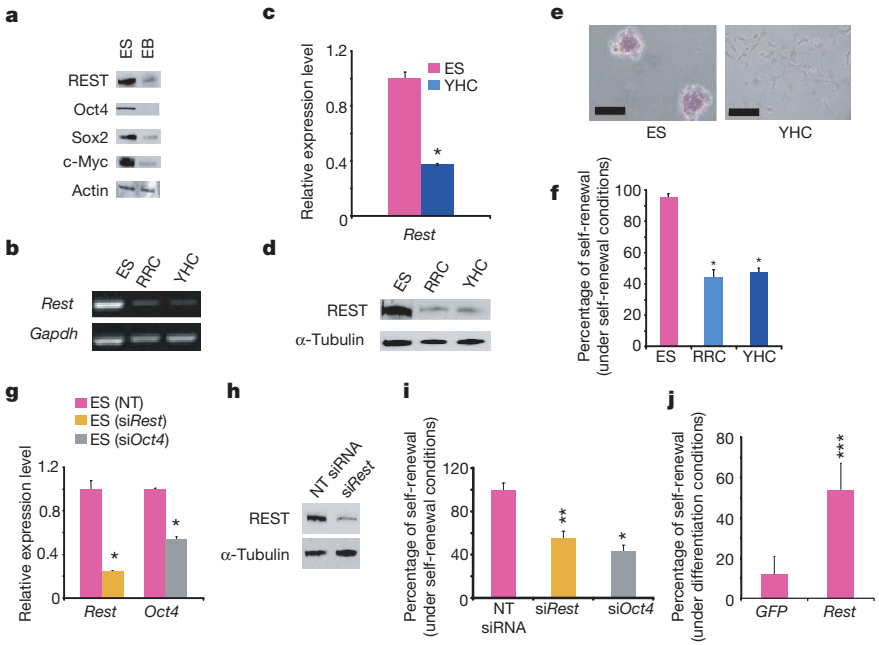


Figure 1 | REST regulates self-renewal in mouse ES cells. **a**, EB cells (ES cells grown in the absence of LIF under non-adherent conditions for 4 days) showed reduced levels of self-renewal markers and REST compared with ES cells (grown in the presence of LIF under adherent conditions). Western blot analysis of whole-cell extracts prepared from ES cells and EB cells showed an association between expression of REST and self-renewal markers. Actin was used as a loading control. **b–f**, Mouse ES cell lines with heterozygous deletion of *Rest* (RRC and YHC) show loss of self-renewal. **b**, **c**, RT–PCR (**b**) and qRT–PCR (**c**) analyses with different primer sets of total RNA isolated from ES, RRC and YHC cells show reduced *Rest* transcripts in *Rest*$^{+/-}$ cells. *Gapdh* was used as a loading control. *, $P < 0.0001$. The values are represented as mean ± s.d. ($n = 3$). **d**, Western blot analysis of whole-cell extracts from ES, RRC and YHC cells shows reduced levels of REST protein in *Rest*$^{+/-}$ cells. α-Tubulin was used as a loading control. **e**, Alkaline phosphatase staining of ES colonies shows loss of self-renewal in *Rest*$^{+/-}$ cells as compared with wild-type ES cells. **f**, Percentages of self-renewing colonies of ES, RRC and YHC cells calculated after alkaline phosphatase assays when cultured under self-renewing conditions show significant reductions in the self-renewal capacity of both *Rest*$^{+/-}$ cell lines compared with ES cells. *, $P < 0.0001$. The error bars correspond to three replicates ($n = 3$) and show mean ± s.d. **g–i**, siRNA-mediated knockdown of REST causes loss of self-renewal in mouse ES cells. **g**, Specific knockdown of targeted genes was achieved using siRNA. qRT–PCR of total RNA purified from mouse ES cells treated with si*Rest* or si*Oct4* shows knockdown of specific genes. Analysis was performed 5 days after transfection. *, $P < 0.0001$. The values are represented as mean ± s.d. ($n = 3$). **h**, Western blotting shows reduced REST protein levels in si*Rest*-treated cells compared with control (NT siRNA). **i**, si*Rest*- and si*Oct4*-treated cells show less self-renewal than NT-siRNA-treated cells. Mouse ES cell colonies were screened by alkaline phosphatase assays. *, $P < 0.0001$; **, $P < 0.001$. The error bars correspond to three replicates ($n = 3$) and show mean ± s.d. **j**, Exogenously added *Rest*, but not *GFP*, maintained self-renewal in mouse ES cells cultured under differentiation conditions. Mouse ES cells were transfected with plasmids encoding GFP or REST and grown in the absence of LIF. Percentages of self-renewing colonies from three independent experiments were averaged after alkaline phosphatase assay and are shown for each transfected gene. ***, $P < 0.01$. The error bars correspond to three replicates ($n = 3$) and show mean ± s.d.

©2008 Nature Publishing Group

conditions to RT–PCR assays to examine the transcript levels of various self-renewal markers, we found that the decreased REST levels in *Rest*$^{+/-}$ cells corresponded with decreased transcript levels of several other self-renewal genes, such as *Oct4*, *Nanog*, *Sox2*, *Tbx3* and *c-myc* (Fig. 3a), suggesting that REST protects the expression of these self-renewal genes. Western blot analysis of the same cells also indicated that *Rest*$^{+/-}$ cells have substantially reduced levels of c-Myc and Oct4 proteins compared with the parental mouse ES cells (Fig. 3b), indicating that the downregulation of REST negatively affects self-renewal signals. To confirm further the role of REST in maintaining the self-renewal signals, we performed the same experiment shown in Fig. 1j, in which exogenously added REST maintained the self-renewal of mouse ES cells growing under differentiation conditions, and then analysed the expression of c-Myc and Oct4 by western blotting. As shown in Fig. 3c, the exogenously added REST, but not GFP, maintained c-Myc and Oct4 protein levels in mouse ES cells growing under differentiation conditions. Taken together, these results indicate that REST protects the expression of several self-renewal genes.

To determine how REST maintains the expression of multiple self-renewal genes, we examined the possible role of miRNAs in the self-renewal of REST-mediated mouse ES cells. We examined the miRNA profiles of wild-type and *Rest*$^{+/-}$ mouse ES cells growing under self-renewal conditions. The miRNAs from self-renewing wild-type mouse ES cells fell into two major groups, representing expressed and repressed miRNAs (Supplementary Fig. 1 and Supplementary Table 1). *Rest*$^{+/-}$ YHC and RRC cell lines had similar miRNA profiles, which were strikingly opposite to those of wild-type mouse ES cells. Furthermore, miR-124a, which was recently shown to be a target of REST[14], and miR-106a and miR-106b, whose predicted targets include *Rest*, were upregulated in *Rest*$^{+/-}$ cells, suggesting that regulation involving double-negative feedback loops may exist to maintain homeostasis. Using miRNA databases (http://cbio.mskcc.org/mirnaviewer and http://pictar.bio.nyu.edu/), we found a set of miRNAs that can potentially target self-renewal genes such as *Nanog*, *Sox2*, *Tbx3* and *c-myc* (Supplementary Table 2). These miRNAs were present only at low levels in self-renewing mouse ES cells and were present at higher levels in both the *Rest*$^{+/-}$ heterozygous cells and EB cells. To determine whether REST directly binds to the regulatory sequences (chromatin) of these miRNA-containing genes, we determined such potential REST binding sites (Supplementary Table 3) using the MatInspector module of the Genomatix database[15]. We found several potential REST-binding sites for each of these genes. Chromatin immunoprecipitation assays (Fig. 3d) and quantitative chromatin immunoprecipitation assays (Fig. 3e) in ES cells further revealed that REST was bound only to specific sites of the gene chromatin for each of these miRNAs. In a gain-of-function experiment similar to those shown in Figs 1j and 3c, in which exogenously added REST restored the self-renewal function of mouse ES cells growing under differentiation conditions, we analysed the expression of the miRNAs in these cells by qRT–PCR. As shown in Fig. 3f, levels of these miRNAs were compared with self-renewing mouse ES cells, expressed at higher levels in both *Rest*$^{+/-}$ heterozygous cells and EB cells, and expressed at lower levels in EB cells transfected with exogenous REST. To confirm further REST-mediated repression of the miRNAs, we performed a loss-of-function experiment in which we treated mouse ES cells with si*Rest* and analysed the expression of the miRNAs by qRT–PCR. As shown, si*Rest*-mediated knockdown of REST resulted in increased levels of miRNA expression when compared with non-targeting siRNA (Fig. 3g). To confirm that the si*Rest*-mediated knockdown of REST, which increased the expression of the miRNAs, also decreased the expression of the self-renewal genes, we

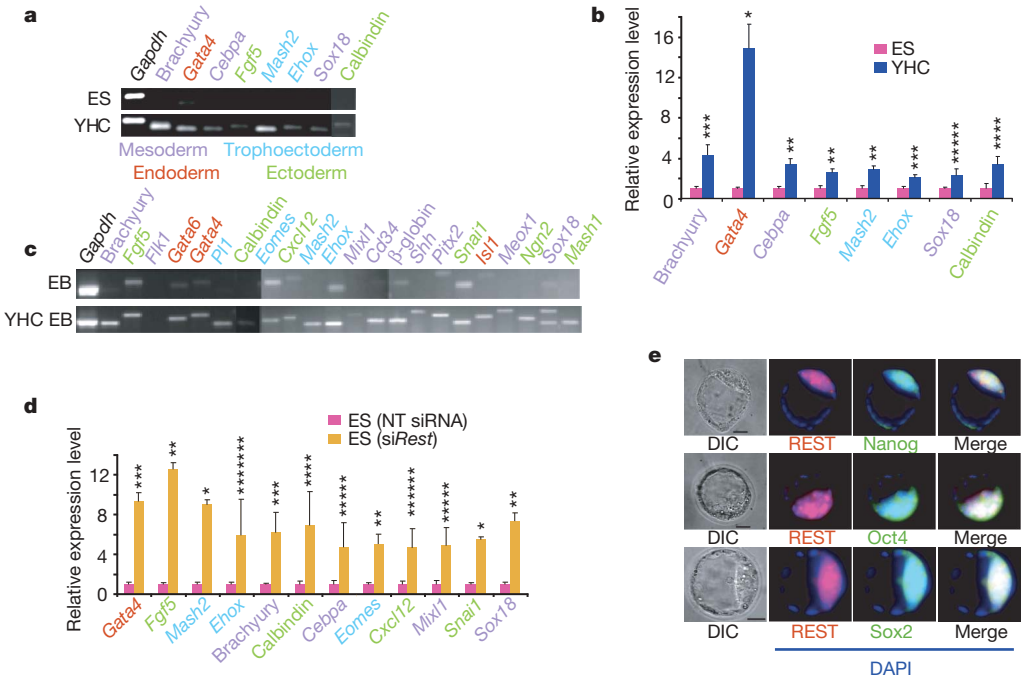


Figure 2 | Mouse ES cells with heterozygous deletion of *Rest* or mouse ES cells treated with si*Rest* differentiate into multiple lineages. **a**, **b**, RT–PCR (**a**) and qRT–PCR (**b**) analyses of total RNA isolated from ES and YHC cells grown under self-renewing conditions were performed using primers specific for markers for ectoderm, mesoderm, endoderm and trophectoderm. *Gapdh* was used as an internal control. *, $P = 0.001$; **, $P = 0.003$; ***, $P = 0.006$; ****, $P = 0.012$. The values are represented as mean ± s.d. ($n = 3$). **c**, EB cells derived from YHC cells show lineage markers that are either absent or significantly reduced in EB cells derived from wild-type mouse ES cells. RT–PCR analysis of total RNA isolated from wild-type EB cells and YHC EB cells was performed using specific primers for various differentiated lineage genes (shown above each lane). *Gapdh* was used as a loading control. **d**, qRT–PCR analysis of total RNA isolated from mouse ES cells treated with either non-targeting siRNA (NT) or si*Rest* was performed using primers specific for different lineage markers. Analysis was performed 5 days after transfection. *, $P < 0.0001$; **, $P = 0.001$; ***, $P = 0.002$; ****, $P = 0.004$; *****, $P = 0.005$; ******, $P = 0.01$. The values are represented as mean ± s.d. ($n = 3$). **e**, The inner cell mass of blastocysts showed co-expression of REST and self-renewal markers Oct4, Nanog and Sox2. Blastocysts were stained for nuclei (4,6-diamidino-2-phenylindole (DAPI), blue), REST (red) and self-renewal markers (green). Merged and differential interference contrast (DIC) images are also shown. Scale bar, 25 μm.

©2008 Nature Publishing Group

performed a qRT–PCR assay. As shown in Fig. 3h, si*Rest*-treated mouse ES cells showed decreased levels of *Rest*, *Oct4*, *Nanog* and *Sox2* expression. Thus, taken together, these results indicate that REST represses this set of miRNAs, and strongly suggest that at least some of the miRNAs, in turn, interfere with the expression of critical self-renewal regulators such as Oct4, Nanog and Sox2.

To determine whether the REST-regulated miRNAs are directly involved in maintaining self-renewal, we performed self-renewal assays in mouse ES cells after transfecting cells with individual precursors of these miRNAs. At least one of these precursor miRNAs, pre-miR-21, markedly decreased the self-renewing capacity of mouse ES cells by 60%, compared with that of an NT siRNA. In contrast, pre-miR-26a did not effectively alter the self-renewal efficiency when compared with the same NT siRNA (Fig. 4a). To confirm that the decreased self-renewal seen in mouse ES cells was actually due to the intracellular increase in the corresponding miRNAs, we measured levels of miR-21 and miR-26a by qRT–PCR. Indeed, the pre-miR-transfected cells showed more than tenfold and 300-fold increases in miR-21 and miR-26a expression, respectively (Fig. 4b).

To confirm that the effect of pre-miR-21 on self-renewal was specific, we transfected mouse ES cells with NT siRNA or a mixture of pre-miR-21 plus either NT siRNA or anti-miR-21 and performed self-renewal assays. As shown in Fig. 4c, pre-miR-21 plus NT siRNA lowered the self-renewal capacity by 30% when compared with NT siRNA alone. This lowering of self-renewal by 50 nM pre-miR-21 was half of what we had observed with 100 nM pre-miR-21 (Fig. 4a), indicating that under these conditions, 50–100 nM pre-miR-21 was in the linear range of its activity in suppressing self-renewal in mouse ES cells. Furthermore, the suppression of self-renewal by pre-miR-21 was rescued by anti-miR-21 but not by NT siRNA, indicating that the lowering of self-renewal was pre-miR-21-specific. We then performed qRT–PCR assays to determine the expression levels of the self-renewal markers *Oct4*, *Nanog*, *Sox2* and *c-myc* in pre-mir-21-treated cells. The results showed a corresponding decrease in the expression levels of these markers in pre-mir-21-treated cells (Fig. 4d). This is the first evidence of a single miRNA as a regulator of self-renewal in mouse ES cells. Interestingly, an earlier report indicated that miR-21 was one of the miRs whose expression was higher in differentiated compared with undifferentiated ES cells[16]. Taken together, our results indicate that REST-repressed miR-21 regulates the self-renewal of mouse ES cells. Thus, we report that REST has a newly discovered role in maintaining self-renewal and pluripotency of mouse ES cells. Furthermore, although REST can perform this function partly by direct repression of some of its target differentiation genes, such as calbindin, we provide evidence that REST functions via repressing the transcription of a specific miRNA, miR-21, which suppresses self-renewal by means of a corresponding loss of expression of the critical self-renewing regulators

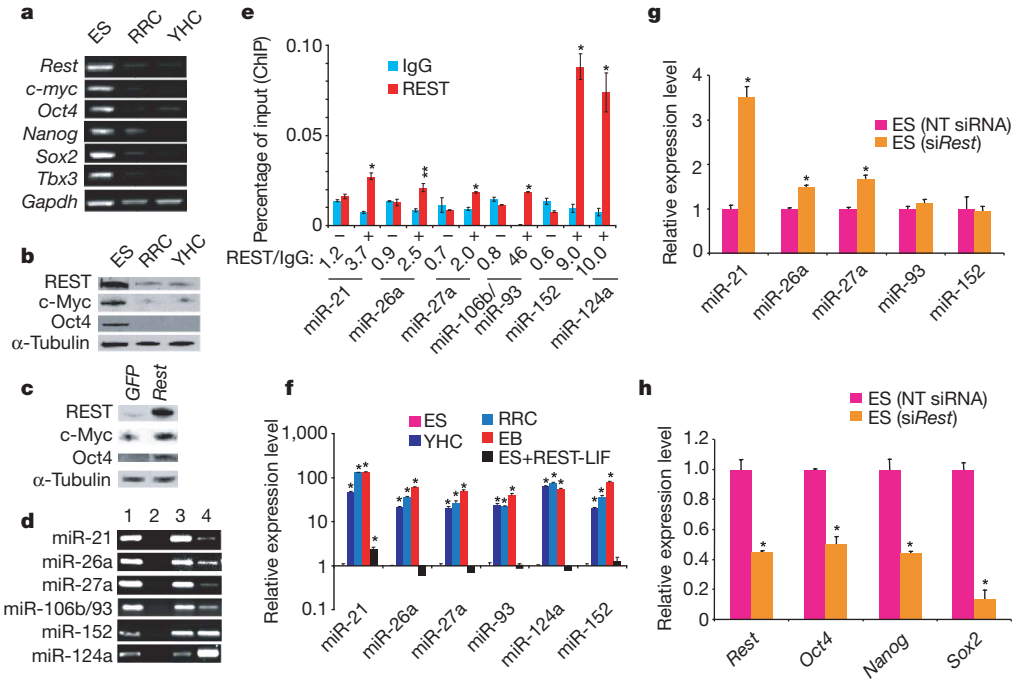


Figure 3 | REST maintains the expression of critical self-renewal regulators and represses expression of a set of miRNAs in mouse ES cells. **a**, **b**, Heterozygous deletion of *Rest* results in decreased expression of self-renewal genes. **a**, RT–PCR analysis of total RNA from ES, RRC and YHC cells. Genes are indicated on the left side of each panel. *Gapdh* was used as a loading control. **b**, Western blot analysis of whole-cell extracts from ES, RRC and YHC cells. Antibodies specific for REST, c-Myc and Oct4 proteins were used for the analysis. α-Tubulin was used as a loading control. **c**, Exogenously added *Rest*, but not *GFP*, maintained the expression of c-Myc and Oct4 in ES cells growing without LIF. Shown is a western blot analysis of whole-cell extracts from mouse ES cells transfected with plasmids encoding GFP or REST and then grown in the absence of LIF. Antibodies specific for REST, c-Myc and Oct4 proteins were used for the analysis. α-Tubulin was used as a loading control. **d**–**h**, REST suppresses the expression of miRNAs, which are associated with the absence of self-renewal regulators. **d**, **e**, ChIP (**d**) and quantitative ChIP (**e**) assays showed that REST binds to the gene chromatin of a set of miRNAs in ES cells (1, input; 2, IgG, negative control; 3, anti-H3 antibody, positive control; 4, anti-REST antibody; + and − represent predicted RE-1 binding sites occupied and unoccupied by REST, respectively). *, $P < 0.0001$; **, $P = 0.001$. The values of three replicates are represented as mean ± s.d. **f**, qRT–PCR analysis of ES, YHC, RRC, EB and EB plus REST showed that expression of the miRNAs (shown in **d** and **e**) was lower in ES than in YHC, RRC and EB cells. The higher expression of miRNAs in EB cells could be extinguished in a gain-of-function experiment in the presence of exogenously added REST. *, $P < 0.0001$. The values are represented as mean ± s.d. ($n = 3$). **g**, The expression of the miRNAs shown in **d** was upregulated in a loss-of-function experiment when mouse ES cells were treated with si*Rest* but not when they are treated with non-targeting siRNA (NT). *, $P < 0.0001$. The values are represented as mean ± s.d. ($n = 3$). Analysis was performed 3 days after transfection. **h**, si*Rest*-mediated knockdown of REST that produced increased expression of the miRNAs in **g** corresponded to the decreased expression of *Rest* and the known self-renewal genes *Oct4*, *Nanog* and *Sox2*. A qRT–PCR assay with mouse ES cells treated with si*Rest* and NT siRNA is shown. Analysis was performed 5 days after transfection. *, $P < 0.0001$. The values are represented as mean ± s.d. ($n = 3$).

©2008 Nature Publishing Group

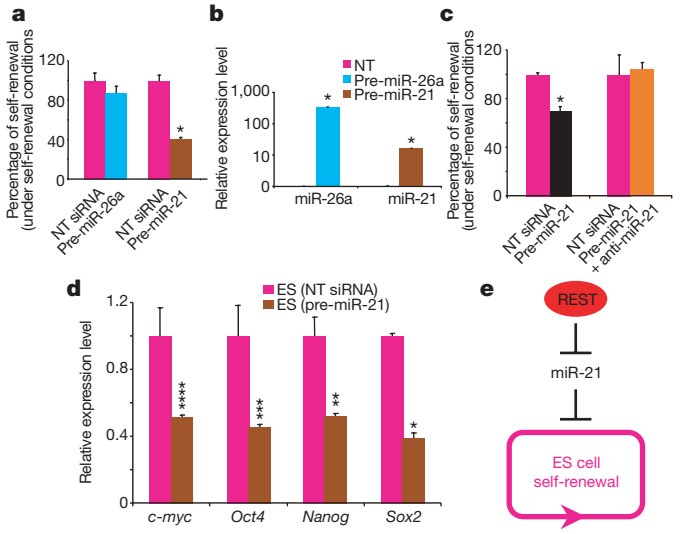


Figure 4 | miR-21 regulates self-renewal in mouse ES cells. **a**, Addition of exogenous precursor miR-21 (pre-miR-21), but not non-targeting siRNA (NT) or pre-miR-26a, caused a decrease of self-renewal in mouse ES cells. *, $P < 0.0001$. The error bars correspond to three replicates ($n = 3$) and show mean ± s.d. **b**, Addition of exogenous pre-miR-21 and exogenous pre-miR-26a, but not NT siRNA, to mouse ES cells resulted in increased levels of their corresponding miRNA expression. Results of qRT–PCR assays performed with NT-siRNA-, pre-miR-26a- and pre-miR-21-transfected cells 1 day after transfection are shown. *, $P < 0.0001$. The values are represented as mean ± s.d. ($n = 3$). **c**, The effect of miR-21 on loss of self-renewal was specific. Addition of exogenous pre-miR-21 along with anti-miR-21 rescued the loss of self-renewal by pre-miR-21 alone. *, $P < 0.0001$ (there was no significant difference between NT siRNA and pre-miR-21 plus anti-miR-21 values). The error bars correspond to three replicates ($n = 3$) and show mean ± s.d. **d**, Addition of exogenous pre-miR-21, but not NT siRNA, to mouse ES cells resulted in loss of self-renewal markers *Oct4*, *Nanog*, *Sox2* and *c-myc*. qRT–PCR after transfection of mouse ES cells with pre-miR-21 and NT siRNA was performed 5 days after transfection to detect the expression levels of the self-renewal markers. *, $P < 0.0001$; **, $P = 0.002$; ***, $P = 0.007$; ****, $P = 0.008$. The values are represented as mean ± s.d. ($n = 3$). **e**, Proposed model of REST-mediated maintenance of self-renewal through miR-21. Rest maintains self-renewal and pluripotency of mouse ES cells by suppressing miR-21 expression.

Oct4, Nanog, Sox2 and c-Myc (Fig. 4e). Whether miR-21 directly suppresses the expression of these self-renewing regulators, thereby inhibiting self-renewal, or whether miR-21 inhibits self-renewal by a different mechanism, with consequent suppression of these self-renewal regulators, is unclear. Of note, bioinformatics-predicted miR-21-binding sites are present in the mRNAs of *Sox2* and *Nanog* but not those of *Oct4* and *c-myc* (Supplementary Table 2). If these sites truly function as miR-21 target sites, and if miR-21 target sites are indeed not present in *Oct4* and *c-myc* mRNAs, then these observations indicate that the miR-21-mediated regulation of mouse ES cell self-renewal involves a regulatory cascade initiated by the loss of Sox2 and Nanog expression that leads to the loss of expression of Oct4 and c-Myc. It is noteworthy that Oct4, Nanog and Sox2 were found to co-occupy *Rest* promoter/enhancer sequences, apparently as an activator complex[17]. Furthermore, REST formed part of the Nanog–Oct4 complex[18] and was predicted to be a major component of an Oct4–Sox2–Nanog network in ES cells[19]. Taken together with our data presented here, these observations indicate that REST is a new element of the interconnected regulatory network required to maintain the self-renewal and pluripotency of cultured mouse ES cells.

METHODS SUMMARY

Embryonic stem cells (E14Tg2a) were cultured in the presence of LIF on gelatin-coated dishes. Embryoid bodies were grown on bacterial dishes in the absence of LIF and β-mercaptoethanol. Alkaline phosphatase staining after loss- and gain of function was carried out using a detection kit from Chemicon. Chromatin immunoprecipitation (ChIP) assays were performed as described in the Methods. For the miRNA microarrays, total RNA was prepared using TRIzol, and RNA samples were analysed by LC Sciences on their microarray platform with a probe set based on Sanger version 9.0. Blastocysts were obtained from C57BL/6 NCr mice, fixed in 4% paraformaldehyde, permeabilized with 0.25% Triton X-100, and were then double immunostained with antibodies against REST, Sox2, Nanog and Oct4. Primers used for RT–PCR and ChIP are available in the Methods.

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

Received 21 November 2007; accepted 22 February 2008.
Published online 23 March 2008.

1. Ballas, N., Grunseich, C., Lu, D. D., Speh, J. C. & Mandel, G. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* **121**, 645–657 (2005).
2. Ballas, N. & Mandel, G. The many faces of REST oversee epigenetic programming of neuronal genes. *Curr. Opin. Neurobiol.* **15**, 500–506 (2005).
3. Ooi, L. & Wood, I. C. Chromatin crosstalk in development and disease: lessons from REST. *Nature Rev. Genet.* **8**, 544–554 (2007).
4. Coulson, J. M. Transcriptional regulation: cancer, neurons and the REST. *Curr. Biol.* **15**, R665–R668 (2005).
5. Majumder, S. REST in good times and bad: roles in tumor suppressor and oncogenic activities. *Cell Cycle* **5**, 1929–1935 (2006).
6. Su, X., Kameoka, S., Lentz, S. & Majumder, S. Activation of REST/NRSF target genes in neural stem cells is sufficient to cause neuronal differentiation. *Mol. Cell. Biol.* **24**, 8018–8025 (2004).
7. Watanabe, Y. *et al.* Conversion of myoblasts to physiologically active neuronal phenotype. *Genes Dev.* **18**, 889–900 (2004).
8. Lawinger, P. *et al.* The neuronal repressor REST/NRSF is an essential regulator in medulloblastoma cells. *Nature Med.* **6**, 826–831 (2000).
9. Fuller, G. N. *et al.* Many human medulloblastoma tumors overexpress repressor element-1 silencing transcription (REST)/neuron-restrictive silencer factor, which can be functionally countered by REST-VP16. *Mol. Cancer Ther.* **4**, 343–349 (2005).
10. Su, X. *et al.* Abnormal expression of REST/NRSF and Myc in neural stem/progenitor cells causes cerebellar tumors by blocking neuronal differentiation. *Mol. Cell. Biol.* **26**, 1666–1678 (2006).
11. Pan, G. & Thomson, J. A. Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res.* **17**, 42–49 (2007).
12. Pritsker, M., Ford, N. R., Jenq, H. T. & Lemischka, I. R. Genomewide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **103**, 6946–6951 (2006).
13. Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38**, 431–440 (2006).
14. Conaco, C., Otto, S., Han, J. J. & Mandel, G. Reciprocal actions of REST and a microRNA promote neuronal identity. *Proc. Natl Acad. Sci. USA* **103**, 2422–2427 (2006).
15. Cartharius, K. *et al.* MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics* **21**, 2933–2942 (2005).
16. Houbaviy, H. B., Murray, M. F. & Sharp, P. A. Embryonic stem cell-specific microRNAs. *Dev. Cell* **5**, 351–358 (2003).
17. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
18. Wang, J. *et al.* A protein interaction network for pluripotency of embryonic stem cells. *Nature* **444**, 364–368 (2006).
19. Zhou, Q., Chipperfield, H., Melton, D. A. & Wong, W. H. A gene regulatory network in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **104**, 16438–16443 (2007).

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

Acknowledgements We would like to thank the National Institutes of Health for funding (grants CA97124 and CA81255 to S.M.). M.N.K. and S.K.S. were both recipients of the Dodie Hawn Fellowship in Cancer Genetics.

Author Contributions S.K.S. and M.N.K. participated in all aspects including manuscript preparation; J.P.-T. conducted blastocyst experiments; H.A. carried out immunofluorescence studies; S.M. did project planning and manuscript preparation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.M. (smajumder@mdanderson.org).

©2008 Nature Publishing Group

METHODS

ES cell lines and culture conditions. Mouse ES cells (E14Tg2A) obtained from Bay Genomics were cultured under self-renewal conditions (in the presence of LIF, cultured on adherent tissue culture dishes) without feeder cells in 1,000 units ml^{-1} LIF (ESGRO) on gelatin-coated tissue culture dishes[20]. Embryoid bodies in the absence of LIF and β-mercaptoethanol on non-adherent tissue culture dishes were grown for 3 days in bacterial dishes as described[1]. The $Rest^{+/-}$ cell lines YHC334 (YHC) and RRC160 (RRC) that have gene traps at *Rest* exons 1 and 2, respectively, were obtained from Bay Genomics. Both of these lines were derived from parental E14Tg2A mouse ES cells, the same cells that we used for all of our wild-type mouse ES cell experiments.

ChIP assays. ChIP assays of cultured mouse ES cells and EBs were performed as described[21]. We used antibodies against mouse IgG (2 μl; catalogue number I 5381, Sigma), REST (2 μl; 07-579, Upstate) and histone H3 (2 μl; ab1791, Abcam). One microgram of each antibody and approximately 15 micrograms of chromatin were added to each reaction mixture and incubated overnight at 4 °C. The mixtures were then incubated with 33 μl of protein A-agarose beads (Amersham Pharmacia Biotech AB; pre-equilibrated with 1 mg ml^{-1} bovine serum albumin and 0.3 mg ml^{-1} salmon sperm DNA).

The PCR primer sequences used are as follows: miR-93/miR-106b forward 5′-CCCCCAAAACCAGTATCCTT-3′, reverse 5′-TCGTACTTCCCGGATCACTC-3′; miR-27a forward 5′-AAATCCCCAGAAGCTGGAGT-3′, reverse 5′-AGCAC-TTGAAAGGCAAAGGA-3′; miR-21 forward 5′-AGGGCAGGAAGATGACA-CAC-3′, reverse 5′-GGGCTGATGAGCACTAAGGA-3′; miR-26a forward 5′-GCCTAACCCAAGAAGGGAAA-3′, reverse 5′-TCCCTCTCATCTGGACAACC-3′; miR-152 forward 5′-TGGTCTCTGTCCAGCACAAC-3′, reverse 5′-CAGGT-CACAGCTGCACTCAT-3′; miR-124a forward 5′-CTCTGCGTGTTCACAGC-GG-3′, reverse 5′-CTCTTGGCATTCACCGCGTG-3′.

Quantitative ChIP PCR was performed using the same extracts used above. Briefly, qPCR was performed and analysed using real-time PCR (ABI prism 7500). Primers were designed by Primer Express 2.0 (Applied Biosystems) to amplify 60- to 150-bp amplicons. Amplicons were measured by SYBR Green fluorescence (SYBR Green Master mix, Applied Biosystems) in 10-μl reactions. Reactions were performed in triplicates. The amount of product was determined relative to a standard curve of input chromatin. Melting curves showed that PCRs yielded single product. C_t values for each sample were normalized against that of input DNA and percentage recovery is plotted. The primers used for qPCR will be available on request.

Western blotting, RT–PCR and miRNA microarray. Whole-cell extracts were prepared, and approximately 10 μg of proteins were resolved on SDS–PAGE using antibodies against REST (07-579, Upstate), c-Myc (sc-764, Santa Cruz Biotechnology), Oct4 (ab19857, Abcam), Sox2 (ab15830, Abcam), actin (sc-1616, Santa Cruz) and α-tubulin (MMS-407R, Covance), following the manufacturers' recommendations. Total RNA was extracted using TRIzol (Invitrogen), and approximately 50 ng of total RNA was used as a template. RT–PCR was performed using a OneStep RT–PCR kit (Qiagen).

The following gene-specific primers were designed: *Rest* forward 5′-AGCGAGTACCACTGGAGGAA-3′, reverse 5′-CTGAATGAGTCCGCATGTGT-3′; calbindin forward 5′-GCTTCTATCTGGCGGAAGG-3′, reverse 5′-TGTCATC-TGGCTACCTTCCC-3′; *Pitx2* forward 5′-CGGCAGAGGACTCATTTCAC-3′, reverse 5′-GTACGAATAGCCGGGGTACA-3′; *Snai1* forward 5′-CTTGTGT-CTGCACGACCTGT-3′, reverse 5′-CTTCACATCCGAGTGGGTTT-3′; *Sox18* forward 5′-AACAAAATCCGGATCTGCAC-3′, reverse 5′-CGAGGCCGGTA-CTTGTAGTT-3′; *Cebpa* forward 5′-CCGACTTCTACGAGGTGGAG-3′, reverse 5′-TGGCCTTCTCCTGCTGTC-3′. The following primers were based on published sequences: *Gapdh*, *Oct4*, *Nanog*, *Gata6*, *Fgf5*, brachyury, *Flk1* and *Pl1* (ref. 22); *Mash1*, *Mash2*, β-globin, *Cd34*, *Cxcl12*, *Ehox*, *Eomes*, *Gata4*, *Isl1*, *Meox1*, *Mixl1*, *Ngn2*, *Shh* and *Tbx3* (ref. 23).

For quantitative RT–PCR, first-strand synthesis was performed using Applied Biosystems' high-capacity cDNA reverse transcription kit (4368814). Real-time PCR was carried out using cDNAs with SYBR Green PCR master mix from Applied Biosystems (4309155). Reactions were carried out in triplicates using ABI prism 7500 and the analysis was performed similar to that described for quantitative ChIP PCR above, except that ΔC_t was obtained after normalization with C_t of *Gapdh* for mouse-specific primers and that of actin for human-specific primers. Mouse-specific primers will be provided on request; human-specific primers are as published[24].

For the miRNA microarrays, total RNA was prepared using TRIzol, and RNA samples were analysed by LC Sciences on their microarray with a probe set based on Sanger version 9.0. Multiple sample analysis involves normalization, data adjustment, *t*-test/analysis of variance (ANOVA) analysis, and clustering. Normalization is carried out using a cyclic Lowess (locally weighted regression) method. The normalization is to remove system-related variations, such as sample amount variations, different labelling dyes, and signal gain differences of scanners so that biological variations can be faithfully revealed. Data adjustment includes data filtering, $\log_2$ transformation, and gene centring and normalization. The data filtering removes genes (or miRNAs) with (normalized) intensity values below a threshold value of 32 across all samples. The $\log_2$ transformation converts intensity values into the $\log_2$ scale. Gene centring and normalization transform the $\log_2$ values using the mean and the standard deviation of individual genes across all samples using the following formula:

$$\text{Value} = [(\text{value}) - \text{mean}(\text{gene})]/[\text{s.d.}(\text{gene})]$$

In certain cases, expression ratios of dual-sample arrays are used in clustering analysis. Gene normalization would then be performed using the following formula:

$$\text{Value} = [\log_2(\text{ratio})]/[\text{s.d.}(\log_2(\text{ratio}))]$$

The *t*-test is performed between 'control' and 'test' sample groups, with each group containing at least two samples. *t*-values are calculated for each miRNA, and *P*-values are computed from the theoretical *t*-distribution. MicroRNAs with *P*-values below a critical *P*-value (typically 0.01) are selected for cluster analysis. The clustering is done using a hierarchical method and is performed with average linkage and euclidean distance metric.

ANOVA is an extension of the *t*-test to more than two experimental conditions. It picks out miRNAs that have significant differences in means across three or more groups of samples, with each group containing at least two samples. In this analysis, *P*-values are computed from the *F*-distribution. In certain cases, such as time point experiments, in which only one sample is collected and assayed for each condition or time point, repeating sets of reporters (or probes) on each array may be used as a 'group' in an ANOVA analysis.

All data processes, except clustering plot, were carried out using in-house developed computer programs. The clustering plot is generated using TIGR MeV (multiple experimental viewer) software from The Institute for Genomic Research.

miRNA validation was carried out using Ambion's *mir*Vana qRT–PCR miRNA detection kit (AM1558). *mir*Vana qRT–PCR primer sets for the miRNA of interest were also obtained from Ambion (30000–30999). Reactions were performed as per the manufacturer's recommendations. The values obtained were normalized with primer set corresponding to 5S rRNA (AM 30302).

RNAi-mediated knockdown, exogenous expression of miR-21, anti-miR-21 and gain-of-function experiments. RNAi-mediated knockdown of REST and Oct4 were carried out per the manufacturer's protocol (Dharmacon). si*Rest* (60 pmol), si*Oct4* (60 pmol) and non-targeting siRNA (60 pmol) were transfected using the Amaxa nucleofection method. Colonies were assayed for alkaline phosphatase activity 3 days after transfection.

Amaxa nucleofection was also used to deliver precursor miR-21 and anti-miR-21 molecules into mouse ES cells. Transfection was carried out using 100 pmol of precursor or 50 pmol of miR-21 and anti-miR-21 each using 4×10^6 cells. Non-targeting siRNA was used as a control. Delivery and mature miR-21 product was inspected in cells by qRT–PCR 1 day after transfection, as described above.

For the gain-of-function experiments, mouse ES cells were transfected with 2 μg of plasmid DNA using Lipofectamine 2000. The Flag–REST construct was a gift from G. Mandel. The transfected cells were selected with G418 (125 μg ml^{-1}) under self-renewing conditions in the presence of LIF for 3 days. After selection, we grew the cells under differentiation conditions in the absence of LIF for 4 days, and subjected them to self-renewal assays. An alkaline phosphatase assay was then performed using an alkaline phosphatase detection kit (Chemicon) after 4 days. Percentages of self-renewing colonies from three independent experiments were plotted. Standard deviation was calculated and is depicted by error bars in the figures.

Immunofluorescence microscopy of blastocysts. Blastocysts from C57BL/6 NCr mice were fixed in 4% paraformaldehyde, permeabilized with 0.25% Triton X-100, and blocked with 1% goat serum in PBS[25]. The blastocysts were then double immunostained with mouse anti-REST antibody (raised against the amino-terminal portion of human REST, A372) in combination with antibodies against Sox2 (Abcam, AB15830), Oct4 (Abcam, AB19857), Nanog (Abcam, AB21603) followed by Alexa-Fluor-488-conjugated secondary antibody detecting mouse IgG and Fluor-555-conjugated secondary antibody detecting rabbit IgG (Molecular Probes). Images were captured with a Nikon microscope (Nikon Eclipse TE 2000U).

20. Nichols, J., Evans, E. P. & Smith, A. G. Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development* **110**, 1341–1348 (1990).

©2008 Nature Publishing Group

21. Cui, R. *et al.* Family members p53 and p73 act together in chromatin modification and direct repression of α-fetoprotein transcription. *J. Biol. Chem.* **280,** 39152–39160 (2005).
22. Kaji, K. *et al.* The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nature Cell Biol.* **8,** 285–292 (2006).
23. Pritsker, M., Ford, N. R., Jenq, H. T. & Lemischka, I. R. Genomewide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells. *Proc. Natl Acad. Sci. USA* **103,** 6946–6951 (2006).
24. Cai, J. *et al.* Assessing self-renewal and differentiation in human embryonic stem cell lines. *Stem Cells* **24,** 516–530 (2006).
25. Zhang, J. *et al.* Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of *Pou5f1*. *Nature Cell Biol.* **8,** 1114–1123 (2006).

©2008 **Nature Publishing Group**

LETTERS

Long-term haematopoietic reconstitution by $Trp53^{-/-}p16^{Ink4a-/-}p19^{Arf-/-}$ multipotent progenitors

Omobolaji O. Akala[1,2], In-Kyung Park[3], Dalong Qian[2], Michael Pihalja[3], Michael W. Becker[4] & Michael F. Clarke[2]

Haematopoiesis is maintained by a hierarchical system where haematopoietic stem cells (HSCs) give rise to multipotent progenitors, which in turn differentiate into all types of mature blood cells[1]. HSCs maintain themselves for the lifetime of the organism because of their ability to self-renew. However, multipotent progenitors lack the ability to self-renew, therefore their mitotic capacity and expansion potential are limited and they are destined to eventually stop proliferating after a finite number of cell divisions[1,2]. The molecular mechanisms that limit the proliferation capacity of multipotent progenitors and other more mature progenitors are not fully understood[2,3]. Here we show that bone marrow cells from mice deficient in three genes genetically downstream of Bmi1—*$p16^{Ink4a}$*, *$p19^{Arf}$* and *Trp53* (triple mutant mice; *$p16^{Ink4a}$* and *$p19^{Arf}$* are alternative reading frames of the same gene (also called *Cdkn2a*) that encode different proteins)—have an approximately 10-fold increase in cells able to reconstitute the blood long term. This increase is associated with the acquisition of long-term reconstitution capacity by cells of the phenotype c-kit$^+$Sca-1$^+$Flt3$^+$CD150$^-$CD48$^-$Lin$^-$, which defines multipotent progenitors in wild-type mice[4–6]. The pattern of triple mutant multipotent progenitor response to growth factors resembles that of wild-type multipotent progenitors but not wild-type HSCs. These results demonstrate that p16^{Ink4a}/p19^{Arf} and Trp53 have a central role in limiting the expansion potential of multipotent progenitors. These pathways are commonly repressed in cancer, suggesting a mechanism by which early progenitor cells could gain the ability to self-renew and become malignant with further oncogenic mutations.

Bmi1 is necessary for the maintenance of adult HSCs and neural stem cells[7–9].The *$p16^{Ink4a}$*/*$p19^{Arf}$* locus is one of the major targets repressed by Bmi1 (refs 8, 10). In neural stem cells, *$p16^{Ink4a}$* or *$p19^{Arf}$* deficiency partially restored the ability of *Bmi1*-deficient stem cells to self-renew[7,11]. We analysed the effect of individually deleting either *$p16^{Ink4a}$*, *$p19^{Arf}$* or *Trp53* in *Bmi1*-deficient mice and found that in the blood system, *Trp53* deficiency, but not *$p19^{Arf}$* or *$p16^{Ink4a}$* deficiency, partially rescues *$Bmi1^{-/-}$* HSCs (Supplementary Figs 2 and 3 and Supplementary Table 1).

Because the deletion of either the *$p16^{Ink4a}$*/*$p19^{Arf}$* locus[12] or *Trp53* seems to have a role in the rescue of HSCs in *Bmi1*-deficient mice, we examined the impact of deleting *$p19^{Arf}$*, *$p16^{Ink4a}$*/*$p19^{Arf}$*, *Trp53* and *$p16^{Ink4a}$*/*$p19^{Arf}$*/*Trp53* on the mature cells of the haematopoietic system. Analysis of the liver, spleen, bone marrow and peripheral blood from mutant mice showed no defects in the output of mature cells (Supplementary Figs 4 and 5). Neutrophil, lymphocyte and monocyte counts from mutant mice were normal (Supplementary Fig. 4), as were haematocrit and platelet counts (data not shown). In the bone marrow of mutant mice, levels of B220$^+$ cells, Ter119$^+$ cells, Gr-1$^+$ cells and Mac-1$^+$ cells were normal (Supplementary Fig. 4). Analysis of spleen and thymus also showed normal levels of mature haematopoietic cells (Supplementary Fig. 4). Therefore, the deletion of *$p19^{Arf}$*, *Trp53*, *$p16^{Ink4a}$*/*$p19^{Arf}$* and *$p16^{Ink4a}$*/*$p19^{Arf}$*/*Trp53* does not seem to compromise haematopoietic differentiation.

We wished to know whether *$p16^{Ink4a}$*, *$p19^{Arf}$* or *Trp53* regulate self-renewal in blood cells. To do this, limiting dilution analysis was done to determine the frequencies of bone marrow cells responsible for long-term haematopoietic reconstitution in mutant mice. Whole bone marrow cells from approximately 8-week-old wild-type and mutant mice were injected at different doses into lethally irradiated recipient mice along with a radioprotective dose of recipient bone marrow (Supplementary Table 2). In *$p19^{Arf-/-}$*, *$Trp53^{-/-}$* and *$p16^{Ink4a-/-}p19^{Arf-/-}$* bone marrow the frequency of cells responsible for long-term reconstitution was not significantly different from that of wild-type mice (Table 1). These results agree with previous studies that show that in young mice, deficiency of *$p16^{Ink4a}$*, *$p19^{Arf}$*, *Trp53* and both *$p16^{Ink4a}$* and *$p19^{Arf}$* does not significantly affect the long-term reconstitution frequency of HSCs and has more of an effect in the context of stem-cell ageing[13–15]. Notably, the frequency of long-term reconstituting cells in triple mutant bone marrow was at least 10-fold higher than that of wild-type bone marrow (Table 1).

Our results suggest that triple deletion of *$p16^{Ink4a}$*, *$p19^{Arf}$* and *Trp53* caused a striking increase in the frequency of long-term blood-repopulating cells. We therefore decided to examine further the bone marrow of triple mutant mice. Because there was only a modest increase in the frequency of bone marrow cells with a phenotype of HSCs in triple mutant mice (Fig. 1a, b), we concluded that the increased repopulating activity in the triple mutant mice was probably not due to the expansion of the overall numbers of HSCs but rather to acquisition of stem-cell-like properties by other cells. Therefore, we compared the long-term reconstitution capacity of

Table 1 | Frequency of long-term reconstituting cells determined by limiting dilution analysis

Genotype	One long-term reconstituting cell in:	Range	Frequency (%)
Wild type	92,049	70,278–120,564	0.0011
$p19^{Arf-/-}$	55,072	43,857–69,154	0.0018
$Trp53^{-/-}$	50,317	35,056–72,222	0.0020
$p16^{Ink4a-/-}p19^{Arf-/-}$	53,002	41,344–67,948	0.0019
$p16^{Ink4a-/-}p19^{Arf-/-}Trp53^{-/-}$	7,657	5,221–11,230	0.0131*

Using the limiting dilution analysis from Supplementary Table 2, the frequency of long-term reconstituting cells was calculated according to Poisson statistics using L-Calc software.
* *$p16^{Ink4a-/-}p19^{Arf-/-}Trp53^{-/-}$* bone marrow showed a 10-fold increase in the frequency of cells capable of long-term reconstitution when measured 20 weeks after transplantation compared to wild-type cells (two-tailed *t*-test; $P < 0.0005$). The other mutant bone marrow did not show a marked difference in the frequency of long-term reconstituting cells.

[1]Cellular and Molecular Biology Graduate Program, University of Michigan, 2966 Taubman Medical Library, Ann Arbor, Michigan 48109-0619, USA. [2]Stanford Institute for Stem Cell Biology and Regenerative Medicine and Division of Hematology/Oncology, Internal Medicine, Stanford University, 1050 Arastradero Road, Palo Alto, California 93404, USA. [3]Department of Hematology/Oncology, Internal Medicine, University of Michigan, 1500 East Medical Center Drive, Ann Arbor, Michigan 48109, USA. [4]Division of Hematology/Oncology, University of Rochester, Rochester, New York 14642, USA.

©2008 Nature Publishing Group

both HSCs (defined as CD150$^+$Sca-1$^+$c-kit$^+$CD48$^-$Lin$^-$) and multipotent progenitors (defined as Sca-1$^+$c-kit$^+$CD150$^-$CD48$^-$Lin$^-$) from wild-type and triple mutant mice. We performed two independent experiments where we injected either 100 or 500 HSCs or multipotent progenitors from four different triple mutant mice into lethally irradiated recipients along with a radioprotective dose of recipient-type bone marrow. Contribution to the generation of lymphocytes, monocytes and granulocytes after 12 weeks is an indicator of long-term reconstitution[1]. Consistent with previous reports[4,16,17], wild-type HSCs, but not multipotent progenitors, contributed to the long-term generation of monocytes and granulocytes (Fig. 2a, c). Recipient mice injected with HSCs from the bone marrow of triple mutant mice showed long-term reconstitution for multiple lineages (Fig. 2b and Supplementary Table 3). Notably, recipient mice injected with multipotent progenitors from each of the four triple mutant mice were also long-term multi-lineage reconstituted (Fig. 2d and Supplementary Table 3). Contribution to haematopoiesis was observed at 14 to 20 weeks (Fig. 2b, d and Supplementary Figs 6 and 7). This suggests that the greater frequency of cells capable of long-term reconstituting ability observed during limiting dilution analysis of triple mutant bone marrow was due to both the HSC and multipotent progenitor populations of the mutant mice. On secondary transplantation of approximately 2×10^6 whole bone marrow cells from recipient mice engrafted for 20 weeks with 100 triple mutant HSCs or multipotent progenitors, we observed contribution to donor CD3$^+$ lymphocytes and Gr-1$^+$ myeloid cells in lethally irradiated recipients (Supplementary Table 3 and Supplementary Fig. 8) 12 weeks after transplantation, indicating that cells with the phenotype of multipotent progenitors in triple mutant mice

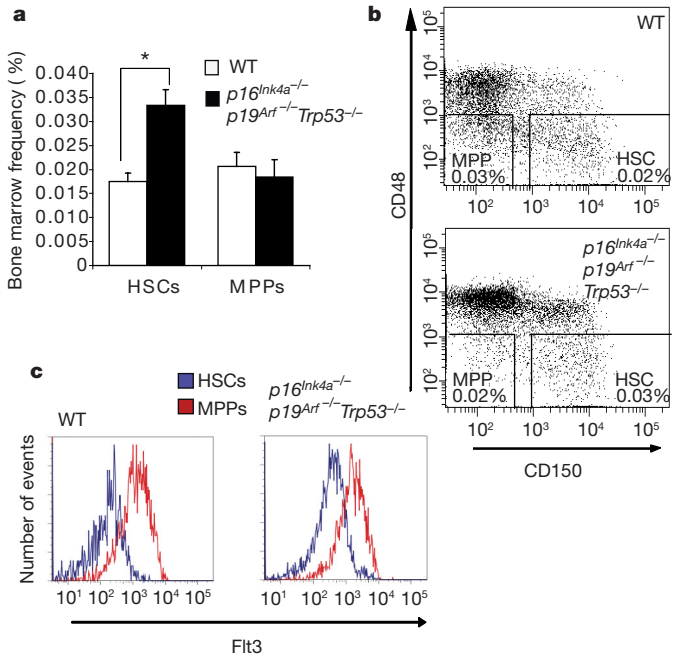


Figure 1 | Immunophenotypic frequency of HSCs and multipotent progenitors in wild-type and *p16$^{Ink4a-/-}$p19$^{Arf-/-}$Trp53$^{-/-}$* mice. a, The immunophenotypic frequency of HSCs (CD150$^+$Sca-1$^+$c-kit$^+$CD48$^-$Lin$^-$) and multipotent progenitors (MPPs; Sca-1$^+$c-kit$^+$CD150$^-$CD48$^-$Lin$^-$) was assessed by flow cytometry. There was a modest increase in HSC frequency in *p16$^{Ink4a-/-}$p19$^{Arf-/-}$Trp53$^{-/-}$* mice compared to wild type (WT) and no significant difference in the multipotent progenitor frequency. The averages of wild-type ($n = 7$) and triple mutant mice ($n = 5$) are shown. Asterisk, $P < 0.005$ (Student's *t*-test). Error bars denote s.e.m. **b**, A representative fluorescence-activated cell sorting (FACS) plot gated on Sca-1$^+$c-kit$^+$Lin$^-$ bone marrow cells depicting HSC and multipotent progenitor frequency. **c**, Mean intensity fluorescence of Flt3 expression of wild-type ($n = 9$) and triple mutant ($n = 2$) HSCs and multipotent progenitors.

can generate mature blood cells of both lymphoid and myeloid lineages for extended periods of time.

Immunophenotypic analysis by flow cytometry showed that there was at most twice the frequency of HSCs by phenotype (CD150$^+$Sca-1$^+$c-kit$^+$CD48$^-$Lin$^-$) and a relatively identical frequency of multipotent progenitors by phenotype (Sca-1$^+$c-kit$^+$CD150$^-$CD48$^-$Lin$^-$) in *p16$^{Ink4a-/-}$p19$^{Arf-/-}$Trp53$^{-/-}$* bone marrow compared to wild-type marrow (Fig. 1a, b). We also used Flt3, which is expressed by normal multipotent progenitors but not HSCs, as an alternative marker to distinguish HSCs from multipotent progenitors[5,6]. We found that the wild-type and triple mutant Sca-1$^+$c-kit$^+$CD150$^-$CD48$^-$Lin$^-$ cells expressed higher levels of this marker than did cells with a stem cell phenotype of CD150$^+$Sca-1$^+$c-kit$^+$CD48$^-$Lin$^-$ (Fig. 1c). This further suggests that the phenotype of HSCs and multipotent progenitors is not altered in the triple mutant mice.

Next, we performed three independent *in vitro* proliferation experiments to provide evidence that triple mutant CD150$^+$Sca-1$^+$c-kit$^+$Flt3$^+$CD48$^-$Lin$^-$ cells were indeed multipotent progenitors and to understand the mechanisms by which their lifespan was extended. We used culture conditions that allow HSCs to expand *in vitro* in StemSpan serum-free media containing the growth factors SCF, TPO, IGF-2, FGF-1 and Angptl3 (STIFA media)[18–20]. When cultured in STIFA medium, wild-type multipotent progenitors proliferated fivefold less than wild-type HSCs (Fig. 3a), and the addition of Flt3 ligand to the media doubled the proliferative rate of the multipotent progenitors (Fig. 3a). Notably, when cultured in STIFA media, triple mutant multipotent progenitors also proliferated about fivefold less than triple mutant HSCs and still responded to Flt3 ligand (Fig. 3a). This suggests that triple mutant multipotent progenitors, although capable of long-term reconstitution, display an *in vitro* functional behaviour that mirrors that of wild-type multipotent progenitors and are different from HSCs. In two independent experiments we also determined the percentage of cells undergoing apoptosis in STIFA media in the case of HSCs and STIFA media plus Flt3 ligand for multipotent progenitors, measured by annexin V staining. Triple mutant HSCs and multipotent progenitors had threefold and twofold decreases, respectively, in the percentage of cells undergoing apoptosis compared to wild-type cells (Supplementary Fig. 9). The proliferation advantage of triple mutant HSCs and multipotent progenitors could be partly due to a decrease in the rate of apoptosis in the triple mutants compared to their wild-type counterparts.

To examine further the proliferative capacity of wild-type and triple mutant HSCs and multipotent progenitors we observed the colony-forming capacity of single cells double-sorted into 96-well plates in Methocult GF M3434 media (Methocult) and Methocult GF M3434 media supplemented with interleukin (IL)-11, granulocyte–macrophage colony-stimulating factor (GM-CSF), TPO and Flt3 ligand (Methocult plus) (Fig. 3b) in three independent experiments. Triple mutant HSCs formed more secondary colonies than wild-type HSCs in Methocult plus media, whereas the triple mutant multipotent progenitors formed significantly more colonies than wild-type multipotent progenitors in both Methocult plus and Methocult media (Fig. 3b). This clearly shows that triple mutant multipotent progenitors have an enhanced proliferative capacity, and in the Methocult plus media they proliferate and form secondary colonies better than even wild-type HSCs (Fig. 3b, c).

We next asked whether loss of the *p16^{Ink4a}/p19^{Arf}* and *Trp53* loci conferred long-term reconstituting ability on all proliferating blood cells. To do this, we transplanted three recipient mice with 100 c-kit$^+$Sca-1$^-$Lin$^-$ myeloid progenitors, a more differentiated progenitor population. We could not detect contribution to mature lineages at 4 weeks and 10 weeks, and we stopped analysis at 16 weeks (Fig. 2e). Next, we transplanted eight recipient mice with 500 or 1,000 triple mutant common myeloid progenitors and another eight recipient mice with 500 or 1,000 granulocyte–macrophage progenitors[21]. After 3 weeks, we were unable to detect mature myeloid cells that were descended from the transplanted common myeloid

©2008 Nature Publishing Group

progenitors or granulocyte–macrophage progenitors (Supplementary Table 3). These results suggest that as progenitor cells mature the constraints on self-renewal increase (Supplementary Fig. 1).

The cellular hierarchy seen in many tissues with self-renewing stem cells and short-lived progenitors probably evolved to limit the number of cells that can accumulate oncogenic mutations. Our data demonstrates that loss of both *p16^{Ink4a}/p19^{Arf} and Trp53* loci results in substantial expansion of self-renewing cells. Our results suggest that triple mutant multipotent progenitors have acquired long-term reconstituting ability, implying that the compound deletion of *p16^{Ink4}*, *p19^{Arf}* and *Trp53* removes constraints limiting self-renewal in early progenitors[2]. However, triple mutant common myeloid progenitors and granulocyte–macrophage progenitors, more differentiated progenitor cells, still have limited lifespans (Supplementary Fig. 1), suggesting that there must be other genetic pathways that also prevent self-renewal in mature progenitors. This further supports the notion that there are multiple genetic determinants that regulate the frequency of self-renewing haematopoietic cells *in vivo*[3].

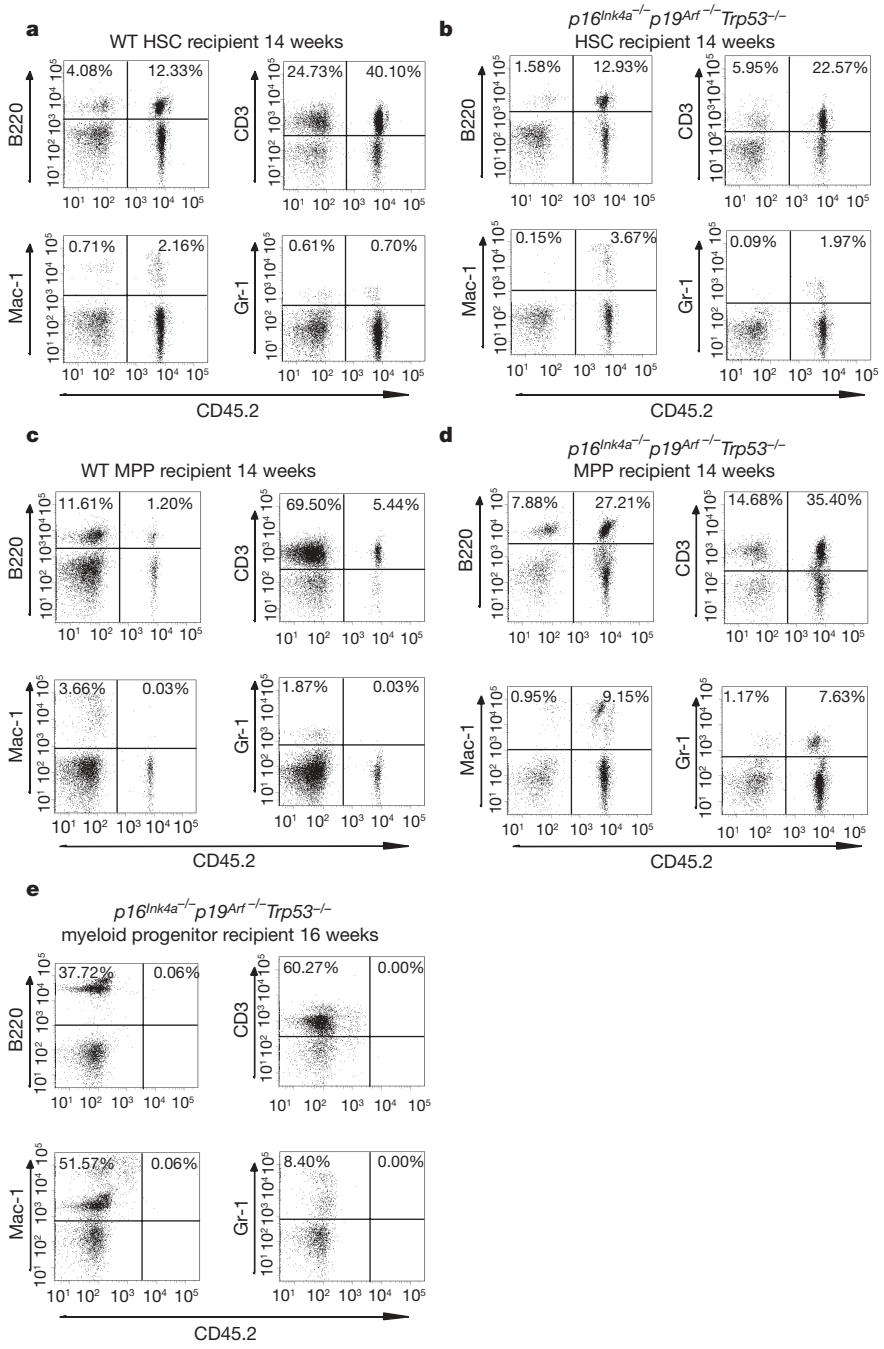


Figure 2 | Long-term multi-lineage reconstitution by HSCs and multipotent progenitors from *p16$^{Ink4a-/-}$p19$^{Arf-/-}$Trp53$^{-/-}$* mice. **a**, Long-term reconstitution by double-sorted HSCs (CD150$^+$Sca-1$^+$c-kit$^+$CD48$^-$Lin$^-$) from wild-type mice. **b**, Transplantation of HSCs from *p16$^{Ink4a-/-}$p19$^{Arf-/-}$Trp53$^{-/-}$* mice results in long-term multi-lineage reconstitution of recipient mice when analysed 14–20 weeks after transplantation (four of four mice). **c**, Double-sorted multipotent progenitors (Sca-1$^+$c-kit$^+$CD150$^-$CD48$^-$Lin$^-$) from wild-type mice are only capable of contributing to long-lived lymphocyte B220$^+$ and CD3$^+$ populations and not to short-lived myeloid Mac-1$^+$ and Gr-1$^+$ populations. **d**, Transplanted double-sorted multipotent progenitors from triple mutant mice also resulted in long-term multi-lineage reconstitution of recipient mice when analysed 14–20 weeks after transplantation (four of four mice). **e**, Double-sorted c-kit$^+$Sca-1$^-$Lin$^-$ myeloid progenitors do not contribute to short-lived myeloid Mac-1$^+$ and Gr-1$^+$ populations.

©2008 Nature Publishing Group

This expansion of long-lived cells potentially reveals a cellular mechanism by which these genes contribute to oncogenesis and may explain why they, or other components of their regulatory pathways, are so commonly mutated or dysregulated in multiple tumour types[22–26]. The regulation of self-renewal in wild-type HSCs by epigenetically repressing the $p16^{Ink4a}/p19^{Arf}$ locus and *Trp53* pathway, in contrast to that of triple mutant multipotent progenitors achieved by genetic deficiency of the $p16^{Ink4a}/p19^{Arf}$ locus and *Trp53*, hints at differences in the regulation of self-renewal in normal cells and self-renewing cancer cells that harbour mutations in these genes. Such differences could be exploited as therapeutic targets.

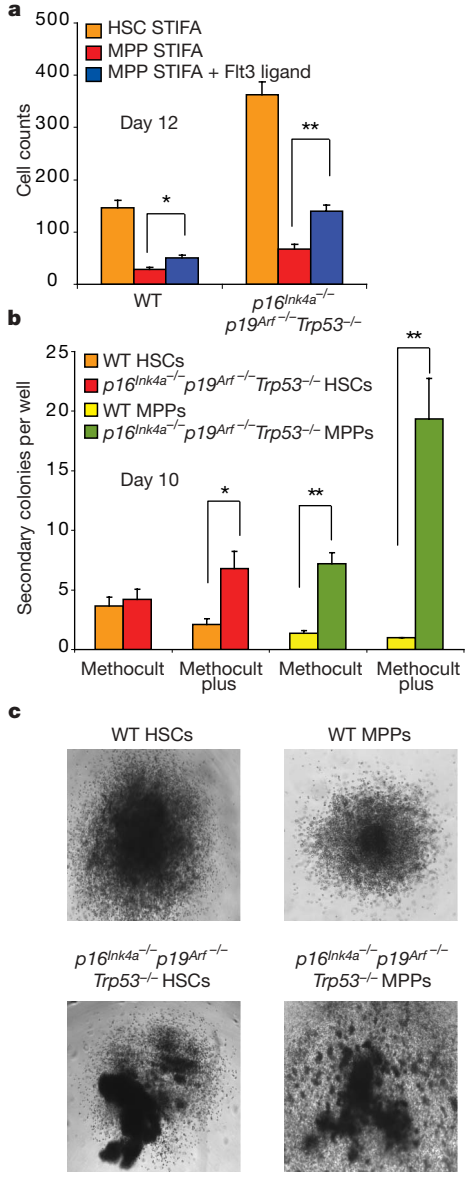


Figure 3 | *In vitro* proliferation and colony formation of wild-type and triple mutant HSCs and multipotent progenitors. a, Proliferation of wild-type and triple mutant HSCs and multipotent progenitors in expansion media. Wild-type and triple mutant HSCs are responsive to cytokines that promote their expansion and proliferate considerably better than wild-type and triple mutant multipotent progenitors in serum-free StemSpan media containing SCF, TPO, IGF-2, FGF-1 and Angptl3 (STIFA) media. Wild-type and triple mutant multipotent progenitors show a doubling in their proliferation rates with the addition of Flt3 ligand to the STIFA media. Five cells were double-sorted into U-bottom 96-well plates. Replicates of at least 21 wells for each experimental group were counted. **b**, Triple mutant multipotent progenitors show an increase in the number of secondary colonies that they form compared to wild-type multipotent progenitors in Methocult and Methocult plus media, demonstrating their proliferative advantage. Triple mutant multipotent progenitors are also very responsive to the additional cytokines IL-11, GM-CSF, Flt3 ligand and TPO added to the Methocult media and can form secondary colonies even better than wild-type HSCs. Single cells were double-sorted into U-bottom 96-well plates for the colony formation experiments. Replicates of at least 20 wells were scored in the methylcellulose colony formation experiments. For **a** and **b**: asterisk, $P < 0.005$; double asterisk, $P < 0.0001$ (Student's *t*-test). Error bars denote s.e.m. **c**, Representative images of wild-type and triple mutant colonies from single double-sorted HSCs and multipotent progenitors taken at ×25 magnification.

METHODS SUMMARY

Flow cytometry. Isolation and analysis of bone marrow, thymus, spleen and peripheral blood have been previously described[4,6,21] and additional details can be found in Methods.

Cell proliferation, apoptosis and colony formation. Proliferation of HSCs was assayed via *in vitro* culture in STIFA expansion media[20] and in the case of multipotent progenitors in STIFA media with or without Flt3 ligand. Apoptosis during *in vitro* expansion of HSCs and multipotent progenitors was assayed by annexin V and 4,6-diamidino-2-phenylindole (DAPI) staining. Colony formation was performed as previously described[27].

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

Received 20 December 2007; accepted 25 February 2008.
Published online 16 April 2008.

1. Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1**, 661–673 (1994).
2. Clarke, M. F. & Fuller, M. Stem cells and cancer: two faces of eve. *Cell* **124**, 1111–1115 (2006).
3. Morrison, S. J. *et al.* A genetic determinant that specifically regulates the frequency of hematopoietic stem cells. *J. Immunol.* **168**, 635–642 (2002).
4. Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109–1121 (2005).
5. Adolfsson, J. *et al.* Upregulation of Flt3 expression within the bone marrow $Lin^{-}Sca1^{+}c\text{-}kit^{+}$ stem cell compartment is accompanied by loss of self-renewal capacity. *Immunity* **15**, 659–669 (2001).
6. Christensen, J. L. & Weissman, I. L. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc. Natl Acad. Sci. USA* **98**, 14541–14546 (2001).
7. Molofsky, A. V. *et al.* Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. *Nature* **425**, 962–967 (2003).
8. Park, I.-K. *et al.* Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. *Nature* **423**, 302–305 (2003).
9. Lessard, J. & Sauvageau, G. Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells. *Nature* **423**, 255–260 (2003).
10. Jacobs, J. J., Kieboom, K., Marino, S., DePinho, R. A. & van Lohuizen, M. The oncogene and Polycomb-group gene *bmi-1* regulates cell proliferation and senescence through the *ink4a* locus. *Nature* **397**, 164–168 (1999).
11. Molofsky, A. V., He, S., Bydon, M., Morrison, S. J. & Pardal, R. Bmi-1 promotes neural stem cell self-renewal and neural development but not mouse growth and survival by repressing the $p16^{Ink4a}$ and $p19^{Arf}$ senescence pathways. *Genes Dev.* **19**, 1432–1437 (2005).
12. Oguro, H. *et al.* Differential impact of Ink4a and Arf on hematopoietic stem cells and their bone marrow microenvironment in Bmi1-deficient mice. *J. Exp. Med.* **203**, 2247–2253 (2006).
13. Janzen, V. *et al.* Stem-cell ageing modified by the cyclin-dependent kinase inhibitor $p16^{INK4a}$. *Nature* **443**, 421–426 (2006).
14. Stepanova, L. & Sorrentino, B. P. A limited role for $p16^{Ink4a}$ and $p19^{Arf}$ in the loss of hematopoietic stem cells during proliferative stress. *Blood* **106**, 827–832 (2005).
15. Dumble, M. *et al.* The impact of altered p53 dosage on hematopoietic stem cell dynamics during aging. *Blood* **109**, 1736–1742 (2007).
16. Kim, I., He, S., Yilmaz, O. H., Kiel, M. J. & Morrison, S. J. Enhanced purification of fetal liver hematopoietic stem cells using SLAM family receptors. *Blood* **108**, 737–744 (2006).
17. Yilmaz, O. H., Kiel, M. J. & Morrison, S. J. SLAM family markers are conserved among hematopoietic stem cells from old and reconstituted mice and markedly increase their purity. *Blood* **107**, 924–930 (2006).
18. Zhang, C. C. & Lodish, H. F. Insulin-like growth factor 2 expressed in a novel fetal liver cell population is a growth factor for hematopoietic stem cells. *Blood* **103**, 2513–2521 (2004).
19. Zhang, C. C. & Lodish, H. F. Murine hematopoietic stem cells change their surface phenotype during *ex vivo* expansion. *Blood* **105**, 4314–4320 (2005).
20. Zhang, C. C. *et al.* Angiopoietin-like proteins stimulate *ex vivo* expansion of hematopoietic stem cells. *Nature Med.* **12**, 240–245 (2006).
21. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–195 (2000).

©2008 Nature Publishing Group

22. Berggren, P. *et al.* Detecting homozygous deletions in the *CDKN2A(p16^{INK4a})/* ARF(*p14^{ARF}*) gene in urinary bladder cancer using real-time quantitative PCR. *Clin. Cancer Res.* **9,** 235–242 (2003).
23. Esteller, M. *et al.* Hypermethylation-associated inactivation of p14^{ARF} is independent of p16^{INK4a} methylation and p53 mutational status. *Cancer Res.* **60,** 129–133 (2000).
24. Weber, A., Bellmann, U., Bootz, F., Wittekind, C. & Tannapfel, A. INK4a-ARF alterations and p53 mutations in primary and consecutive squamous cell carcinoma of the head and neck. *Virchows Arch.* **441,** 133–142 (2002).
25. Burke, L. *et al.* Prognostic implications of molecular and immunohistochemical profiles of the Rb and p53 cell cycle regulatory pathways in primary non-small cell lung carcinoma. *Clin. Cancer Res.* **11,** 232–241 (2005).
26. Zolota, V., Sirinian, C., Melachrinou, M., Symeonidis, A. & Bonikos, D. S. Expression of the regulatory cell cycle proteins p21, p27, p14, p16, p53, mdm2, and cyclin E in bone marrow biopsies with acute myeloid leukemia. Correlation with patients' survival. *Pathol. Res. Pract.* **203,** 199–207 (2007).
27. Jamieson, C. H. *et al.* Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. *N. Engl. J. Med.* **351,** 657–667 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature. A summary figure is also included.

Acknowledgements This work was supported by grants from the National Cancer Institute and National Institute of Diabetes and Digestive and Kidney Diseases (to M.F.C).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.F.C. (mfclarke@stanford.edu).

©2008 Nature Publishing Group

METHODS

Mice. *Trp53*$^{-/-}$ mice (B6.129S2-*Trp53*^{tm1Tyj}) were purchased from the Jackson Laboratory, *p16*$^{Ink4a-/-}$ mice (FVB/N.129-*Cdkn2a*^{tm2Rdp}) and *p16*$^{Ink4a-/-}$ *p19*$^{Arf-/-}$ mice (B6.129-*Cdkn2a*^{tm1Rdp}) were obtained from Mouse Models of Human Cancers Consortium (NCI-Frederick). *p19*$^{Arf-/-}$ mice were obtained from C. Sherr. *Bmi1*$^{+/-}$ mice were obtained from T. Magnuson. All mice were bred with BA mice (C57Bl/Ka-CD45.2/Thy1.1) at least five times, and verified for Thy1.1, H-2^{b} and CD45.2 homozygosity by peripheral blood analysis using flow cytometry. Recipient mice in transplant assays were adult C57Bl/Ka-CD45.1/Thy1.2 mice. All mice used for this study were maintained at the University of Michigan Animal Facility or Stanford Animal Facility in accordance with the guidelines of both Institutional Animal Care Use Committees.

Flow cytometry. Antibodies used for sorting bone marrow cells and analysis were lineage markers (CD3, CD5, CD8, Gr-1, B220 and Ter119), Sca-1, c-kit, CD135 (Flt3), FcγR (CD16/CD32), CD34, IL-7R, CD150 and CD48. Cells were analysed or sorted using a Vantage fluorescence-activated cell sorter (BD Biosciences) or FACSAria cell sorter (BD Biosciences). For analysis of wild-type and mutant mice in a Bmi1-deficient background we used Lin^{-}CD135^{-} Thy-1.1^{lo}Sca-1^{+}c-kit^{+} HSCs[6]. For analysis and sorting of wild-type and triple mutant HSCs and multipotent progenitors we used CD150^{+}CD48^{-} Sca-1^{+}Lin^{-}c-kit^{+} HSCs and CD150^{-}CD48^{-}Sca-1^{+}Lin^{-}c-kit^{+} multipotent progenitors as previously described[4,16,17]. For sorting of wild-type and triple mutant myeloid progenitors we used c-kit^{+}Sca-1^{-}Lin^{-} myeloid progenitors, whereas for further fractionation we used CD34^{+}FcγR^{lo}IL-7R^{-}Sca-1^{-}Lin^{-} c-kit^{+} common myeloid progenitors and CD34^{+}FcγR^{hi}IL-7R^{-}Sca-1^{-}Lin^{-} c-kit^{+} granulocyte–macrophage progenitors[21]. For lineage analysis, both bone marrow cells and splenocytes were stained with antibodies against Gr-1, Mac-1, CD3, B220 and Ter119, and thymocytes with antibodies against CD3, CD4 and CD8. For peripheral blood analysis, red blood cells were lysed with hypotonic buffer, and nucleated cells were stained with antibodies against CD45.2, Gr-1, Mac-1, CD3 and B220. Antibodies were directly conjugated or biotinylated and purchased from e-Bioscience, BD Biosciences, or Biolegend.

Long-term competitive reconstitution. C57Bl/Ka-CD45.1/Thy1.2 congenic mice were lethally irradiated (1,140 rad) at a dose rate of approximately 3 Gy min^{-1}, delivered in two doses 4 h apart. The next day, mice were competitively reconstituted by retro-orbital venous sinus injection of whole bone marrow cells from donor mice mixed with a radioprotective dose of 2×10^5 bone marrow cells from unirradiated C57Bl/Ka-CD45.1/Thy1.2 mice. Peripheral blood was drawn monthly up to 20 weeks to monitor reconstitution by donor type (CD45.2) myeloid and lymphoid cells, as described above. Mice that had more than 1% donor-derived (Ly5.1^{+}) cells in both lymphoid (CD3^{+} and B220^{+}) and myeloid (Gr-1^{+} and Mac-1^{+}) subpopulations were considered to be repopulated by donor cells. The secondary bone marrow transplant was performed using 1×10^6 or 2×10^6 whole bone marrow cells. Frequency of long-term reconstituting cells from limiting dilution experiments was calculated using L-Calc software (StemCell Technologies).

***In vitro* proliferation, colony formation and annexin V analysis.** We double-sorted 1 or 5 wild-type or triple mutant HSCs or multipotent progenitors into U-bottom 96-well plates containing 200 µl of StemSpan serum-free media (StemCell technologies) supplemented with 10 µg ml^{-1} of heparin (Sigma), 10 ng ml^{-1} mouse SCF, 20 ng ml^{-1} mouse TPO, 20 ng ml^{-1} mouse IGF-2, 100 ng ml^{-1} mouse Angptl3 (all from R&D) and 10 ng ml^{-1} human FGF-1 as previously described[18–20]. In the case of both wild-type and triple mutant multipotent progenitors we also clone-sorted 1 or 5 cells into the StemSpan serum-free media supplemented as above with the addition of 30 ng ml^{-1} mouse Flt3 ligand (R&D). The cell counts were performed on day 12.

For colony formation we double-sorted single wild-type or triple mutant HSCs or multipotent progenitors into U-bottom 96-well plates containing 100 µl of Methocult GF M3434 media (StemCell Technologies) or 100 µl of Methocult GF M3434 media supplemented with 10 ng ml^{-1} mouse IL-11, 50 ng ml^{-1} mouse GM-CSF, 50 ng ml^{-1} mouse TPO and 10 ng ml^{-1} mouse Flt3 ligand (all from R&D) as previously described[27]. Colonies were scored on day 10 of culture. Pictures of colonies were taken with a Leica DMI 6000B microscope using Image Pro-plus software version 5.1.

For annexin V analysis of apoptotic cells we sorted 1,000 wild-type or triple mutant HSCs into StemSpan serum-free media supplemented with the cytokines SCF, TPO, IGF-2, FGF-1 and Angptl3, or 1,000 wild-type or triple mutant multipotent progenitors sorted into StemSpan serum-free media supplemented with the cytokines SCF, TPO, IGF-2, FGF-1, Angptl3 and Flt3 ligand. The cells were cultured for 12 days and stained with annexin V-Fitc (BD Biosciences) and the viability marker DAPI (Molecular Probes).

©2008 Nature Publishing Group

LETTERS

Discrete fixed-resolution representations in visual working memory

Weiwei Zhang[1,2] & Steven J. Luck[2]

Limits on the storage capacity of working memory significantly affect cognitive abilities in a wide range of domains[1], but the nature of these capacity limits has been elusive[2]. Some researchers have proposed that working memory stores a limited set of discrete, fixed-resolution representations[3], whereas others have proposed that working memory consists of a pool of resources that can be allocated flexibly to provide either a small number of high-resolution representations or a large number of low-resolution representations[4]. Here we resolve this controversy by providing independent measures of capacity and resolution. We show that, when presented with more than a few simple objects, human observers store a high-resolution representation of a subset of the objects and retain no information about the others. Memory resolution varied over a narrow range that cannot be explained in terms of a general resource pool but can be well explained by a small set of discrete, fixed-resolution representations.

To separately measure the number of items stored in working memory and the precision of each representation, we used a short-term recall paradigm[5,6] in which subjects report the remembered colour of a probed item by clicking on a colour wheel (Fig. 1a). If the probed item has been stored in working memory, the recalled value will tend to be near the original colour. If the probed item has not been stored, then the observer will have no information about the colour, and the responses should be random. These two types of trials are mixed together in the data (Fig. 1b), but the components can be recovered via standard estimation methods. This produces one parameter (P_m) representing the probability that the probed item was present in memory at the time of the probe and another parameter (s.d.) representing the precision of the representation when the cued item was present in memory.

Experiment 1 ($N = 8$) tested this model using set sizes of 3 or 6 coloured squares (Fig. 1c). s.d. did not vary significantly across set sizes ($F < 1$), whereas P_m was approximately twice as great at set size 3 as at set size 6 ($F(1,7) = 761.26$, $P < 0.001$). Our simple fixed-resolution model provided an excellent quantitative fit to the data, whereas a model in which all items are encoded could not fit the data (see Supplementary Notes). This result rules out the entire class of working memory models in which all items are stored but with a resolution or noise level that depends on the number of items in memory[5]. Control experiments demonstrated that these results cannot be explained by a lack of time to encode the items or by a lack of sensitivity, and additional analyses demonstrated that the observers remembered continuous colour values rather than colour categories (see Supplementary Notes).

These results demonstrate that observers store a small number of representations with good precision. However, it is possible that performance is influenced both by a limited number of 'storage slots' and a limited pool of resources[7]. As an analogy, consider three cups (the slots) and a bottle of juice (the resource). It would be impossible to serve juice to more than three people at a time, but it would be possible to pour most of the juice into a single cup, leaving only a few drops for the other two cups. Thus, allocating most of the resources to a single representation could increase the precision of that representation, leaving 'only a few drops' of resources for the other representations, which would then be highly imprecise. We call this the 'slots+resources' model.

The storage of information in visual working memory could instead be an all-or-none process that either creates a representation of a given precision or creates no representation at all. This would be analogous to a limited set of prepackaged juice boxes of a fixed size. The juice boxes are still a type of resource, but one that is highly constrained by the small number and fixed size of each box. That is, if three juice boxes are available, an individual could be given 0, 1, 2 or 3 boxes. Similarly, if three memory slots are available, all three could be used to represent a single object. If each representation stores an independent sample of the stimulus, and observers simply report the average of the three representations at the time of test, this will lead to an increase in the precision of the report. We call this the 'slots+averaging' model. Note that storing a single object in multiple

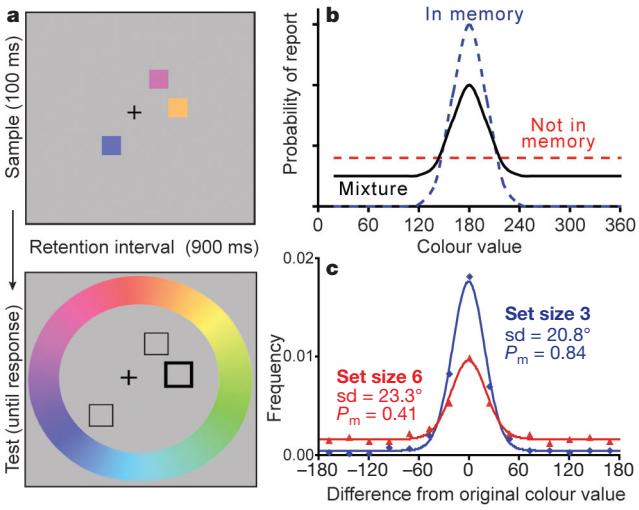


Figure 1 | Experimental approach and results of experiment 1. a, Colour recall task. **b**, Mixture model of performance, showing the probability of reporting each colour value given a sample colour at 180°. When the probed item is present in memory, the reported colour tends to be near the original colour (blue broken line). When the probed item is not present in memory, the observer is equally likely to report any colour value (red broken line). When collapsed across trials, the data comprise a mixture of these two trial types (solid line), weighted by the probability that the probed item was stored in memory. **c**, Results of experiment 1 ($N = 8$). P_m and s.d. are defined in the text.

[1]Department of Psychology, University of Iowa, Iowa City, Iowa 52242, USA. [2]Center for Mind & Brain, University of California, Davis, California 95618, USA.

©2008 Nature Publishing Group

slots would be a rational strategy, and this sort of averaging is common in models of perception[8–10].

For both the slots+resources and slots+averaging models, s.d. will be improved when the set size is reduced below the number of available slots. Moreover, both models predict that this improvement will follow a square root function (see Supplementary Notes). This is exactly what was observed in experiment 2 (Fig. 2), in which observers ($N = 8$) were presented with 1, 2, 3 or 6 objects: s.d. increased as the set size increased from 1 to 3 but then remained constant as the set size increased to 6. In contrast, P_m declined very slowly as set size increased from 1 to 3 and then decreased suddenly at set size 6. This pattern of results can be explained quantitatively by both the slots+resources model (adjusted $r^2 = 0.96$) and the slots+averaging model (adjusted $r^2 = 0.99$) (see Fig. 2 and Supplementary Notes), but it differs significantly from the predictions of a pure resource model ($P < 0.001$, χ^2 test).

The slots+resources and slots+averaging models make different predictions about the range over which precision can vary. Specifically, the slots+resources model posits that the majority of resources can be devoted to one representation (leading to a very small s.d.), leaving 'only a few drops' of resources for other representations (leading to a very large s.d.). In contrast, the slots+averaging model posits that the observed s.d. is never worse than the s.d. of a single slot and is never better than the s.d. for a single slot divided by the square root of the number of slots. To distinguish between these models, experiment 3 ($N = 22$) used a line in the sample array to cue one of four coloured squares (Fig. 3a). The cued square was probed on 70% of the trials, and each uncued square was probed on 10% of trials. Neutral trials were also included, in which all four locations were cued. The cue was simultaneous with the sample array so that it would not influence perceptual processing[11], and the duration of the sample array was increased to 300 ms to provide adequate time for resource allocation[10,12].

The slots+resources model predicts that observers will devote the lion's share of resources to the cued item, leading to a large difference in s.d. between valid, neutral and invalid trials, but only a small difference in P_m. In contrast, the slots+averaging model predicts that observers will devote most of their slots to the cued location, which would lead to a large difference in P_m between valid and invalid trials. This should also lead to a somewhat smaller s.d. on valid trials than on neutral trials because of the benefits of averaging. However, this should lead to no difference in s.d. between neutral and invalid trials, because a given item receives either 0 or 1 slots on both neutral and invalid trials.

We found that P_m was substantially greater on valid trials than on invalid trials ($F(1,21) = 203.87$, $P < 0.001$; Fig. 3a), demonstrating that the observers attempted to maximize performance for the cued item by devoting more slots to it. s.d. was slightly but significantly smaller on valid trials than on neutral trials ($F(1,21) = 13.49$, $P < 0.001$), and the magnitude of this difference was within the small range that can result from averaging slots. In addition, s.d. was virtually identical on neutral and invalid trials ($F < 1$), indicating that the improvement in s.d. on valid trials was not achieved by taking resources away from the uncued items. Thus, despite the fact that the cued item was seven times more likely to be probed than each uncued item, s.d. was only slightly improved for the cued item (compared to the neutral trials) and s.d. was not reduced for the uncued items (compared with the neutral trials). It does not appear to be possible to provide a representation with 'only a few drops' of resources and thereby produce an imprecise representation.

Computational neuroscience theories suggest that an all-or-none, fixed-resolution encoding process may be required to create durable representations that can survive new sensory inputs[13,14]. To assess the encoding process, experiment 4 ($N = 8$) used a masking manipulation that emulates the masking effects of eye movements in natural vision. Specifically, we presented masks at the locations of the coloured squares either 110 or 340 ms after the onset of the squares (Fig. 3b). At these intervals, masks interfere with working memory encoding but not with perceptual processing[15]. If working memory representations gradually become more precise over time, then presenting a mask array at an early time point could potentially reveal the existence of low-precision representations. If, however, the process of creating durable memory representations that can survive new visual inputs involves an all-or-none step, as suggested by studies of the 'attentional blink' phenomenon[16], then the masks will only influence P_m. We observed that decreasing the masking interval produced a large decline in P_m ($F(1,7) = 47.70$, $P < 0.001$) but no change in s.d. ($F < 1$). Thus, the creation of working memory representations that can survive new inputs involves an all-or-none step for simple objects (although it is possible that some gradual accumulation of information occurs before this step and is available in the absence of masking).

To demonstrate that the present results can generalize to other stimulus dimensions, we repeated experiments 2 and 3 with shapes rather than colours. We used shapes defined by Fourier descriptors[17], which vary along continuous quantitative dimensions. The results were largely identical to the results obtained for colour, with approximately the same P_m for these shapes as for the simple colours in experiment 2 (see Supplementary Notes). Most notably, s.d. did not increase as the set size increased from three to six items and was virtually identical for neutral and invalid trials. Because the Fourier descriptor method provides a mathematically[17], perceptually[18,19] and neurally[20] meaningful way to describe shapes of any complexity, this analytic approach could be used to determine whether the present pattern of results would be obtained with more complex objects. Object complexity can have a large impact on performance in change detection tasks[7], but this may reflect greater sample-test similarity for complex objects[21] or the need to store each

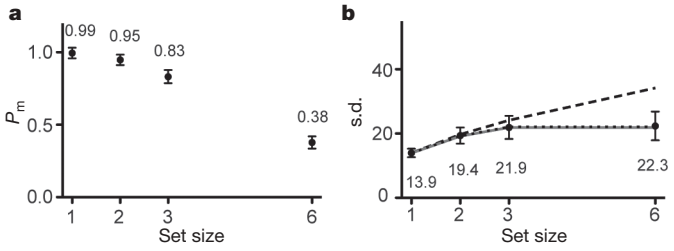


Figure 2 | P_m and s.d. results from experiment 2 ($N = 8$). **a**, P_m; **b**, s.d.; the numbers within the panels provide the means. In **b**, the lines show the s.d. predictions of a pure resource model (black dashed line), the slots+averaging model (grey solid line), and the slots+resources model (black dotted line). Error bars show within-subjects 95% confidence intervals[26].

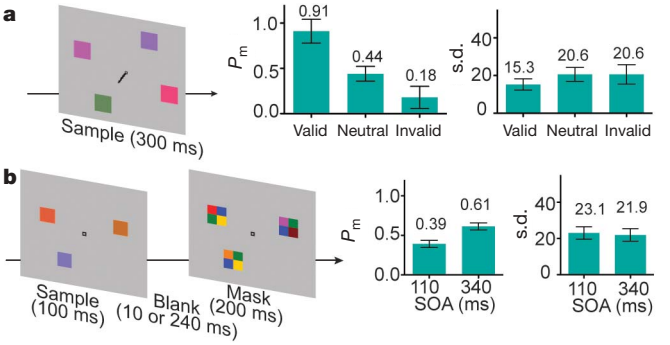


Figure 3 | Stimuli and results from experiments 3 and 4. **a**, Experiment 3 ($N = 22$), which included valid, neutral and invalid trials. The cue appeared simultaneously with the sample array. **b**, Experiment 4 ($N = 8$), in which a mask array followed the sample array with a stimulus onset asynchrony (SOA) of 110 or 340 ms. A colour wheel and probe array appeared at the end of the trial, 900 ms after sample offset. Error bars show within-subjects 95% confidence intervals[26], and numbers within the panels provide the means.

©2008 Nature Publishing Group

part of a multipart object in a separate slot[22,23]. Alternatively, complex objects may require some kind of limited resource that is not needed for the simple objects studied here.

Together, the present experiments resolve an issue that has been debated for decades[4,5,24,25], showing that a model with a small set of discrete, fixed-resolution representations can provide a quantitative account of memory performance across a broad range of experimental manipulations. This model does not completely eliminate the concept of resources, because the slots themselves are a type of resource. However, the slots+averaging model defines exactly what the resource is and describes strict limits on how flexibly this resource can be allocated.

METHODS SUMMARY

The stimuli and task are shown in Fig. 1a. Subjects viewed a sample array and then, following a brief delay, reported the colour of one item from this array (indicated by a thick outlined box) by clicking on a colour wheel. There is good agreement between this procedure and the more commonly used change-detection procedure (see Supplementary Notes).

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

Received 2 November 2007; accepted 14 February 2008.
Published online 2 April 2008.

1. Baddeley, A. D. *Working Memory* (Clarendon, Oxford, 1986).
2. Miyake, A. & Shah, P. (eds) *Models of Working Memory* (Cambridge Univ. Press, Cambridge, UK, 1999).
3. Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390,** 279–281 (1997).
4. Frick, R. W. Issues of representation and limited capacity in the visuospatial sketchpad. *Br. J. Psychol.* **79,** 289–308 (1988).
5. Wilken, P. & Ma, W. J. A detection theory account of change detection. *J. Vis.* **4,** 1120–1135 (2004).
6. Prinzmetal, W., Amiri, H., Allen, K. & Edwards, T. Phenomenology of attention: I. Color, location, orientation, and spatial frequency. *J. Exp. Psychol. Hum. Percept. Perform.* **24,** 261–282 (1998).
7. Alvarez, G. A. & Cavanagh, P. The capacity of visual short-term memory is set both by information load and by number of objects. *Psychol. Sci.* **15,** 106–111 (2004).
8. Palmer, J. Attentional limits on the perception and memory of visual information. *J. Exp. Psychol. Hum. Percept. Perform.* **16,** 332–350 (1990).
9. Newsome, W. T. *et al.* in *The Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 401–414 (MIT Press, Cambridge, Massachusetts, 1995).
10. Bonnel, A. M. & Miller, J. Attentional effects on concurrent psychophysical discriminations: Investigations of a sample-size model. *Percept. Psychophys.* **55,** 162–179 (1994).
11. Vogel, E. K., Woodman, G. F. & Luck, S. J. Pushing around the locus of selection: Evidence for the flexible-selection hypothesis. *J. Cogn. Neurosci.* **17,** 1907–1922 (2005).
12. Luck, S. J., Hillyard, S. A., Mouloua, M. & Hawkins, H. L. Mechanisms of visual-spatial attention: Resource allocation or uncertainty reduction? *J. Exp. Psychol. Hum. Percept. Perform.* **22,** 725–737 (1996).
13. Wang, X.-J. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* **24,** 455–463 (2001).
14. Raffone, A. & Wolters, G. A cortical mechanism for binding in visual working memory. *J. Cogn. Neurosci.* **13,** 766–785 (2001).
15. Vogel, E. K., Woodman, G. F. & Luck, S. J. The time course of consolidation in visual working memory. *J. Exp. Psychol. Hum. Percept. Perform.* **32,** 1436–1451 (2006).
16. Sergent, C. & Dehaene, S. Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychol. Sci.* **15,** 720–728 (2004).
17. Zahn, C. T. & Roskies, R. Z. Fourier descriptors for plane closed curves. *IEEE Trans. Comput. C* **21,** 269–281 (1972).
18. Alter, I. & Schwartz, E. L. Psychophysical studies of shape with Fourier descriptor stimuli. *Perception* **17,** 191–202 (1988).
19. Cortese, J. M. & Dyre, B. P. Perceptual similarity of shapes generated from Fourier descriptors. *J. Exp. Psychol. Hum. Percept. Perform.* **22,** 133–143 (1996).
20. Schwartz, E. L., Desimone, R., Albright, T. D. & Gross, C. G. Shape recognition and inferior temporal neurons. *Proc. Natl Acad. Sci. USA* **80,** 5776–5778 (1983).
21. Awh, E., Barton, B. & Vogel, E. K. Visual working memory represents a fixed number of items regardless of complexity. *Psychol. Sci.* **18,** 622–628 (2007).
22. Xu, Y. Limitations of object-based feature encoding in visual short-term memory. *J. Exp. Psychol. Hum. Percept. Perform.* **28,** 458–468 (2002).
23. Sakai, K. & Inui, T. A feature-segmentation model of short-term visual memory. *Perception* **31,** 579–590 (2002).
24. Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* **63,** 81–97 (1956).
25. Simon, H. A. How big is a chunk? *Science* **183,** 482–488 (1974).
26. Cousineau, D. Confidence intervals in within-subjects designs: A simpler solution to Loftus and Masson's method. *Tutor. Quant. Methods Psychol.* **1,** 42–45 (2007).
27. Myung, I. J. Tutorial on maximum likelihood estimation. *J. Math. Psychol.* **47,** 90–100 (2003).
28. Fisher, N. I. *Statistical Analysis of Circular Data* (Cambridge Univ. Press, New York, 1993).

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

Acknowledgements This work was made possible by grants from NIMH to S.J.L.

Author Contributions W.Z. and S.J.L. together developed the general conceptualization of this study and wrote the paper. W.Z. designed and implemented the experimental and analytical approach.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to W.Z. (wwzhang@ucdavis.edu).

©2008 Nature Publishing Group

METHODS

Subjects. Eight subjects between 18 and 35 years old participated in each colour memory experiment except the cuing experiment, in which 22 subjects participated owing to the low probability of the invalid trials. Subjects provided informed consent and received course credit or monetary compensation. All reported having normal colour vision and normal or corrected-to-normal visual acuity.

Experiments. The stimuli in all experiments were presented on a CRT monitor with a grey background ($15.5\,cd\,m^{-2}$) at a viewing distance of 57 cm. The monitor was calibrated with a Tektronix J17 LumaColour colorimeter. Each coloured square in the sample array subtended $2 \times 2^{\circ}$ of visual angle. Each square was centred on an invisible circle with a radius of 4.5°. The positions were randomly chosen from a set of eight locations equally spaced along the circle. The colour wheel was 2.2° thick and was centred on the monitor with a radius of 8.2°. It consisted of 180 colour values that were evenly distributed along a circle in the CIE $L^*a^*b^*$ colour space. This circle was centred in the colour space at ($L = 70$, $a = 20$, $b = 38$) with a radius of 60. Its centre was chosen to maximize its radius and therefore the discriminability of the colours. All colours had equal luminance and varied mainly in hue and slightly in saturation. The sample array colours were randomly selected from this set of colours. The colour wheel was presented at a random rotation on each trial to minimize contributions from spatial memory.

In the basic paradigm (experiments 1 and 2), each trial consisted of a 100-ms sample array followed by a 900-ms blank delay period and then a probe display that remained present until a response was made (Fig. 1a). The probe display contained the colour wheel and an outlined square at the location of each item from the sample array. One of these squares was thicker (0.20°) than the others (0.04°), which cued the subject to recall the colour of the corresponding item from the sample array by clicking the appropriate colour on the colour wheel with the computer mouse. Accuracy was stressed, and the responses were not timed. Except as noted below, 150 trials were tested in each experimental condition (for example, each set size). The different trial types (for example, different set sizes) were presented in an unpredictable order in each experiment.

In the control experiment that involved varying the level of perceptual noise (Supplementary Fig. 1), the set size was held constant at three items and the duration of the sample array was reduced to 30 ms to ensure that the masks would be effective. Each coloured square in the sample array was covered with a set of either 75 or 150 simultaneously presented coloured dots, randomly distributed over a circular region with a diameter of 4.4° that was centred on the coloured square. Each dot subtended $0.2 \times 0.2^{\circ}$ of visual angle and was drawn in a colour that was randomly sampled from the set of 180 colour values used for the coloured squares.

In the cuing experiment (Fig. 3a), the sample display contained a 1.6° cue line extending from fixation towards one of the four coloured squares (on valid and invalid trials) or four lines extending towards all four squares (on neutral trials). The duration of the sample display was increased to 300 ms in this experiment to provide the observers sufficient time to shift attention to the cued item; the interval between sample onset and probe onset remained 900 ms. Each observer received 350 valid trials, 150 invalid trials (50 per uncued location) and 150 neutral trials. These trial types were randomly intermixed.

In the backward masking experiment (Fig. 3b), the sample array always contained three items. An array of masks was presented 110 or 340 ms after the onset of the sample array, with a duration of 200 ms. The interval between sample offset and probe onset remained constant at 900 ms. Each mask consisted of a 2×2 arrangement of coloured squares, each of which measured $0.55 \times 0.55^{\circ}$. Each mask was centred at the location of one of the three items in the sample display.

Data analysis. The data from a given observer in the colour experiments consisted of a set of distances between the reported colour value and the original colour value in each condition, which reflects the degree of error in the reported colour. Histograms of these error values were used to visualize the distribution of responses (as in Fig. 1c). Maximum likelihood estimation[27] was used to decompose the data from each subject in each condition into three parameters that represent a mixture of a uniform distribution of errors (for trials on which the probed item was not encoded in memory) and a von Mises distribution of errors (for trials on which the probed item was encoded). The von Mises distribution is the circular analogue of the gaussian distribution and was used because the tested colour space was circular[28]. The uniform distribution was represented by a single parameter, P_m, which is the probability that the probed item was present in memory at the time of the probe (which is inversely related to the height of the uniform distribution). The von Mises distribution was represented by two parameters, its mean (μ) and its standard deviation (s.d.). μ reflects any systematic shift of the distribution away from the original colour value. No systematic shifts were expected or observed in any of the present experiments, so this parameter will not be considered further. s.d. reflects the width of the distribution of errors on trials when the probed item was encoded in memory, which in turn reflects the precision or resolution of the memory representation.

The slots+averaging model was fitted to the estimated P_m and s.d. parameters in the experiment in which set sizes 1, 2, 3 and 6 were tested. We computed the total number of slots (K_i) by multiplying P_m by the set size (using the data from set size 3). We then assumed that the slots were randomly distributed among the available items in the sample array, allowing multiple slots to be assigned to a given object if the set size was lower than the number of slots. The s.d. at set size 3 was used to estimate the precision of a single slot. The s.d. from a set of N samples is equal to the s.d. from a single sample divided by the square root of N (see Supplementary Notes). Thus, by knowing the s.d. of a single slot and the average number of slots assigned to the probed item in a given condition, it is possible to predict the s.d. for that condition.

In the slots+resources model, the s.d. at set size 1 was used to estimate the maximum precision when all resources are devoted to a single object. For modelling the data from larger set sizes, the s.d. simply increases as a function of the square root of the number of objects being represented, up to the number of slots (which is estimated as in the slots+averaging model). For simple manipulations of set size, the predictions of the slots+resources model are equivalent to those of the slots+averaging model except that the s.d. values are estimated on the basis of the data at set size 1 rather than the data at set size 3.

©2008 Nature Publishing Group

LETTERS

TGF-β-induced Foxp3 inhibits T_H17 cell differentiation by antagonizing RORγt function

Liang Zhou[1], Jared E. Lopes[3,4], Mark M. W. Chong[1], Ivaylo I. Ivanov[1], Roy Min[1,2], Gabriel D. Victora[1], Yuelei Shen[1], Jianguang Du[3,4], Yuri P. Rubtsov[5], Alexander Y. Rudensky[5], Steven F. Ziegler[3,4] & Dan R. Littman[1,2]

T helper cells that produce IL-17 (T_H17 cells) promote autoimmunity in mice and have been implicated in the pathogenesis of human inflammatory diseases. At mucosal surfaces, T_H17 cells are thought to protect the host from infection, whereas regulatory T (T_{reg}) cells control immune responses and inflammation triggered by the resident microflora[1–5]. Differentiation of both cell types requires transforming growth factor-β (TGF-β), but depends on distinct transcription factors: RORγt (encoded by *Rorc(γt)*) for T_H17 cells and Foxp3 for T_{reg} cells[6–8]. How TGF-β regulates the differentiation of T cells with opposing activities has been perplexing. Here we demonstrate that, together with pro-inflammatory cytokines, TGF-β orchestrates T_H17 cell differentiation in a concentration-dependent manner. At low concentrations, TGF-β synergizes with interleukin (IL)-6 and IL-21 (refs 9–11) to promote IL-23 receptor (*Il23r*) expression, favouring T_H17 cell differentiation. High concentrations of TGF-β repress *IL23r* expression and favour Foxp3$^+$ T_{reg} cells. RORγt and Foxp3 are co-expressed in naive CD4$^+$ T cells exposed to TGF-β and in a subset of T cells in the small intestinal lamina propria of the mouse. *In vitro*, TGF-β-induced Foxp3 inhibits RORγt function, at least in part through their interaction. Accordingly, lamina propria T cells that co-express both transcription factors produce less IL-17 (also known as IL-17a) than those that express RORγt alone. IL-6, IL-21 and IL-23 relieve Foxp3-mediated inhibition of RORγt, thereby promoting T_H17 cell differentiation. Therefore, the decision of antigen-stimulated cells to differentiate into either T_H17 or T_{reg} cells depends on the cytokine-regulated balance of RORγt and Foxp3.

When T lymphocytes are exposed to microbial antigens, they acquire diverse effector functions depending on which cytokines are produced by activated cells of the innate immune system[12]. Differentiation of pro-inflammatory T_H17 cells requires the presence of IL-23, which is produced by activated dendritic cells[13–15]. *In vitro*, however, T_H17 cell differentiation is independent of IL-23 and is induced by TGF-β plus IL-6 or IL-21 (refs 6, 9–11). Both *in vitro* and *in vivo* differentiation of the T_H17 cell lineage require the upregulation of the orphan nuclear receptor RORγt[7]. TGF-β is also required to restrain inflammatory autoimmune responses[16]. Among its numerous properties is its ability to induce expression of Foxp3 in naive antigen-stimulated T cells, endowing the cells with regulatory or suppressor function[8]. Thus, TGF-β can induce both regulatory and pro-inflammatory T cells, depending on whether pro-inflammatory

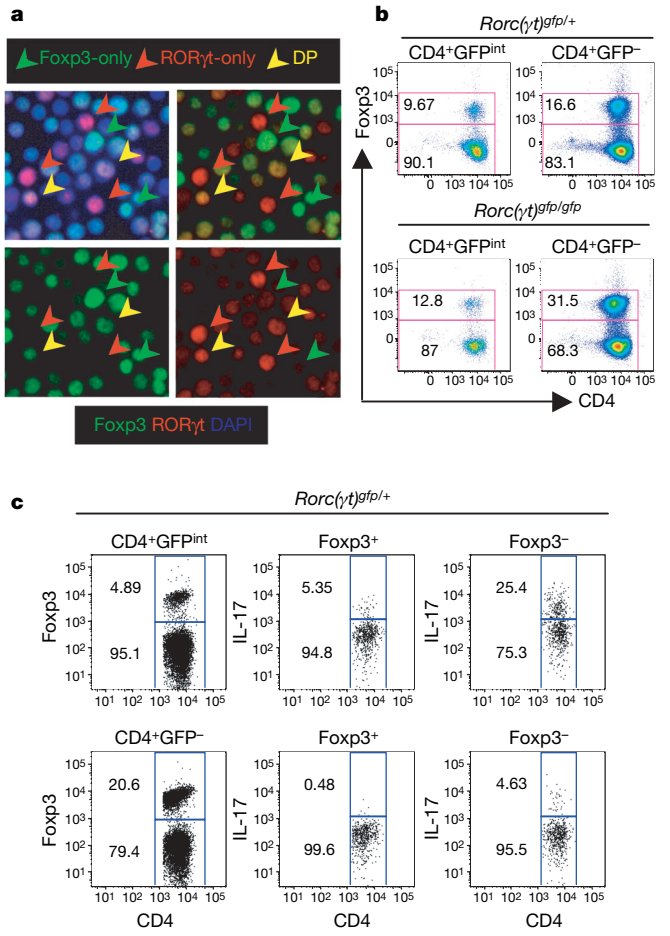


Figure 1 | Co-expression of Foxp3 and RORγt *in vitro* and *in vivo*. **a**, Naive CD4$^+$ T cells were stimulated with anti-T-cell receptor (TCR) and 5 ng ml^{-1} TGF-β for 48 h, and were stained with 4,6-diamidino-2-phenylindole (DAPI; blue nuclear stain; UV channel), anti-RORγ (red; Cy3 channel) and anti-Foxp3 (green; Cy5 channel) monoclonal antibodies. All panels are of the same section. The bottom-left image shows Foxp3-only, bottom-right shows RORγ-only, top-right shows overlay of Foxp3 and RORγ channels, and top-left shows overlay of all three channels. Foxp3, RORγt and double-expressing (DP) cells are indicated with coloured arrows. **b**, Analysis of Foxp3$^+$RORγt$^+$ cells from the small intestinal lamina propria. CD4$^+$GFP^{int} and CD4$^+$GFP$^-$ cells were sorted from lamina propria of *Rorc(γt)*$^{+/gfp}$ and *Rorc(γt)*$^{gfp/gfp}$ mice, and Foxp3 expression was examined by intracellular staining. Results are representative of three experiments. The numbers indicate the percentage of total cells in each gate. **c**, Expression of IL-17 in Foxp3$^+$RORγt$^+$ and Foxp3$^-$RORγt$^+$ T cells from small intestine. Foxp3 and IL-17 expression was examined by intracellular staining of sorted TCRβ$^+$CD4$^+$GFP^{int} cells from the lamina propria of *Rorc(γt)*$^{+/gfp}$ mice.

[1]The Kimmel Center for Biology and Medicine of the Skirball Institute, and [2]Howard Hughes Medical Institute, Departments of Microbiology and Pathology, New York University School of Medicine, New York, New York 10016, USA. [3]Immunology Program Benaroya Research Institute Seattle, Washington 98101, USA. [4]Department of Immunology University of Washington School of Medicine Seattle, and [5]Howard Hughes Medical Institute, Department of Immunology, University of Washington Seattle, Washington 98195, USA.

©2008 Nature Publishing Group

cytokines such as IL-6 and, potentially, IL-23 are present[11,17]. Treatment of antigen-receptor-stimulated T cells with TGF-β alone induces expression of both Foxp3 and RORγt, but not of IL-17 (refs 7, 11). After such treatment, a significant proportion of cells co-expressed the two transcription factors (Fig. 1a and Supplementary Fig. 1a). To determine whether co-expression also occurs *in vivo*, we examined $CD4^+$ T cells from the small intestinal lamina propria of heterozygous RORγt–GFP (green fluorescent protein) knock-in mice, in which IL-17 is produced by TCR^+GFP^{int} lymphocytes[7]. Foxp3 was expressed in about 10% of sorted GFP^{int} ($ROR\gamma t^+$) cells (Fig. 1b and Supplementary Fig. 1b). In addition, Foxp3 was expressed in approximately 17–20% of GFP^- lamina propria $CD4^+$ T cells, consistent with the relatively large proportion of T_{reg} cells in the intestine.

We next performed a fate-mapping analysis to determine the proportion of $IL\text{-}17^+$ small intestinal T cells that had expressed Foxp3 during their ontogeny. Mice expressing Cre recombinase under the regulation of the *Foxp3* locus (Y.P.R. *et al.*, submitted) were crossed with Rosa26-stop–YFP reporter mice[18], and female progeny ($Rosa26^{stop-YFP/+}$; $Foxp3^{cre/+}$) were analysed for expression of yellow fluorescent protein (YFP). When inactivation of X-linked *Foxp3* was taken into account, we found that approximately 15% of $IL\text{-}17^-$ cells and 25% of $IL\text{-}17^+$ cells had expressed Cre at some stage of development (Supplementary Fig. 2). The former represent $Foxp3^+$ T_{reg} cells, whereas the latter are the minimal proportion of T_H17 cells that had expressed Foxp3 at some stage of their differentiation. These data suggest that $Foxp3^+$ T cells can differentiate into T_H17 cells *in vivo* in the presence of pro-inflammatory cytokines.

Examination of IL-17 expression in heterozygous RORγt–GFP knock-in mice revealed that $ROR\gamma t^+Foxp3^+$ lamina propria T cells produced much less IL-17 than $ROR\gamma t^+Foxp3^-$ cells, suggesting that Foxp3 may interfere with the ability of RORγt to induce IL-17 (Fig. 1c). This is consistent with findings showing a >1,000-fold increase in *Il17* messenger RNA, but little change in *Rorc(γt)*, in T_{reg} lineage cells that differentiate in the absence of Foxp3 (ref. 19). To investigate how Foxp3 may influence T_H17 cell differentiation, we asked whether its induction would influence the expression of IL-17 in TGF-β-stimulated T cells. In naive T cells that had been transduced with a retroviral vector encoding RORγt, we found that, whereas IL-6 augmented the proportion of RORγt-IRES-GFP^+ cells that expressed IL-17, TGF-β had a profound inhibitory effect even when added one day after transduction (Fig. 2a, b). Addition of TGF-β was followed by a sharp increase in expression of Foxp3 in the $CD4^+$ T cells, and both the level of *Foxp3* mRNA and proportion of $Foxp3^+$ cells were not affected by the expression of RORγt (Fig. 2c).

To determine whether the inhibitory effect of TGF-β on RORγt is mediated by Foxp3, we knocked down expression of Foxp3 by using a short hairpin RNA (shRNA) vector. TGF-β-induced Foxp3 expression was reduced by the *Foxp3*-specific shRNA vector, but not by control hairpin vectors (Fig. 2c). Accordingly, TGF-β-mediated inhibition of RORγt-directed IL-17 expression was partially reversed by Foxp3 knockdown (Fig. 2d). Consistent with the idea that this inhibition was mediated by Foxp3 upregulation, the most pronounced rescue of IL-17 expression occurred in cells that had lost the most Foxp3 expression (Supplementary Fig. 3). Thus, Foxp3 induced by TGF-β inhibits the function of RORγt.

These results prompted us to ask whether Foxp3 interacts with RORγt to inhibit its function. Using a yeast two-hybrid screen, we previously found that human FOXP3 interacts with RAR-related

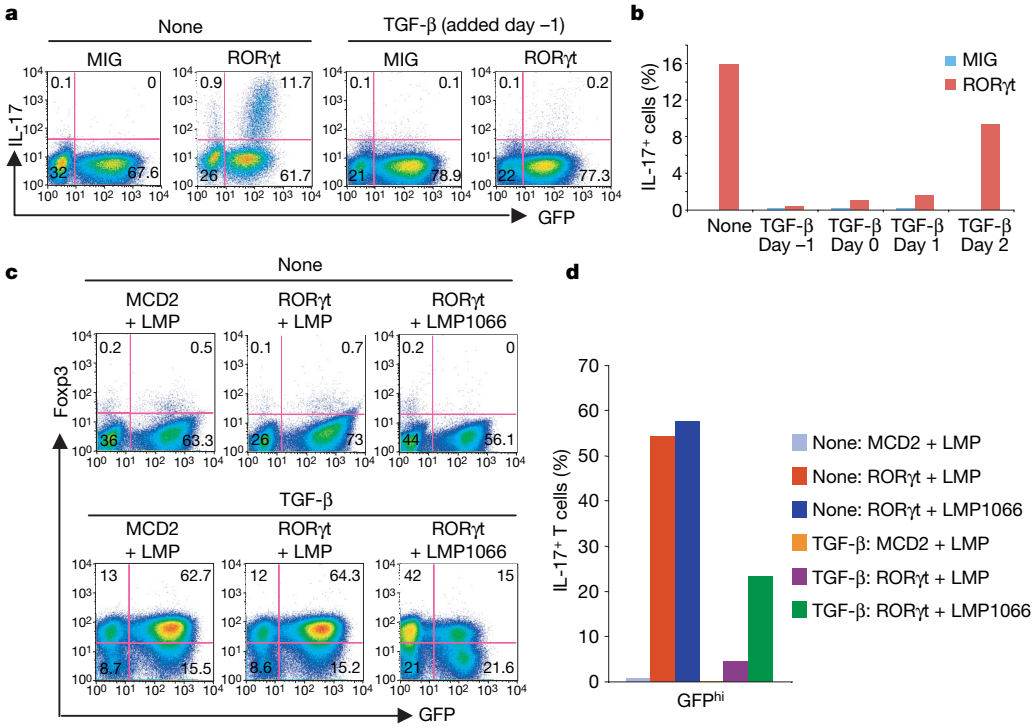


Figure 2 | TGF-β inhibits RORγt-directed IL-17 production by upregulating Foxp3. **a**, The effect of TGF-β on IL-17 expression after transduction of RORγt. Naive $CD4^+$ T cells incubated with TGF-β from day −1 were transduced with control vector MSCV-IRES-GFP (MIG) or RORγt-IRES-GFP (RORγt) on day 0 (24 h after TCR activation), and IL-17 intracellular staining was performed on day 5. **b**, The inhibitory effect of TGF-β when included at different times relative to transduction of RORγt. The percentage of $IL\text{-}17^+$ cells among GFP ($ROR\gamma t$)$^+$ cells is shown. **c**, Knockdown of TGF-β-induced Foxp3 expression with an shRNA against *Foxp3* (LMP1066). Naive $CD4^+$ T cells were stimulated as in **a** and co-transduced on days 0 and 1 with control retroviral construct MSCV-IRES-hCD2 (MCD2) or RORγt-IRES-hCD2 (RORγt) and the specific shRNA vector (LMP1066) or control vector (LMP). After transduction, the cells were cultured with or without TGF-β, and Foxp3 expression was measured by intracellular staining on day 5. **d**, Restoration of RORγt-induced IL-17 expression on knockdown of Foxp3. IL-17 expression was assessed in cells co-transduced as in **c** and gated for the level of GFP expression. The percentage of $IL\text{-}17^+$ T cells in GFP^{hi} cell populations is shown. Results with additional shRNA vectors that failed to downregulate Foxp3 expression were similar to those with the control LMP vector. Representative data from three experiments are shown.

©2008 Nature Publishing Group

orphan receptor A (RORA), and that an alternatively spliced isoform of FOXP3, lacking exon 2 (ref. 20), was deficient in this interaction[21]. We therefore examined whether mouse and human Foxp3 could similarly bind to RORγt, and whether such interaction was necessary for inhibition of the RORγt-mediated induction of IL-17. When Flag-epitope-tagged mouse Foxp3 was co-expressed with mouse RORγt in 293T cells, the two proteins were co-immunoprecipitated (Fig. 3a), even in the presence of DNase I or ethidium bromide, suggesting that the interaction does not involve DNA. A similar interaction was observed between human RORγT and FOXP3 (Fig. 3b). However, both mouse and human Foxp3 lacking the conserved exon 2-encoded sequence (Foxp3ΔEx2) had a substantially reduced association with RORγt (Fig. 3a, b). We examined the localization of the two proteins by confocal microscopy of HeLa cells transfected with Flag-tagged mouse Foxp3 constructs with or without mouse RORγt. Both Foxp3 and RORγt were localized in the nucleus, but Foxp3 lacking the DNA-binding forkhead domain (Foxp3ΔFKH) remained in the cytoplasm due to deletion of the nuclear localization signal in FKH[22] (Supplementary Fig. 4). However, Foxp3ΔFKH translocated to the nucleus when it was co-expressed with RORγt, indicating that the Foxp3–RORγt interaction is independent of FKH. Accordingly, Foxp3ΔFKH co-immunoprecipitated with RORγt in extracts of transfected 293T cells (data not shown). A combined Foxp3ΔEx2/ΔFKH mutant remained in the cytoplasm even when it was co-expressed with RORγt, further indicating that Foxp3 interacts with RORγt by way of the exon 2-encoded sequence (Supplementary Fig. 4).

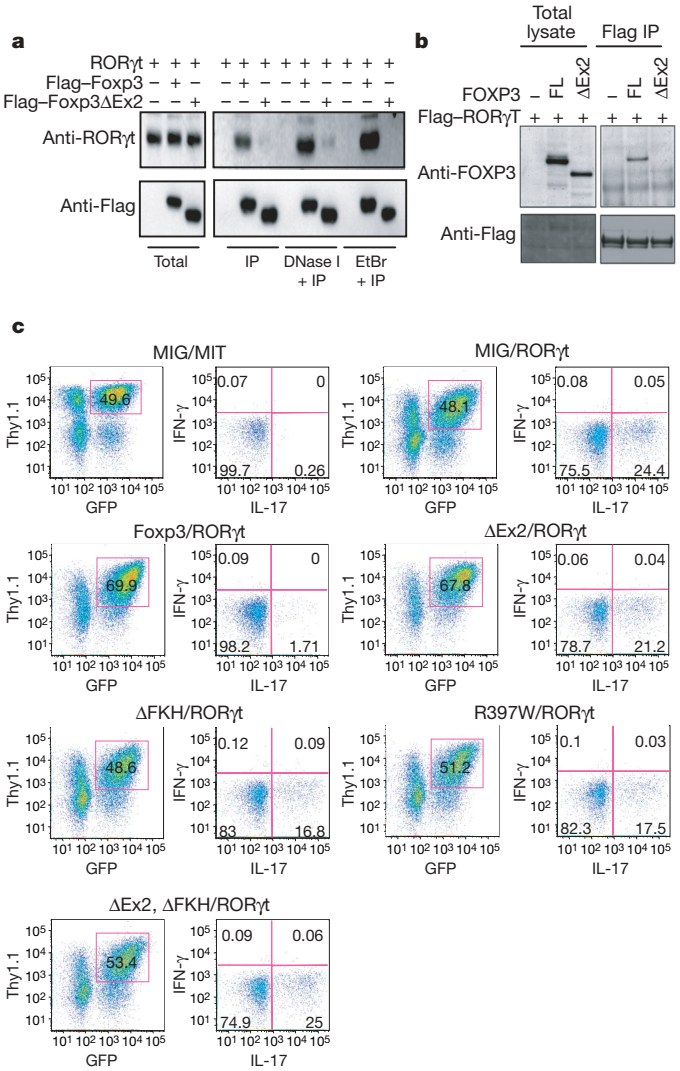


Figure 3 | Foxp3 interacts with RORγt and inhibits RORγt-directed IL-17 expression. a, Co-immunoprecipitation of Foxp3 and RORγt from extracts of co-transfected 293T cells with or without DNaseI or ethidium bromide (EtBr). Cells were transfected with mouse RORγt and Flag-tagged wild-type (Flag–Foxp3) or exon 2-deleted (Flag–Foxp3ΔEx2) Foxp3. Anti-Flag immunoprecipitates (IP) and total lysates were immunoblotted with anti-RORγt antibody and anti-Flag antibody. **b**, Cells were transfected with Flag-tagged human RORγT and full-length (FL) or the exon 2-deleted isoform of human FOXP3 (ΔEx2). Anti-Flag immunoprecipitates and total lysates were probed with anti-FOXP3 and anti-Flag antibodies. **c**, Naive $CD4^+$ T cells were co-transduced with retroviruses encoding RORγt (MIT vector, Thy1.1 reporter) and various mouse Foxp3 constructs (MIG vector, GFP reporter). IL-17 expression was assessed on day 4 in cells gated for expression of both Thy1.1 and GFP. Representative data from at least three experiments are shown in each of the panels.

To investigate the role of the interaction between Foxp3 and RORγt in the repression of RORγt-induced transcription, we co-expressed these transcription factors in naive $CD4^+$ T cells and examined expression of IL-17. Both mouse and human Foxp3 blocked RORγt-directed IL-17 expression, but full suppression required the presence of the exon 2-encoded sequence in Foxp3, suggesting that the interaction between Foxp3 and RORγt is essential (Fig. 3c and Supplementary Fig. 5). The ability of both mouse and human Foxp3 to repress RORγt-induced IL-17 expression was abrogated by deletion of the FKH domain or a point mutation in this domain (R397W) that impairs FOXP3 DNA-binding activity and was identified in X-linked immunodeficiency, polyendocrinopathy, enteropathy (IPEX) syndrome in humans[23,24] (Fig. 3c and Supplementary Fig. 5). Therefore, Foxp3 can block the activity of RORγt at least in part through an interaction involving a sequence encoded by exon 2, but the requirement for an intact FKH domain suggests that its DNA-binding activity also contributes to inhibition of IL-17 expression. Thus, Foxp3 may inhibit RORγt-directed transcription through a mechanism similar to that proposed for its inhibition of IL-2 expression, involving its association with NFAT1 and Runx1 (refs 25 and 26). However, Foxp3ΔEx2 was as effective as the full-length protein in suppressing expression of IL-2 and interferon-γ (IFN-γ) in primary mouse T cells, indicating that, like the naturally occurring human spliced isoform[20], it retains regulatory functions and can, presumably, associate with both NFAT1 and Runx1 (Supplementary Fig. 6).

Our results suggest that Foxp3 may inhibit RORγt activity on its target genes during T_H17 cell differentiation. To extend our analysis from *Il17* to other potential RORγt transcriptional targets, we examined the effect of TGF-β-induced Foxp3 on *Il23r* expression, which also requires the activity of RORγt[10,11]. Forced expression of wild-type mouse Foxp3 inhibited IL-6/IL-21-induced *Il23r* expression, whereas Foxp3ΔEx2 had less inhibitory activity (Fig. 4a); this is consistent with the notion that Foxp3 inhibits the function of RORγt through an interaction involving the sequence encoded by exon 2. Similar results were observed with *Il22* expression in response to IL-6 or IL-21 (data not shown). Expression of *Il22* and of *Il23r* in response to either IL-6 or forced expression of RORγt was also inhibited by high concentrations of TGF-β (refs 11, 27 and data not shown). However, at low concentrations, TGF-β synergized with IL-6 and IL-21 to enhance expression of *Il23r* mRNA (Fig. 4b). As a consequence, addition of IL-23 to cultures containing high concentrations of TGF-β had no effect on IL-17 expression, but significantly increased the number of $IL\text{-}17^+$ cells and the level of IL-17 expression per cell when low concentrations of TGF-β were used (Figs 4c, d and Supplementary Fig. 7). In contrast, induction of T_{reg} ($Foxp3^+$) cells was optimal at high concentrations of TGF-β, but there was little induction at TGF-β concentrations at which IL-23 had a synergistic effect on expression of IL-17 (Fig. 4e).

TGF-β-induced Foxp3 expression is inhibited by IL-6 (ref. 17), IL-21 (ref. 10) and IL-23 (Supplementary Fig. 8). However, a substantial number of $Foxp3^+$ cells differentiated in response to TGF-β, even in the presence of IL-6, and many of these cells also expressed IL-17 (Supplementary Fig. 9a). Conversely, many of the $IL\text{-}17^+$ cells also

©2008 Nature Publishing Group

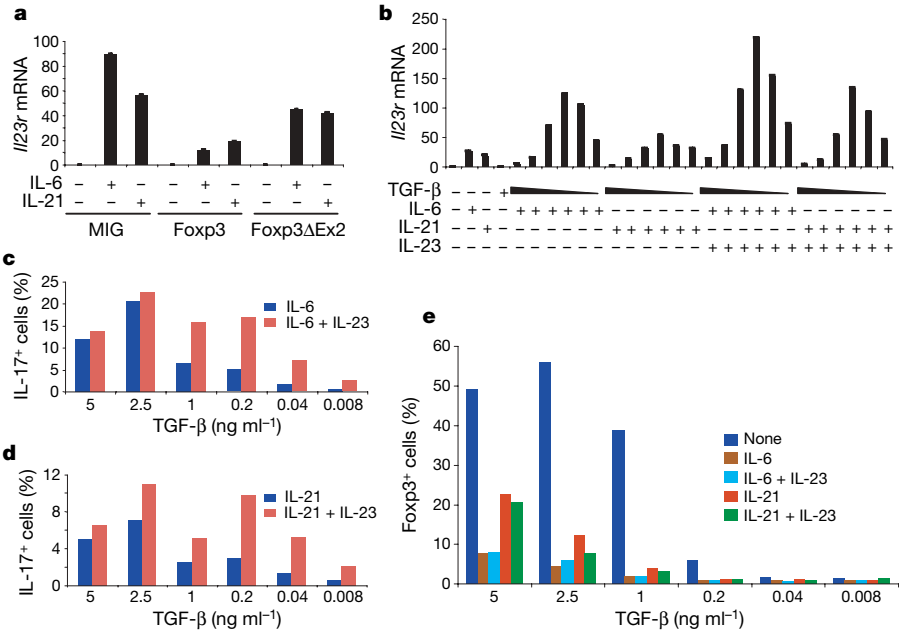


Figure 4 | TGF-β concentration influences *Il23r* expression and levels of IL-17 in response to T_H17-inducing cytokines. a, Foxp3-mediated inhibition of IL-6/IL-21-induced *Il23r* expression. This was measured in arbitrary units relative to expression of transcripts encoding actin. Naive $CD4^+$ T cells were transduced with MIG, full length Foxp3 or Foxp3ΔEx2 viruses, and were treated with the indicated cytokines. RNA was isolated from GFP^+ cells at day 2. *Il23r* expression was measured by real-time RT–PCR and was normalized to the actin level. Error bars represent standard deviations obtained using the standard curve method. **b**, Induction of *Il23r* mRNA in response to cytokines. Naive $CD4^+$ T cells were stimulated with anti-CD3 and anti-CD28 throughout the culture period in the presence of the indicated combinations of cytokines. TGF-β was titrated into the cultures at the concentrations of 5 ng ml^{-1}, 2.5 ng ml^{-1}, 1 ng ml^{-1}, 200 pg ml^{-1}, 40 pg ml^{-1} or 8 pg ml^{-1}. *Il23r* mRNA expression was measured after 48 h by real-time RT–PCR and was normalized to the actin expression level. **c**, **d**, IL-23 enhancement of IL-17 expression at low concentrations of TGF-β. The percentage of $IL\text{-}17^+$ cells at 96 h of stimulation with the indicated cytokines is shown. Results in histogram format are shown in Supplementary Fig. 7. **e**, Induction of Foxp3 at different concentrations of TGF-β. Representative data from at least three experiments are shown for each set of panels.

expressed Foxp3. Thus, RORγt-dependent IL-17 expression can occur in the presence of Foxp3, but the level of Foxp3 may be insufficient to block RORγt function or, alternatively, IL-6 may overcome the inhibitory function of Foxp3. To examine this possibility, we added IL-6 or IL-21 to cultures of cells transduced with both RORγt and Foxp3. Under these conditions, the inhibitory effect of Foxp3 on IL-17 induction was largely circumvented, even though the level of Foxp3 protein was not affected (Supplementary Fig. 9b and data not shown); this suggests that IL-6 and IL-21 may have an additional post-translational effect on either Foxp3 or RORγt.

Our data collectively suggest that T cells receiving a TGF-β signal can acquire the potential to develop into either the T_{reg} or the T_H17 lineage. Foxp3 induction restrains the differentiation of inflammatory T_H17 cells in response to TGF-β in the absence of other pro-inflammatory cytokines by inhibiting the activity of RORγt. In the presence of pro-inflammatory cytokines, the suppression of Foxp3 expression and inhibitory function, together with the concurrent upregulation or stabilization of RORγt expression, leads to full progression towards the T_H17 lineage (Supplementary Fig. 10). This process may be especially relevant in the intestinal lamina propria, in which TGF-β can promote either T_H17 or T_{reg} cell lineage differentiation, depending on its local concentration. In this setting, a fine balance between RORγt and Foxp3 may be critical for immune homeostasis. In line with the observation that more $Foxp3^+$ T_{reg} cells were present in the gut of RORγt-deficient mice (Fig. 1b), these mutant mice were also protected from autoimmune disease (ref. 7 and data not shown). Conversely, a decrease of Foxp3 expression and function and an increase of RORγt expression tips the T_{reg}/T_H17 balance towards the T_H17 cell lineage. This may occur in some autoimmune diseases, as suggested by the finding that an *Il23r* polymorphism correlates with protection from Crohn's disease[28]. These results therefore have important implications for how peripheral tolerance is maintained in the presence of potentially pro-inflammatory cytokines.

METHODS SUMMARY

Mice. C57BL/6 mice (Taconic), mice with a *GFP* reporter cDNA knocked in at the RORγt translation initiation site[29], mice with an *IRES-YFP-Cre* cDNA knocked into the 3′ UTR of the *Foxp3* locus (Y.P.R. *et al.*, submitted) and *Rosa26*$^{stop-YFP}$ (ref. 18) mice were kept in specific pathogen-free (SPF) conditions at the animal facility of the Skirball Institute. All animal experiments were performed in accordance with approved protocols for the NYU Institutional Animal Care and Usage Committee.

Cell culture. Naive $CD4^+$ T cells were purified and cultured as described previously[7]. In brief, 1.5×10^6 naive $CD4^+$ T cells were cultured in wells of 24-well plates (or 0.7×10^6 cells per well in 48-well plates) containing plate-bound anti-CD3 (5 μg ml^{-1}) and soluble anti-CD28 (1 μg ml^{-1}). Cultures were supplemented with 2 μg ml^{-1} anti-IL-4 (BD Pharmingen), 2 μg ml^{-1} anti-IFN-γ (BD Pharmingen) with or without 80 U ml^{-1} human IL-2 (a gift from S. Reiner), 20 ng ml^{-1} IL-6 (eBioscience), 5 ng ml^{-1} TGF-β (PeproTech), 50 ng ml^{-1} IL-21 (R&D Systems) and 10 ng ml^{-1} IL-23 (eBioscience). Viral transduction was performed as described previously, unless indicated otherwise in the text[7]. T cells were isolated from the small intestinal lamina propria as described previously[7].

General. All DNA constructs were generated by PCR-based methodology and confirmed by sequencing. Retroviral production and transduction were performed as described previously[7]. Protein–protein interaction was detected by co-immunoprecipitation and confocal microscopy in 293T cells and HeLa cells. Gene expression analysis was monitored by real-time PCR with reverse transcription (RT–PCR) using gene-specific primers and probes. IL-17 and Foxp3 protein expression were examined by intracellular staining performed according to the manufacturer's protocol. Co-expression of RORγt and Foxp3 was examined by immunofluorescence using anti-RORγ[30] and anti-Foxp3 antibodies.

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

Received 23 January; accepted 4 March 2008.
Published online 26 March 2008.

1. Weaver, C. T., Harrington, L. E., Mangan, P. R., Gavrieli, M. & Murphy, K. M. Th17: an effector CD4 T cell lineage with regulatory T cell ties. *Immunity* **24**, 677–688 (2006).

©2008 Nature Publishing Group

2. Weaver, C. T., Hatton, R. D., Mangan, P. R. & Harrington, L. E. IL-17 family cytokines and the expanding diversity of effector T cell lineages. *Annu. Rev. Immunol.* **25,** 821–852 (2007).
3. McKenzie, B. S., Kastelein, R. A. & Cua, D. J. Understanding the IL-23–IL-17 immune pathway. *Trends Immunol.* **27,** 17–23 (2006).
4. Lee, E. *et al.* Increased expression of interleukin 23 p19 and p40 in lesional skin of patients with psoriasis vulgaris. *J. Exp. Med.* **199,** 125–130 (2004).
5. Witowski, J., Ksiazek, K. & Jorres, A. Interleukin-17: a mediator of inflammatory responses. *Cell. Mol. Life Sci.* **61,** 567–579 (2004).
6. Veldhoen, M., Hocking, R. J., Atkins, C. J., Locksley, R. M. & Stockinger, B. TGFβ in the context of an inflammatory cytokine milieu supports *de novo* differentiation of IL-17-producing T cells. *Immunity* **24,** 179–189 (2006).
7. Ivanov, I. I. *et al.* The orphan nuclear receptor RORγt directs the differentiation program of proinflammatory IL-17$^+$ T helper cells. *Cell* **126,** 1121–1133 (2006).
8. Chen, W. *et al.* Conversion of peripheral CD4$^+$CD25$^-$ naive T cells to CD4$^+$CD25$^+$ regulatory T cells by TGF-β induction of transcription factor Foxp3. *J. Exp. Med.* **198,** 1875–1886 (2003).
9. Korn, T. *et al.* IL-21 initiates an alternative pathway to induce proinflammatory T_H17 cells. *Nature* **448,** 484–487 (2007).
10. Nurieva, R. *et al.* Essential autocrine regulation by IL-21 in the generation of inflammatory T cells. *Nature* **448,** 480–483 (2007).
11. Zhou, L. *et al.* IL-6 programs T_H-17 cell differentiation by promoting sequential engagement of the IL-21 and IL-23 pathways. *Nature Immunol.* **8,** 967–974 (2007).
12. Murphy, K. M. & Reiner, S. L. The lineage decisions of helper T cells. *Nature Rev. Immunol.* **2,** 933–944 (2002).
13. Langrish, C. L. *et al.* IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *J. Exp. Med.* **201,** 233–240 (2005).
14. Aggarwal, S., Ghilardi, N., Xie, M. H., de Sauvage, F. J. & Gurney, A. L. Interleukin-23 promotes a distinct CD4 T cell activation state characterized by the production of interleukin-17. *J. Biol. Chem.* **278,** 1910–1914 (2003).
15. Cua, D. J. *et al.* Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. *Nature* **421,** 744–748 (2003).
16. Letterio, J. J. & Roberts, A. B. Regulation of immune responses by TGF-β. *Annu. Rev. Immunol.* **16,** 137–161 (1998).
17. Bettelli, E. *et al.* Reciprocal developmental pathways for the generation of pathogenic effector T_H17 and regulatory T cells. *Nature* **441,** 235–238 (2006).
18. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1,** 4 (2001).
19. Gavin, M. A. *et al.* Foxp3-dependent programme of regulatory T-cell differentiation. *Nature* **445,** 771–775 (2007).
20. Allan, S. E. *et al.* The role of 2 FOXP3 isoforms in the generation of human CD4$^+$ Tregs. *J. Clin. Invest.* **115,** 3276–3284 (2005).
21. Du, J., Huang, C., Zhou, B. & Ziegler, S. F. Isoform-specfic inhibition of RORα-mediated transcriptional activation by human FOXP3. *J. Immunol.* (in the press).
22. Schubert, L. A., Jeffery, E., Zhang, Y., Ramsdell, F. & Ziegler, S. F. Scurfin (FOXP3) acts as a repressor of transcription and regulates T cell activation. *J. Biol. Chem.* **276,** 37672–37679 (2001).
23. Wildin, R. S. *et al.* X-linked neonatal diabetes mellitus, enteropathy and endocrinopathy syndrome is the human equivalent of mouse scurfy. *Nature Genet.* **27,** 18–20 (2001).
24. Lopes, J. E. *et al.* Analysis of FOXP3 reveals multiple domains required for its function as a transcriptional repressor. *J. Immunol.* **177,** 3133–3142 (2006).
25. Ono, M. *et al.* Foxp3 controls regulatory T-cell function by interacting with AML1/Runx1. *Nature* **446,** 685–689 (2007).
26. Wu, Y. *et al.* FOXP3 controls regulatory T cell function through cooperation with NFAT. *Cell* **126,** 375–387 (2006).
27. Zheng, Y. *et al.* Interleukin-22, a T_H17 cytokine, mediates IL-23-induced dermal inflammation and acanthosis. *Nature* **445,** 648–651 (2007).
28. Duerr, R. H. *et al.* A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* **314,** 1461–1463 (2006).
29. Eberl, G. *et al.* An essential function for the nuclear receptor RORγt in the generation of fetal lymphoid tissue inducer cells. *Nature Immunol.* **5,** 64–73 (2004).
30. Sun, Z. *et al.* Requirement for RORγ in thymocyte survival and lymphoid organ development. *Science* **288,** 2369–2373 (2000).

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

Acknowledgements We thank P. Lopez and J. Hirst for assistance with cell sorting. We also thank J. Lafaille and D. Unutmaz for critical reading of the manuscript, and members of the Littman laboratory for their suggestions. L.Z., M.M.W.C. and I.I.I. were supported by fellowships from the Irvington Institute for Immunological Research, the Cancer Research Institute and the Crohn's and Colitis Foundation of America, respectively. This work was supported by the Howard Hughes Medical Institute (D.R.L., A.Y.R.), the Sandler Program for Asthma Research (D.R.L.), the National Multiple Sclerosis Society (D.R.L.), the Helen and Martin Kimmel Center for Biology and Medicine (D.R.L.), NIH grant AI48779 (S.F.Z.), and the JDRF Collaborative Center for Cell Therapy (S.F.Z.).

Author Contributions L.Z., J.E.L., M.M.W.C, I.I.I. and Y.S. performed the experiments with assistance from R.M. and G.D.V. Y.P.R. and A.Y.R. provided mice for fate-mapping experiments. L.Z., S.F.Z. and D.R.L. designed the experiments, and L.Z. and D.R.L. wrote the manuscript with input from the co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.R.L. (littman@saturn.med.nyu.edu).

©2008 Nature Publishing Group

METHODS

Plasmids and retrovirus production. MIG, MIT and MCD2 are retrovirus-based vectors containing GFP, Thy1.1 and human CD2, respectively, under the regulation of an internal ribosome entry site (IRES). The *Rorc(γt)* cDNA was amplified by PCR and cloned into MIG (RORγt-IRES-GFP), MIT (RORγt-IRES-Thy1.1) and MCD2 (RORγt-IRES-hCD2). The wild-type full-length *Foxp3* cDNA and various *Foxp3* mutant cDNAs were amplified by PCR with a 5′ Flag-tagged primer and a 3′ corresponding primer, and were cloned into MIG (Foxp3-IRES-GFP). *Il23r* cDNA was PCR-amplified and cloned into MIG (IL-23R-IRES-GFP). Human *FOXP3* cDNAs were PCR-amplified and cloned into MIG. MSCV-LTRmiR30-PIG (LMP) is a commercial vector from Openbiosystems. A double-stranded DNA oligonucleotide that targets the coding region of *Foxp3* was cloned into LMP (LMP1066) according to the manufacturer's protocol (the target sequence of *Foxp3* is 5′-GGCAGAGG-ACACTCAATGAAAT-3′). Retrovirus production was as described previously[7].

Surface and intracellular staining, and CFSE labelling. For intracellular staining, cells obtained from *in vitro* culture or isolated from the small intestinal lamina propria were incubated for 4–5 h with 50 ng ml^{-1} PMA (Sigma) and 500 ng ml^{-1} ionomycin (Sigma), plus 2 μg ml^{-1} Brefeldin A (Sigma) during the last 2 h. The cells were kept in a tissue-culture incubator at 37 °C. Surface staining was performed for 15–20 min with the corresponding cocktail of fluorescently labelled antibodies. After surface staining, the cells were resuspended in a Fixation/Permeabilization solution (BD Pharmingen), and intracellular cytokine staining was performed according to the manufacturer's protocol. For intracellular staining of Foxp3, the Foxp3-Staining Buffer Set (fixation/permeabilization and permeabilization buffers) was used (eBioscience) according to the manufacturer's protocol. For carboxyfluoroscein succinimidyl ester (CFSE)-labelling, sorted naive CD4$^+$ T cells were washed twice with Hank's Buffered Salt Solution (HBSS; Invitrogen), and labelled with 5 μM CFSE (Sigma) in HBSS for 10 min at 20 °C. The labelling was then stopped by adding 1/5 volume of FCS. The labelled cells were washed twice with the T cell culture medium before they were seeded and stimulated as described in the text.

Real-time RT–PCR. Complementary DNA was synthesized and analysed by real-time quantitative PCR as described previously[7]. The starting quantity of the initial cDNA sample was calculated from primer-specific standard curves by using the iCycler Data Analysis Software. The expression level of each gene was normalized to the expression level of actin using the standard curve method. The primer sets and probes for real-time PCR were described elsewhere[7,11].

Co-immunoprecipitation and western blot. Cells (293T cells) were transfected with the indicated constructs using Lipofectamine 2000 (Invitrogen). Forty-eight hours after transfection, whole-cell extracts were made in the lysis buffer, which contained 50 mM Tris-HCl (pH 8.0), 120 mM NaCl, 4 mM EDTA, 1% NP-40, 50 mM NaF, 1 mM Na_3VO_4 and protease inhibitors. After the insoluble material was removed by centrifugation, the lysate was immunoprecipated for 12–16 hours at 4 °C with anti-Flag M2 agarose beads (Sigma). After extensive washes with the lysis buffer, samples were resolved in an SDS–polyacrylamide gel electrophoresis (SDS–PAGE) gel and transferred to a nitrocellulose membrane. Western blotting was performed with an anti-Flag monoclonal antibody (Sigma), an anti-Foxp3 monoclonal antibody (eBioscience) and an anti-RORγt hamster monoclonal antibody[30].

Confocal microscopy. HeLa cells were plated on 8-well glass slides (Lab-Tek II Chamber Slide System) before transfection with the indicated constructs using Lipofectamine 2000 (Invitrogen). Forty-eight hours after transfection, cells were washed once in PBS, fixed for 15 min in 2% paraformaldehyde in phosphate buffer (PBS without saline), and then washed twice in PBS. Cells were blocked and permeabilized in PBS-XG (10% goat serum (Sigma) in PBS containing 0.1% Triton X-100) for 1 h at 20 °C. The cells were then incubated for 12–16 hours at 4 °C with anti-RORγt hybridoma supernatant[30] (1:2 dilution in PBS-XG). After two washes in PBS, the cells were incubated for 1 h at room temperature with Cy3-conjugated goat anti-hamster antibody (Jackson ImmunoResearch Laboratory) at 1:400 dilution in PBS-XG. The cells were then washed three times in PBS and incubated for 1 h at 20 °C with anti-Flag M2 monoclonal antibody (Sigma) at 1:1,000 dilution in PBS-XG. After two washes in PBS, the cells were incubated for 1 h at 20 °C with anti-mouse Alexa 633 (Molecular Probes) at 1:200 dilution in PBS-XG. The cells were then washed twice in PBS and incubated for 5 min at 20 °C with 1 μg ml^{-1} DAPI (Sigma), washed two more times in PBS and mounted with Fluoromount-G (Southern Biotechnology Associates). The cells were examined with a Zeiss ZMD510 microscope with a CCD camera, and images were processed with Zeiss LSM Image Browser 4.0 and Adobe Photoshop 7.0.

Immunofluorescence. Naive T cells were sorted as described previously and stimulated in the presence of the indicated cytokines as described[7]. Lamina propria lymphocytes from *Rorc(γt)*$^{gfp/+}$ small intestines were isolated as described[7], and CD4$^+$GFP^{int} and CD4$^+$GFP$^-$ cells were sorted on a MoFlo cytometer (DAKO Cytomation). Naive T cells or sorted lamina propria T cells were then cytospinned on glass slides and fixed in 2% paraformaldehyde in phosphate buffer (PBS without saline) for 20 min at 20 °C. After blocking, immunofluorescence staining was performed by incubating the cells consecutively with the anti-RORγ antibody[30] (hybridoma supernatant 1:4) for 12- 16 hours at 4 °C and biotin anti-mouse/rat Foxp3 monoclonal antibody (eBioscience clone FJK-16, 1:200 dilution) for 1.5 hours at 20 °C. The blocking solution contained PBS, 0.1% Triton-X100 and 10% goat serum. Secondary goat anti-Armenian-hamster Cy3 conjugated antibody (Jackson Immunoresearch) and streptavidin–APC (eBioscience), both at 1:400 dilution, were used at 20 °C for 1.5 hours to detect the RORγ and Foxp3 primary antibodies, respectively.

©2008 Nature Publishing Group

LETTERS

Imaging of Rab5 activity identifies essential regulators for phagosome maturation

Masahiro Kitano[1,3], Michio Nakaya[2,4], Takeshi Nakamura[1], Shigekazu Nagata[2,4] & Michiyuki Matsuda[1]

Efficient phagocytosis of apoptotic cells is crucial for tissue homeostasis and the immune response[1,2]. Rab5 is known as a key regulator of the early endocytic pathway[3] and we have recently shown that Rab5 is also implicated in apoptotic cell engulfment[4]; however, the precise spatio-temporal dynamics of Rab5 activity remain unknown. Here, using a newly developed fluorescence resonance energy transfer biosensor, we describe a change in Rab5 activity during the engulfment of apoptotic thymocytes. Rab5 activity on phagosome membranes began to increase on disassembly of the actin coat encapsulating phagosomes. Rab5 activation was either continuous or repetitive for up to 10 min, but it ended before the collapse of engulfed apoptotic cells. Expression of a dominant-negative mutant of Rab5 delayed this collapse of apoptotic thymocytes, showing a role for Rab5 in phagosome maturation. Disruption of microtubules with nocodazole inhibited Rab5 activation on the phagosome membrane without perturbing the engulfment of apoptotic cells. Furthermore, we found that Gapex-5 is the guanine nucleotide exchange factor essential for Rab5 activation during the engulfment of apoptotic cells. Gapex-5 was bound to a microtubule-tip-associating protein, EB1, whose depletion inhibited Rab5 activation during phagocytosis. We therefore propose a mechanistic model in which the recruitment of Gapex-5 to phagosomes through the microtubule network induces the transient Rab5 activation.

To make Rab5 activity visible in living cells, we developed a genetically encoded fluorescence resonance energy transfer (FRET) probe, designated Raichu-Rab5. The probe comprised a modified yellow fluorescent protein (YFP) called Venus, the amino-terminal Rab5-binding domain of EEA1, a modified cyan fluorescent protein (CFP) called SECFP, and Rab5 (Fig. 1a). In this probe design, an increase in Rab5-GTP results in an increase in FRET, which can be represented by the 525 nm/475 nm emission ratio. Characterization of Raichu-Rab5 was conducted similarly to that of other Raichu probes reported previously[5,6]. In comparison with the wild-type Rab5 probe, Raichu-Rab5-Q79L—which lacks GTPase activity—had an increased FRET efficiency, whereas Raichu-Rab5-S34N—which shows a reduced affinity for guanine nucleotides—had a decreased FRET efficiency, as expected (Supplementary Fig. 1a). The GTP loading of the Rab5 probes correlated well with that of the authentic Rab5 proteins (Supplementary Fig. 1b), and the GTP loadings obtained here were similar to those reported previously[7]. Next, we examined the sensitivity of Raichu-Rab5 to guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs) (Supplementary Fig. 1c). Rab5 GEFs such as Rabex-5, Rin1 and Gapex-5 induced a dose-dependent increase in the FRET efficiency of Raichu-Rab5. In contrast, the expression of Rab5 GAPs such as RabGAP-5 and RN-tre decreased the FRET efficiency in a dose-dependent manner. These results indicate that Raichu-Rab5 is capable of monitoring the balance between GEF and GAP activities towards Rab5.

We then examined the subcellular distribution of Raichu-Rab5. Rab5 localizes at early endosomes and the plasma membrane. Raichu-Rab5 co-localized with authentic Rab5a at endosomal structures in HeLa cells (Supplementary Fig. 2a, top panel). The distribution of Raichu-Rab5 did not overlap with those of Rab7 and Rab11a, which were used as markers for late and recycling endosomes, respectively (Supplementary Fig. 2a, middle and bottom panels). The subcellular distribution of Raichu-Rab5 was very similar to those of early endosomal markers, namely the FYVE domain of EEA1 (ref. 8) and the PX domain of $p40^{phox}$ (also known as NCF4)[9] (Supplementary Fig. 2b). Finally, Raichu-Rab5 was found to bind to Rab guanine dissociation inhibitor (RabGDI) as efficiently as authentic Rab5a (data not shown), suggesting that Raichu-Rab5 is also subject to regulation by RabGDI.

By using Raichu-Rab5 we revealed Rab5 activation during the milk-fat-globule epidermal growth factor 8 (MFG-E8)-mediated engulfment of apoptotic cells by Swiss3T3 cells stably expressing integrin $\alpha_v\beta_3$ (Fig. 1b and Supplementary Video 1). The progress of phagocytosis was monitored by phase-contrast images, in which the completion of engulfment was recognizable by the transition of the engulfed apoptotic cells from phase-bright to phase-dark[10]. The accumulation of Raichu-Rab5 around the engulfed apoptotic cells, and an increase in FRET, began during the period of phase shift and reached a peak after the completion of engulfment.

During the course of experiments, we noticed that the abundance of cytoplasmic Raichu-Rab5, probably associated with RabGDI, hampered the quantification of FRET at the membrane. We therefore developed a Raichu-Rab5 probe fused to the carboxy terminus of Ki-Ras protein (Raichu-Rab5/PM), which localized at the plasma membrane and on phagocytic vesicles. This alteration did not affect the activation status of Rab5 probe (Supplementary Fig. 3). With this improved Raichu-Rab5 probe, we examined the time course of Rab5 activation more precisely (Figs 1c and 2e, and Supplementary Video 2). We set the zero time point to be the frame immediately before the initiation of the phase shift, which lasted about 3 min on average. Rab5 activation started during this period of phase shift and reached a peak within 4 min on average. Very similar results were obtained in a macrophage cell line, BAM3 (Supplementary Fig. 4). The duration of Rab5 activation varied significantly between different phagosomes. In addition, 37% of phagosomes showed reactivation of Rab5 after the cessation of initial Rab5 activation (Figs 1d and 3c, and Supplementary Video 3). The time course of this Rab5 reactivation also varied between phagosomes: some phagosomes showed a repetitive cycle of activation and inactivation, whereas others showed an abrupt and transient reactivation.

[1]Laboratory of Bioimaging and Cell Signaling, Graduate School of Biostudies, and [2]Department of Medical Chemistry, Graduate School of Medicine, Kyoto University, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. [3]Department of Signal Transduction, Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita-shi, Osaka 565-0871, Japan. [4]Solution Oriented Research for Science and Technology, Japan Science and Technology Corporation, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.

©2008 Nature Publishing Group

To gain insight into the role of Rab5 activation in phagocytosis, we examined the precise timing of actin assembly, Rab5 recruitment around phagocytic vesicles, and breakdown of apoptotic thymocytes in Swiss 3T3/integrin $\alpha_v\beta_3$ cells (Fig. 2). As revealed with enhanced green fluorescent protein (EGFP)–actin, actin fibres could be seen around the phagocytic vesicles from −3 to 2 min, showing that Rab5 is activated during actin disassembly (Fig. 2a, d, e). This observation was shown more directly by simultaneous imaging of monomeric red fluorescent protein (mRFP)–actin and Raichu-Rab5 (Supplementary Fig. 5). Next, accumulation of mEGFP–Rab5 on the phagosome membrane was revealed with mRFP as a reference of entirely cytosolic protein[11] (Fig. 2b, f). The ratio of mEGFP–Rab5 versus mRFP increased on phagosomes after 3–7 min of internalization, indicating that Rab5 was recruited to phagosomes during this period. This time course of Rab5 accumulation was similar to that of Rab5 activation. Finally, we examined the timing of apoptotic cell breakdown by the use of EGFP-expressing thymocytes. Owing to the acid-labile nature of EGFP, the collapse of apoptotic cells within phagosomes could be monitored by the extinction of fluorescence. With this method, we found that EGFP-expressing thymocyte fluorescence disappeared during the period from 10 to 20 min (Fig. 2c, g). We confirmed, by the use of a pH indicator, that the extinction of EGFP fluorescence was correlated with an increase in intracellular pH level (Supplementary Fig. 6). Thus, Rab5 inactivation preceded the decrease in pH and the breakdown of apoptotic cells. The expression of Rab5-S34N delayed, but did not block, the breakdown of apoptotic thymocytes (Fig. 2h). This indicated that Rab5 activation has a function in phagosome maturation during the engulfment of apoptotic cells, although Rab5-independent pathways may exist. A similar role for Rab5 has been proposed in FcγR-mediated phagocytosis[12,13]. This is supposed to come from the basic function of Rab5 mediating the fusion of phagosomes with early endosomes[14,15]. Furthermore, active Rab5 is necessary for the recruitment of Rab7 (ref. 13), which regulates the fusion of phagosomes with late endosomes to initiate the lysis of cell corpses.

In the engulfment process, microtubules were required for Rab5 activation. Treatment with 0.1 μM nocodazole significantly decreased Rab5 activation at the engulfment site (Fig. 3a), indicating that Rab5 activation depends on the integrity of the microtubule network. When microtubules were stabilized with 10 nM paclitaxel, initial Rab5 activation was observed as in the control cells, but repetitive Rab5 activation was inhibited significantly (Fig. 3b, c). Thus, it seems that the repetitive activation of Rab5 as shown in Fig. 1d is due to the dynamic instability of microtubule plus-ends.

To search for the molecular link between microtubule and Rab5 activation during phagocytosis, candidate Rab5 GEFs were depleted by RNA interference, and the resulting phenotypes were examined. The expressions of Rabex-5, Rin1 and Gapex-5 were markedly decreased by transfection of the respective short interfering RNA (siRNAs) in Swiss 3T3/integrin $\alpha_v\beta_3$ cells (Supplementary Fig. 7a). We then monitored the Rab5 activity during phagocytosis in cells lacking each Rab5 GEF. The depletion of Gapex-5 significantly inhibited Rab5 activation at engulfment sites, whereas neither Rabex-5 depletion nor Rin1 depletion showed any effect on Rab5 activation during phagocytosis (Fig. 3d). Expression of siRNA-resistant Gapex-5 restored the Rab5 activation (Fig. 3d and Supplementary Fig. 7c). The involvement of Gapex-5 in Rab5 activation was corroborated by

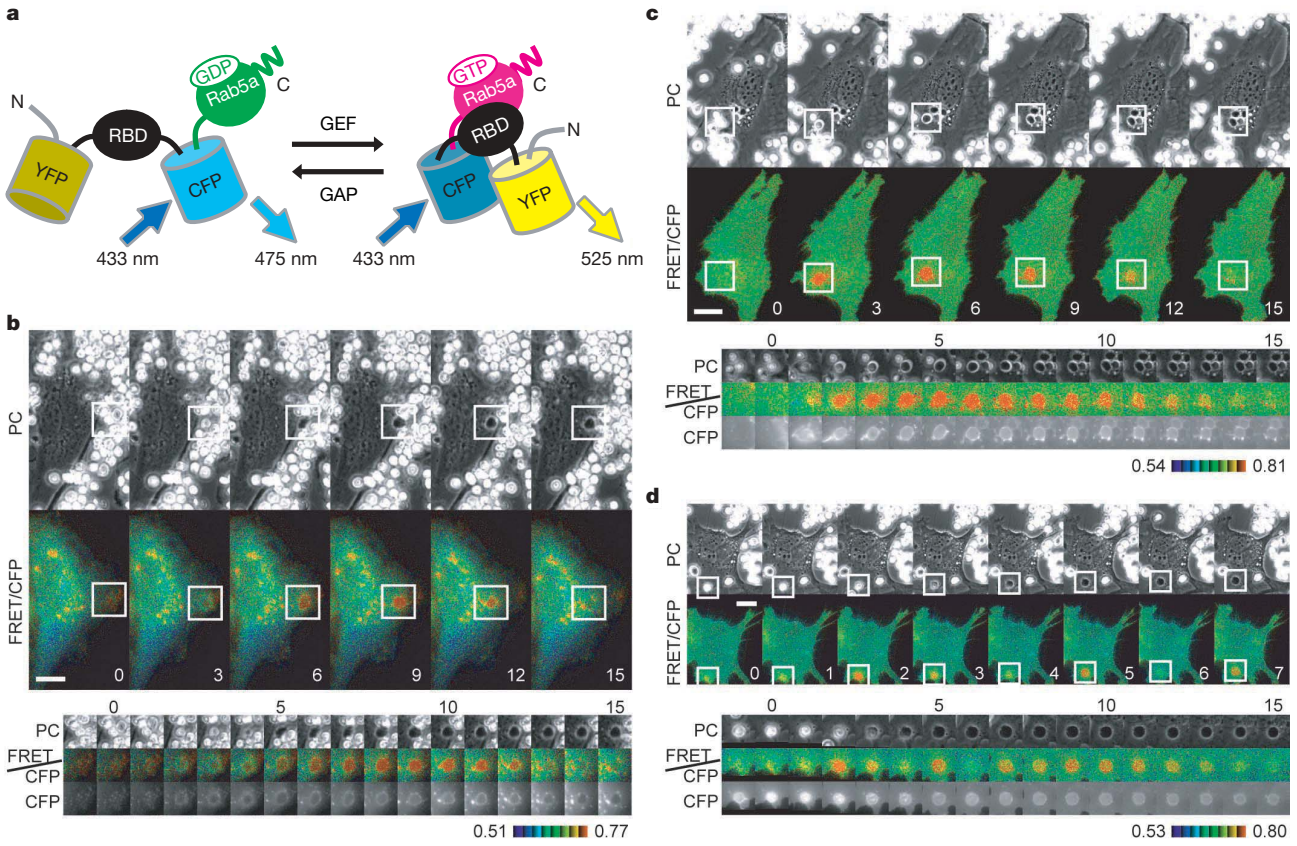


Figure 1 | Transient Rab5 activation during phagocytosis monitored by Raichu-Rab5. **a**, Schematic representation of Raichu-Rab5 bound to GDP or GTP. **b–d**, $\alpha_v\beta_3$ integrin-expressing Swiss3T3 cells were transfected with pRaichu-Rab5 (**b**) or pRaichu-Rab5/PM (**c**, **d**) and cultured together with apoptotic thymocytes in the presence of MFG-E8. Thereafter, images were obtained every 1 min. Top panels show phase-contrast (PC) and FRET/CFP ratio images at the indicated time points (min). In the intensity-modulated display mode shown here, eight colours from red to blue are used to represent the FRET/CFP ratio, with the intensity of each colour indicating the mean intensity of FRET and CFP. The upper and lower limits of the ratio range are shown at the bottom, and the divisions are evenly spaced. Time sequences in the bottom panels show the PC, FRET/CFP ratio and CFP images of the engulfed sites marked by white squares in the top panels. Scale bars, 20 μm.

©2008 Nature Publishing Group

a significant delay in the duration of EGFP-expressing thymocyte fluorescence in Gapex-5-depleted cells (Fig. 3e). These results demonstrated for the first time that Gapex-5 is an essential regulator for Rab5 activation and phagosome maturation during phagocytosis of apoptotic cells. We assumed that Gapex-5 is delivered to phagosomes by a microtubule-dependent mechanism. This idea could explain the observation that the reduced dynamic instability of microtubules in paclitaxel-treated cells suppressed repetitive Rab5 activation on phagosomes (Fig. 3c). Because Gapex-5 did not show a high affinity for microtubules by itself, we searched for the protein that could link Gapex-5 to microtubule and found that a microtubule-tip-associating protein, EB1 (ref. 16), was associated

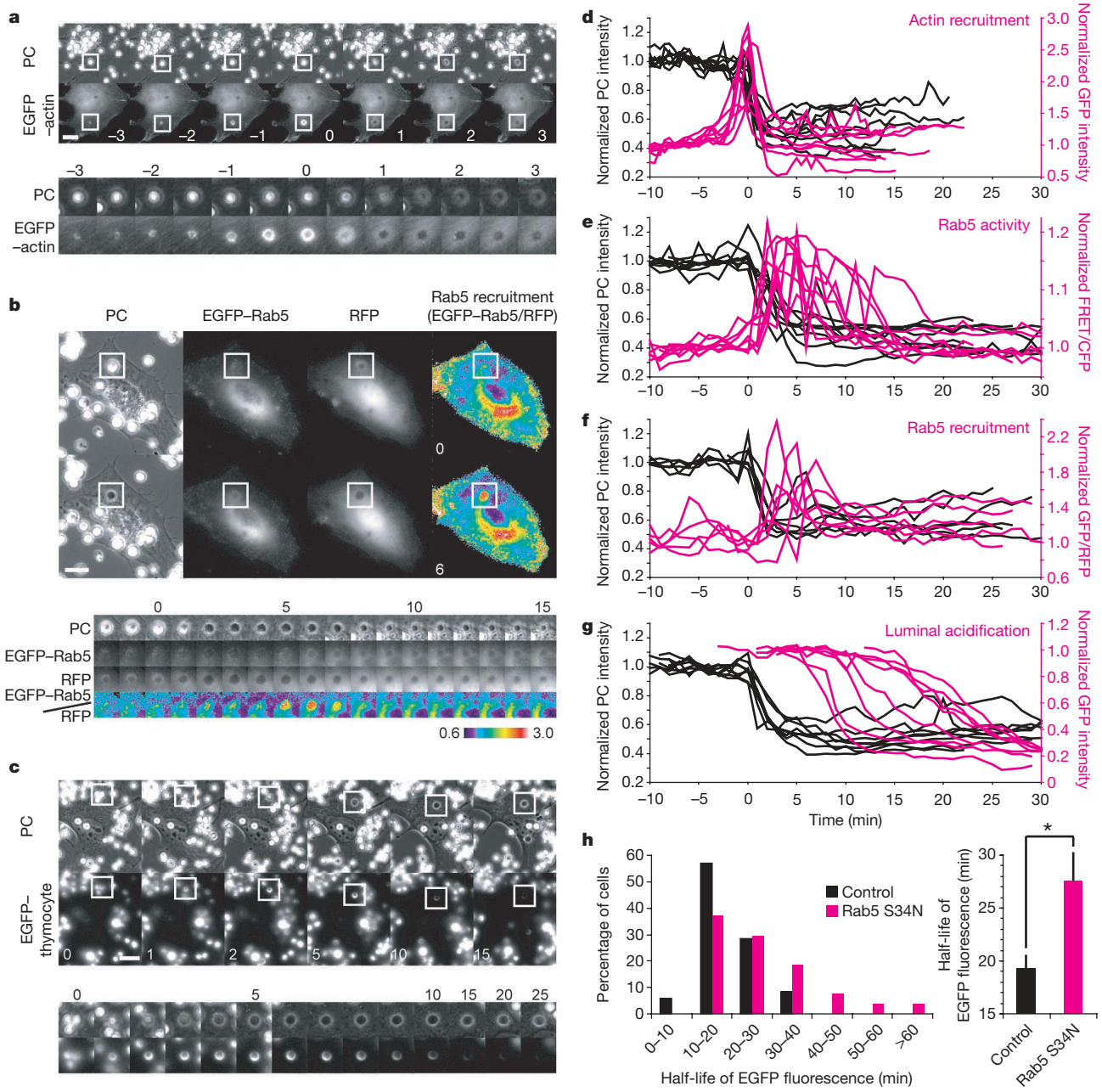


Figure 2 | Time sequence of actin and Rab5 accumulation, Rab5 activation, and breakdown of apoptotic thymocytes. **a**, **b**, Swiss3T3/integrin $\alpha_v\beta_3$ cells expressing EGFP–actin (**a**) or mEGFP–Rab5 and mRFP (**b**) were cultured together with apoptotic thymocytes. **c**, Swiss3T3/integrin $\alpha_v\beta_3$ cells were cultured together with apoptotic thymocytes derived from green mice. In **a** and **c**, the top panels show phase-contrast (PC) and fluorescent images at the indicated time points (min). In the fourth column of **b**, ratio images of mEGFP–Rab5/mRFP are shown in pseudocolour mode. In **a**–**c**, the bottom panels show the images of engulfment sites marked by white squares in the top panels. Scale bars, 20 μm. **d**–**g**, Black lines indicate PC intensity; magenta lines indicate the fluorescence intensity of EGFP–actin (**d**), the FRET/CFP ratio of Raichu-Rab5/PM (**e**), the mEGFP–Rab5/mRFP ratio (**f**), and the fluorescence intensity of EGFP-expressing thymocytes (**g**) in the circular regions (diameter 7.3 μm) containing engulfment sites ($n = 8$ in each panel). PC intensity (**d**–**g**), EGFP–actin intensity (**d**), the FRET/CFP ratio (**e**) and the mEGFP–Rab5/mRFP ratio (**f**) were normalized by dividing by their average values obtained before zero time (for EGFP–actin intensity (**d**) the average value between −10 and −5 min was used). EGFP-expressing thymocyte intensity (**g**) was normalized by dividing by the reference value, which was averaged over the period during which the EGFP signal had been stable before the quenching started. **h**, Control or Rab5-S34N-expressing Swiss3T3/integrin $\alpha_v\beta_3$ cells were cultured together with apoptotic thymocytes derived from green mice. The half-life of EGFP-expressing thymocyte fluorescence was divided into seven equal segments. Left: histogram plotting the percentage of cell numbers in each segment. Right: half-life of EGFP-expressing thymocyte fluorescence (mean and s.e.m.) (*$P < 0.01$; $n > 25$).

©2008 Nature Publishing Group

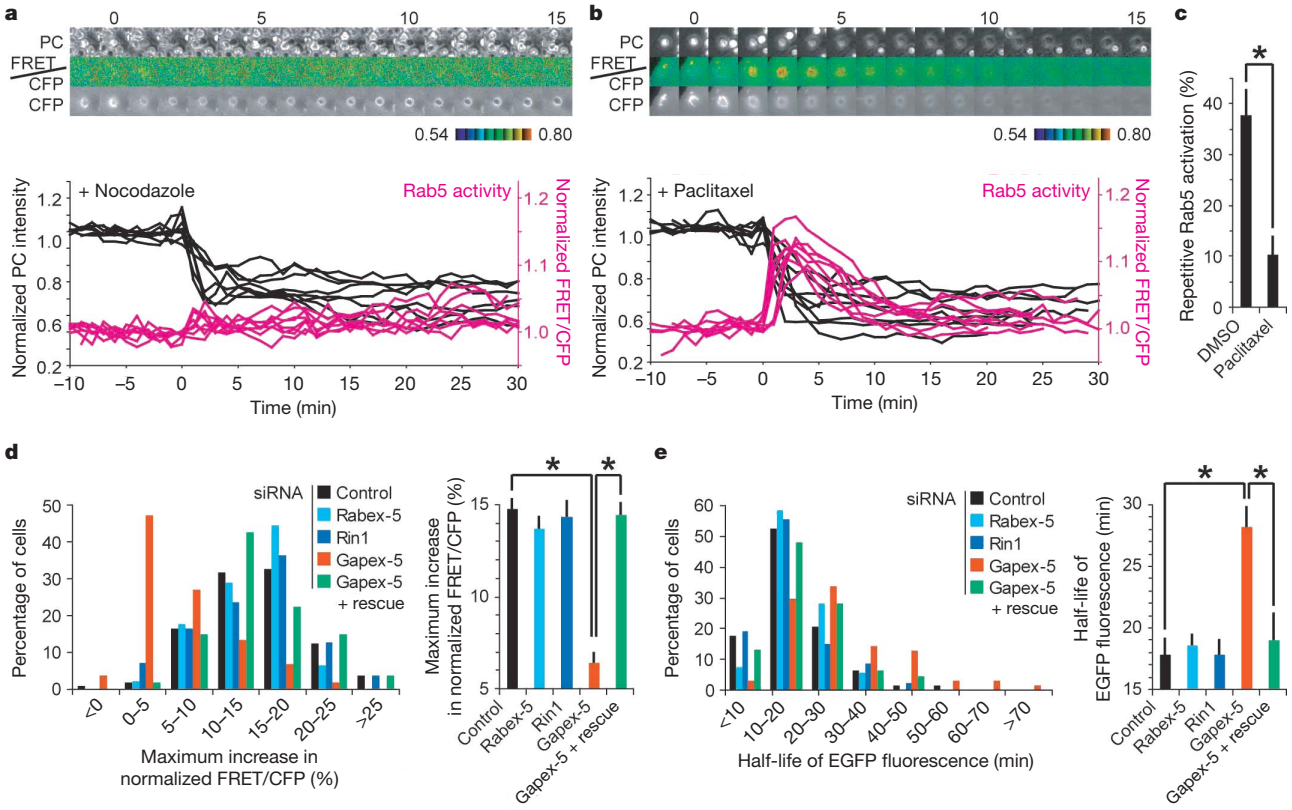


Figure 3 | Involvement of microtubules and Gapex-5 in Rab5 activation at engulfment sites. **a**, **b**, Swiss3T3/integrin $\alpha_v\beta_3$ cells expressing Raichu-Rab5/PM were pretreated with 0.1 μM nocodazole (**a**) or 10 nM paclitaxel (**b**) for 5 min and cultured together with apoptotic thymocytes. Time sequences in the top panels show the phase-contrast (PC), FRET/CFP ratio and CFP images at engulfment sites. The bottom panels show the time course of the normalized PC intensity (black) and FRET/CFP ratio (magenta). **c**, In dimethylsulphoxide (DMSO)-treated and paclitaxel-treated cells, at least 14 events of phagocytosis were observed in each experiment ($n = 4$). The percentage of cells showing reactivation (mean and s.e.m.) is shown ($^*P < 0.01$). **d**, Swiss3T3/integrin $\alpha_v\beta_3$ cells treated with siRNA for the indicated Rab5 GEFs were transfected with pRaichu-Rab5/PM and cultured together with apoptotic thymocytes. The maximum relative increase in Rab5 activity during phagocytosis was divided into seven segments. A histogram plotting the percentage of cell number in each segment is presented in the left panel. In some experiments we also used Swiss3T3/integrin $\alpha_v\beta_3$ cells stably expressing Gapex-5 resistant to Gapex-5 siRNA (+ rescue). The Rab5 peak activation (mean and s.e.m.) is shown in the right panel ($^*P < 0.001$; $n > 40$). **e**, Swiss3T3/integrin $\alpha_v\beta_3$ cells treated with siRNA for the indicated Rab5 GEFs were cultured together with apoptotic thymocytes derived from green mice. The half-life of EGFP-expressing thymocyte fluorescence was divided into eight segments. Left: histogram plotting the percentage of cell numbers in each segment. Right: half-life of EGFP-expressing thymocyte fluorescence (mean and s.e.m.) ($^*P < 0.001$; $n > 40$).

with Gapex-5 (Supplementary Fig. 8a, c). In agreement with the role predicted for EB1, EB1 knockdown significantly reduced Rab5 activation during phagocytosis (Supplementary Fig. 8b).

The present study shows that Gapex-5-mediated Rab5 activation follows disassembly of the actin coat surrounding phagosomes. Recalling that the Raichu-Rab5 probe monitors the balance of activity between GEFs and GAPs on the membrane, we envisage the following model for the regulation of Rab5 during phagocytosis: first, engulfment of apoptotic cells and accumulation of actin filaments around the nascent phagosome; second, actin disassembly during phagosome membrane closure; third, invasion of microtubules into the cell periphery (this probably happens because the actin filaments behave as physical barriers for microtubule extension[17]); fourth, delivery of Gapex-5 to phagosomes by EB1 on microtubules; fifth, a subsequent increase in the ratio of GEF activity to GAP activity on the phagosome membrane resulting not only in the activation of local Rab5 but also in the recruitment of Rab5 released from the Rab5–RabGDI complex in the cytosol. There might be a positive feedback mechanism for local amplification of Rab5 signal, as has been proposed for early endosomes[3,18]. This study shows that live-cell imaging with FRET probes enables us to pinpoint the activation and inactivation of Rab5 and thereby to understand its relationship to the other events of phagocytosis.

METHODS SUMMARY

C57BL/6 mice were purchased from Japan SLC. EGFP-transgenic mice[19] were a gift from M. Okabe. Swiss3T3 fibroblasts stably expressing integrin $\alpha_v\beta_3$ were established by infecting cells with a retrovirus carrying mouse integrin α_v and β_3 cDNAs[20] and maintained in DMEM medium supplemented with 10% fetal bovine serum. Retrovirus expressing EGFP–actin, mEGFP–Rab5, mCherry-Rab5-S34N or mCherry-Gapex-5 resistant to Gapex-5 siRNA was additionally inoculated into Swiss3T3/integrin $\alpha_v\beta_3$ cells. Apoptotic thymocytes were prepared as previously described[4]. In brief, thymocytes of 4–8-week-old C57BL/6 mice or EGFP-transgenic mice were treated for 5 h with 10 μM dexamethasone at 37 °C to induce apoptosis. Transient transfection of siRNA was used to decrease the level of specific proteins. For time-lapse FRET imaging, Swiss3T3/integrin $\alpha_v\beta_3$ cells expressing FRET probes were cultured in phenol-red-free DMEM/F12 medium supplemented with 10% fetal bovine serum, and loaded with 4×10^6 apoptotic thymocytes in the presence of 0.1 μg ml^{-1} MFG-E8. Cells were imaged with an IX81 inverted microscope (Olympus) equipped with a cooled charge-coupled device (CCD) camera (Retiga Exi, QImaging or Cool SNAP-HQ; Roper Scientific), an IX2-ZDC laser-based autofocusing system (Olympus) and an MD-XY30100T-Meta automatically programmable *XY* stage (Sigma KOKI). The following filters used for the dual-emission imaging studies were obtained from Omega Optical: an XF1071 (440AF21) excitation filter, an XF2034 (455DRLP) dichroic mirror and two emission filters (XF3075 (480AF30) for CFP and XF3079 (535AF26) for YFP). After background subtraction, FRET/CFP ratio images were created with MetaMorph software (Universal Imaging), and the images were used to represent FRET efficiency.

©2008 Nature Publishing Group

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

Received 9 January; accepted 19 February 2008.
Published online 2 April 2008.

1. Henson, P. M., Bratton, D. L. & Fadok, V. A. Apoptotic cell removal. *Curr. Biol.* **11,** R795–R805 (2001).
2. Savill, J., Dransfield, I., Gregory, C. & Haslett, C. A blast from the past: clearance of apoptotic cells regulates immune responses. *Nature Rev. Immunol.* **2,** 965–975 (2002).
3. Zerial, M. & McBride, H. Rab proteins as membrane organizers. *Nature Rev. Mol. Cell Biol.* **2,** 107–117 (2001).
4. Nakaya, M., Tanaka, M., Okabe, Y., Hanayama, R. & Nagata, S. Opposite effects of Rho family GTPases on engulfment of apoptotic cells by macrophages. *J. Biol. Chem.* **281,** 8836–8842 (2006).
5. Mochizuki, N. *et al.* Spatio-temporal images of growth-factor-induced activation of Ras and Rap1. *Nature* **411,** 1065–1068 (2001).
6. Takaya, A., Ohba, Y., Kurokawa, K. & Matsuda, M. RalA activation at nascent lamellipodia of epidermal growth factor-stimulated Cos7 cells and migrating Madin–Darby canine kidney cells. *Mol. Biol. Cell* **15,** 2549–2557 (2004).
7. Stenmark, H. *et al.* Inhibition of rab5 GTPase activity stimulates membrane fusion in endocytosis. *EMBO J.* **13,** 1287–1296 (1994).
8. Stenmark, H., Aasland, R., Toh, B. H. & D'Arrigo, A. Endosomal localization of the autoantigen EEA1 is mediated by a zinc-binding FYVE finger. *J. Biol. Chem.* **271,** 24048–24054 (1996).
9. Kanai, F. *et al.* The PX domains of p47phox and p40phox bind to lipid products of PI(3)K. *Nature Cell Biol.* **3,** 675–678 (2001).
10. Diakonova, M., Bokoch, G. & Swanson, J. A. Dynamics of cytoskeletal proteins during Fcγ receptor-mediated phagocytosis in macrophages. *Mol. Biol. Cell* **13,** 402–411 (2002).
11. Yoshizaki, H., Mochizuki, N., Gotoh, Y. & Matsuda, M. Akt-PDK1 complex mediates EGF-induced membrane protrusion through Ral activation. *Mol. Biol. Cell* **18,** 119–128 (2007).
12. Vieira, O. V. *et al.* Modulation of Rab5 and Rab7 recruitment to phagosomes by phosphatidylinositol 3-kinase. *Mol. Cell. Biol.* **23,** 2501–2514 (2003).
13. Niedergang, F. & Chavrier, P. Signaling and membrane dynamics during phagocytosis: many roads lead to the phagos(R)ome. *Curr. Opin. Cell Biol.* **16,** 422–428 (2004).
14. Alvarez-Dominguez, C., Barbieri, A. M., Beron, W., Wandinger-Ness, A. & Stahl, P. D. Phagocytosed live *Listeria monocytogenes* influences Rab5-regulated *in vitro* phagosome-endosome fusion. *J. Biol. Chem.* **271,** 13834–13843 (1996).
15. Duclos, S. *et al.* Rab5 regulates the kiss and run fusion between phagosome and endosomes and the acquisition of phagosomes leishmanicidal properties in RAW 264.7 macrophages. *J. Cell Sci.* **113,** 3531–3541 (2000).
16. Tirnauer, J. S. & Bierer, B. E. EB1 proteins regulate microtubule dynamics, cell polarity, and chromosome stability. *J. Cell Biol.* **149,** 761–766 (2000).
17. Zhou, F. Q., Waterman-Storer, C. M. & Cohan, C. S. Focal loss of actin bundles causes microtubule redistribution and growth cone turning. *J. Cell Biol.* **157,** 839–849 (2002).
18. Horiuchi, H. *et al.* A novel Rab5 GDP/GTP exchange factor complexed to Rabaptin-5 links nucleotide exchange to effector recruitment and function. *Cell* **90,** 1149–1159 (1997).
19. Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. & Nishimune, Y. 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* **407,** 313–319 (1997).
20. Hanayama, R. *et al.* Identification of a factor that links apoptotic cells to phagocytes. *Nature* **417,** 182–187 (2002).

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

Acknowledgements We thank N. Yoshida, N. Fujimoto, A. Nishiyama, K. Fukuhara, Y. Kasakawa for technical assistance; I. Lodhi and A. Saltiel for unpublished information on Gapex-5; Y. Kiyosue for advice on EB1; and members of the Matsuda laboratory for their input. This work was supported by grants from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Author Contributions M.K. and M.N. performed the experimental work and data analysis. T.N., S.N. and M.M. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.N. (tnakamr@path1.med.kyoto-u.ac.jp).

©2008 Nature Publishing Group

METHODS

Plasmids. pRaichu-Rab5, derived from the pCAGGS expression vector, encoded a probe, designated Raichu-Rab5, that comprised Venus (the brightest version of YFP)[21], the amino-terminal Rab5-binding domain of EEA1 (amino-acid residues 36–218), SECFP (a brighter version of CFP obtained from A. Miyawaki) and Rab5a (a gift from Y. Takai). In both Venus and SECFP, an A206K mutation was introduced to prevent dimerization between fluorescent proteins[22]. In pRaichu-Rab5/PM, the carboxy-terminal peptide of Rab5, QCCSN, was replaced with that of Ki-Ras, RKMSKDGKKKKKKSKTKCVIM. Mouse Rabex-5 cDNA was cloned from the mouse spleen cDNA library (a gift from N. Mochizuki) and point-mutated to resist siRNA. The cDNA for mCherry was a gift from R. Tsien. mEGFP–Rab5, mCherry-Rab5-S34N (ref. 23) and mCherry-Gapex-5 resistant to Gapex-5 siRNA were subcloned into pCX4bsr (ref. 24). Retroviral vector pCX4bleo-EGFP–actin was provided by M. Okada.

Mice, cells, reagents and antibodies. C57BL/6 mice were purchased from Japan SLC. EGFP-transgenic mice[19] were a gift from M. Okabe. Swiss3T3 fibroblasts stably expressing integrin $\alpha_v\beta_3$ were established by infecting cells with a retrovirus carrying mouse integrin α_v and β_3 cDNAs[20] and were maintained in DMEM medium supplemented with 10% fetal bovine serum. Retrovirus expressing EGFP–actin, mEGFP–Rab5, mCherry-Rab5-S34N or mCherry-Gapex-5 resistant to Gapex-5 siRNA was additionally inoculated into Swiss3T3/integrin $\alpha_v\beta_3$ cells. For imaging experiments, cells were plated on fibronectin-coated 35-mm glass-base dishes (Asahi Techno Glass).

Preparation of apoptotic thymocytes. Apoptotic thymocytes were prepared as described previously[4]. In brief, thymocytes of 4–8-week-old C57BL/6 mice or EGFP-transgenic mice were treated with 10 μM dexamethasone at 37 °C for 5 h to induce apoptosis.

RNA interference experiments. Swiss3T3/integrin $\alpha_v\beta_3$ cells were transfected with the desired siRNAs by using Lipofectamine RNAiMAX (Invitrogen). siRNA sequences for control, mouse Rabex-5, mouse Rin1 and mouse Gapex-5 were 5′-CACCUAAUCCGUGGUUCAA-3′, 5′-UUUAUAGAGACGCGUCAUGAUGUGC-3′, 5′-UUAUACAUUUGCUUCACACCUAAGC-3′ and 5′-GGAGUGUGAUGAAUCGGAUCUUCAA-3′, respectively. After 24 h, the cells were treated with trypsin and plated on glass-base dishes. For FRET imaging, the cells were then transfected with pRaichu-Rab5/PM. Thereafter, the cells were incubated for at least 72 h after siRNA transfection before imaging experiments.

Time-lapse FRET imaging. Swiss3T3/integrin $\alpha_v\beta_3$ cells expressing FRET probes were cultured in phenol-red-free DMEM/F12 medium supplemented with 10% fetal bovine serum, and loaded with 4×10^6 apoptotic thymocytes in the presence of 0.1 μg ml^{-1} MFG-E8. Cells were imaged with an IX81 inverted microscope (Olympus) equipped with a cooled CCD camera (Retiga Exi, QImaging or Cool SNAP-HQ; Roper Scientific), an IX2-ZDC laser-based autofocusing system (Olympus) and an MD-XY30100T-Meta automatically programmable *XY* stage (Sigma KOKI). The following filters used for the dual-emission imaging studies were obtained from Omega Optical: an XF1071 (440AF21) excitation filter, an XF2034 (455DRLP) dichroic mirror, and two emission filters (XF3075 (480AF30) for CFP and XF3079 (535AF26) for YFP). After background subtraction, FRET/CFP ratio images were created with MetaMorph software (Universal Imaging), and the images were used to represent FRET efficiency.

21. Nagai, T. *et al.* A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature Biotechnol.* **20,** 87–90 (2002).
22. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science* **296,** 913–916 (2002).
23. Imamura, H. *et al.* Rho and Rab small G proteins coordinately reorganiza stress fibers and focal adhesions in MDCK cells. *Mol. Biol. Cell* **9,** 2561–2575 (1998).
24. Akagi, T., Sasai, K. & Hanafusa, H. Refractory nature of normal human diploid fibroblasts with respect to oncogene-mediated transformation. *Proc. Natl Acad. Sci. USA* **100,** 13567–13572 (2003).

©2008 Nature Publishing Group

LETTERS

Chromatin decouples promoter threshold from dynamic range

Felix H. Lam[1,2], David J. Steger[1]† & Erin K. O'Shea[1]

Chromatin influences gene expression by restricting access of DNA binding proteins to their cognate sites in the genome[1–3]. Large-scale characterization of nucleosome positioning in *Saccharomyces cerevisiae* has revealed a stereotyped promoter organization in which a nucleosome-free region (NFR) is present within several hundred base pairs upstream of the translation start site[4,5]. Many transcription factors bind within NFRs and nucleate chromatin remodelling events which then expose other *cis*-regulatory elements[6–9]. However, it is not clear how transcription-factor binding and chromatin influence quantitative attributes of gene expression. Here we show that nucleosomes function largely to decouple the threshold of induction from dynamic range. With a series of variants of one promoter, we establish that the affinity of exposed binding sites is a primary determinant of the level of physiological stimulus necessary for substantial gene activation, and sites located within nucleosomal regions serve to scale expression once chromatin is remodelled. Furthermore, we find that the *S. cerevisiae* phosphate response (*PHO*) pathway exploits these promoter designs to tailor gene expression to different environmental phosphate levels. Our results suggest that the interplay of chromatin and binding-site affinity provides a mechanism for fine-tuning responses to the same cellular state. Moreover, these findings may be a starting point for more detailed models of eukaryotic transcriptional control.

When cells sense changes in environmental inorganic phosphate (P_i), the activity of the transcription factor Pho4 is modulated by phosphorylation[10]: Pho4 is phosphorylated on four sites, cytoplasmic and inactive when cells are grown in P_i-rich medium; it is phosphorylated selectively on one site and localized to the nucleus in intermediate P_i (about 100 μM) conditions; and it is unphosphorylated, nuclear and fully active in P_i starvation[11,12]. The co-activator Pho2 interacts with unphosphorylated Pho4 and is required for induction of many *PHO* genes[13,14]; however, it is not thought to be regulated in response to P_i availability[12]. Despite being controlled by the same activators, the target gene *PHO5* is expressed at a low level in intermediate P_i conditions, whereas *PHO84* is significantly induced[14]. Although both promoters contain a combination of high- and low-affinity Pho4-binding sites[15–17], more Pho4 is recruited to *PHO84* in intermediate P_i conditions than to *PHO5*[14]. We hypothesized that chromatin may influence gene expression by differentially regulating the accessibility of Pho4 sites in the *PHO5* and *PHO84* promoters.

To test this hypothesis, we constructed variants of the *PHO5* promoter controlling transcription of a green fluorescent protein (GFP) reporter gene (Fig. 1a). The *PHO5* promoter contains five positioned nucleosomes (numbered −5 to −1), a low-affinity Pho4 site (CACGTTt) and a Pho2 site in the NFR, and a high-affinity Pho4 site (CACGTGg) and distal Pho2 sites occluded under nucleosome −2 (refs 16–20). The difference in affinity between these two Pho4 sites is estimated to be about 3-fold by electrophoretic mobility shift assay[21], and about 13-fold by a recent high-throughput fluorescence-based assay[22]. The Pho4 site in the NFR is required for transcriptional induction: Pho4 binds this site and recruits ATP-dependent chromatin remodelling complexes that displace nucleosomes to expose additional *cis*-regulatory elements[23,24]. We manipulated the accessibility of Pho4 sites by relocating motifs into and/or out of nucleosomal regions and verified that our genetic perturbations have minimal impact on *PHO5* chromatin structure (Supplementary Figs 1 and 2)[5].

To assess the effects of these alterations quantitatively, we assayed the steady-state expression of *PHO5* promoter variants grown in different concentrations of P_i. We observed surprisingly stereotyped behaviour when the output profile for each variant was normalized to its maximum expression. When grown in intermediate (10–100 μM) P_i concentrations, variants containing a low-affinity motif in the NFR induce slightly (5 ± 4% of maximum levels), whereas variants with an exposed high-affinity motif induce significantly (37 ± 2% of maximum levels) (Fig. 1b). This response is a function of the accessible Pho4 site and is independent of the number or placement of occluded Pho4 motifs: similar behaviour is observed when the high-affinity Pho4 site normally occluded by nucleosome −2 is transplanted or added under nucleosome −3 (variants L2, L3), or when there is a single accessible Pho4 site (variants L1, H1).

On the basis of our steady-state measurements, we expect promoter variants with an exposed high-affinity site to induce when cells reach the intermediate P_i range, whereas we expect promoters with an exposed low-affinity site to remain essentially repressed until near the zero P_i state. Indeed, we observe a difference in induction kinetics when the promoters are scaled to their particular dynamic range (Fig. 1c): the time-to-half-maximal induction of the high-affinity class is 146–157 min, whereas that of the low-affinity class is 186–197 min (Supplementary Table 1, interpolated values). The average 40 min lag observed in the low-affinity variants is not solely a result of inherently slower promoter activation; the kinetic difference is reduced by disabling the vacuolar phosphate buffer which accelerates the transition to the zero P_i state (Supplementary Fig. 3 and Supplementary Table 1)[25]. Rather, the lag is a consequence of how the affinity of the exposed site mediates a different response to the same upstream Pho4 signal in intermediate P_i conditions. Thus, the affinity of the transcription-factor binding site(s) in the NFR also influences the apparent kinetics of gene expression.

To assess the generality of this promoter design principle, we asked if binding-site coordinates and affinities, in combination with nucleosome positions, would allow quantitative prediction of promoter

[1]Howard Hughes Medical Institute, Departments of Molecular and Cellular Biology, and Chemistry and Chemical Biology, Faculty of Arts and Sciences Center for Systems Biology, Harvard University, 7 Divinity Avenue, Bauer 307, Cambridge, Massachusetts 02138, USA. [2]Graduate Group in Biophysics, University of California, San Francisco, San Francisco, California 94158, USA. †Present address: Division of Endocrinology, Diabetes and Metabolism, 611 Clinical Research Building, University of Pennsylvania, 415 Curie Boulevard, Philadelphia, Pennsylvania 19104, USA.

©2008 Nature Publishing Group

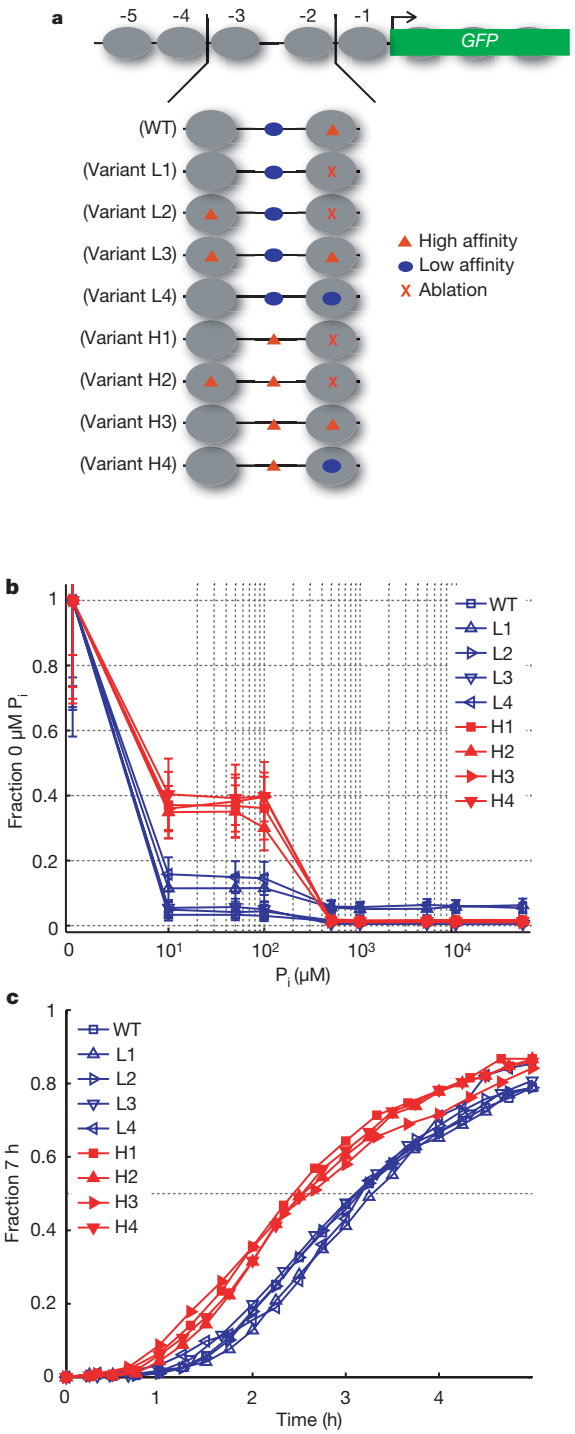


Figure 1 | *PHO5* promoter variants and quantitative expression behaviour. **a**, Schematic of all *PHO5* promoter variants controlling expression of *yeGFP1*. Large grey ovals represent nucleosomes, red triangles the high-affinity Pho4 motif (CACGTGg), blue ovals the low-affinity Pho4 motif (CACGTTt), and X the motif ablations. **b**, Physiological transcriptional response to extracellular inorganic phosphate (P_i) measured by flow cytometry. Data points represent median steady-state expression levels normalized to the median observed in P_i starvation. Error bars represent interquartile ranges, which were observed to encompass the medians of at least three independent measurements. **c**, Induction kinetics in P_i starvation. Data points represent median fluorescence levels scaled between the promoter-specific expression minimum at 0 h and maximum at 7 h. For **b** and **c**, red traces designate variants with an exposed high-affinity site, and blue traces variants with an exposed low-affinity site.

response profiles and kinetics. As a test set, we selected and mapped promoter chromatin (Supplementary Fig. 1) for *PHO84* and five additional Pho4 targets that are highly induced by P_i starvation and have a combination of high- and low-affinity motifs for Pho4 (Supplementary Figs 4 and 5). *PHO89* and *PHM6* exhibit nucleosomal arrays that, like *PHO5*, expose only a low-affinity Pho4 motif(s) in the repressed state (Fig. 2a). The remaining four genes (*PHO84*, *PHO8*, *PHM2* and *PHM4*) have promoters containing at least one accessible high-affinity motif (Fig. 2b). Although many of these promoters likely contain Pho2 sites (which are difficult to recognize by sequence owing to AT-rich composition and degeneracy), in intermediate and high P_i conditions these sites may not have a large influence on Pho4 recruitment because Pho4 is phosphorylated and unable to interact efficiently with Pho2 (ref. 12).

We expected *PHO89* and *PHM6* to behave similarly to *PHO5*, exhibiting low-level expression in intermediate P_i and slow induction in P_i starvation, and promoters containing at least one exposed high-affinity site (*PHO84*, *PHO8*, *PHM2* and *PHM4*) to exhibit substantial expression and rapid induction. We created GFP transcriptional reporter strains and, consistent with our expectations, observed that *PHO89* and *PHM6* induce minimally in intermediate P_i conditions ($3 \pm 1\%$ of maximum) whereas *PHO84*, *PHO8*, *PHM2* and *PHM4* induce significantly ($34 \pm 5\%$ of maximum) (Fig. 2c). Additionally, when starved for P_i, we observed that genes with an exposed high-affinity motif reach half-maximal induction in 121–156 min compared with 198–222 min for the low-affinity class (Fig. 2d and Supplementary Table 2, interpolated values). Therefore, the affinity of the accessible Pho4 site is sufficient to define a promoter threshold from which steady-state and kinetic behaviour follow.

Nucleosome mapping and sequence analysis reveal that about 50% of the evolutionarily conserved Pho4 sites in our selection of *PHO* promoters are inaccessible in the repressed state and become exposed when chromatin is remodelled in response to P_i starvation (Fig. 2a, b and Supplementary Fig. 5). To assess the influence of these sites on quantitative gene expression in a controlled sequence background, we analysed the maximum induction levels of the *PHO5* promoter variants (Fig. 3 and Supplementary Table 3). Consistent with the role of the accessible site in the nucleation of remodelling, variants lacking exposed motifs (A1–A4) are uninducible or severely crippled. Otherwise, maximum transcriptional output differs by a factor of about four to seven and correlates with the number, affinity and placement of Pho4 sites, irrespective of their accessibility in the repressed state. Thus, chromatin enables significant decoupling of the determinants of promoter threshold from determinants of expression capacity.

Collectively, our observations are consistent with a model whereby nucleosomes compete with Pho4 for binding DNA, and nucleosome occupancy is dynamic and determined by a balance of assembly and disassembly activities[26–28]. The chromatin maps and range of promoter outputs observed in different P_i conditions (Figs 1b, 2a–c, 3, Supplementary Fig. 6 and Supplementary Table 3) likely reflect the different extents of chromatin remodelling resulting from a competition between Pho4 binding and nucleosome re-assembly. Even the A4 *PHO5* promoter variant containing two high-affinity sites occluded by nucleosomes can be induced to a reasonable level (Fig. 3). This observation is consistent with the model that chromatin is dynamic, and that multiple buried high-affinity sites are sufficient for Pho4 to compete successfully with nucleosomes.

At the sub-maximal Pho4 activity (that is, nuclear concentration and/or phosphorylation state[14,29]) associated with intermediate P_i conditions, the occupancy of Pho4 at an exposed high-affinity site may be sufficient to nucleate substantial chromatin remodelling such that regulatory sequences that were previously nucleosomal now become more accessible, allowing for appreciable transcription. However, promoters with an exposed low-affinity motif likely do not achieve similar Pho4 occupancy until P_i starvation. Indeed, *in vivo* Pho4 occupancy is substantially higher in intermediate P_i

©2008 Nature Publishing Group

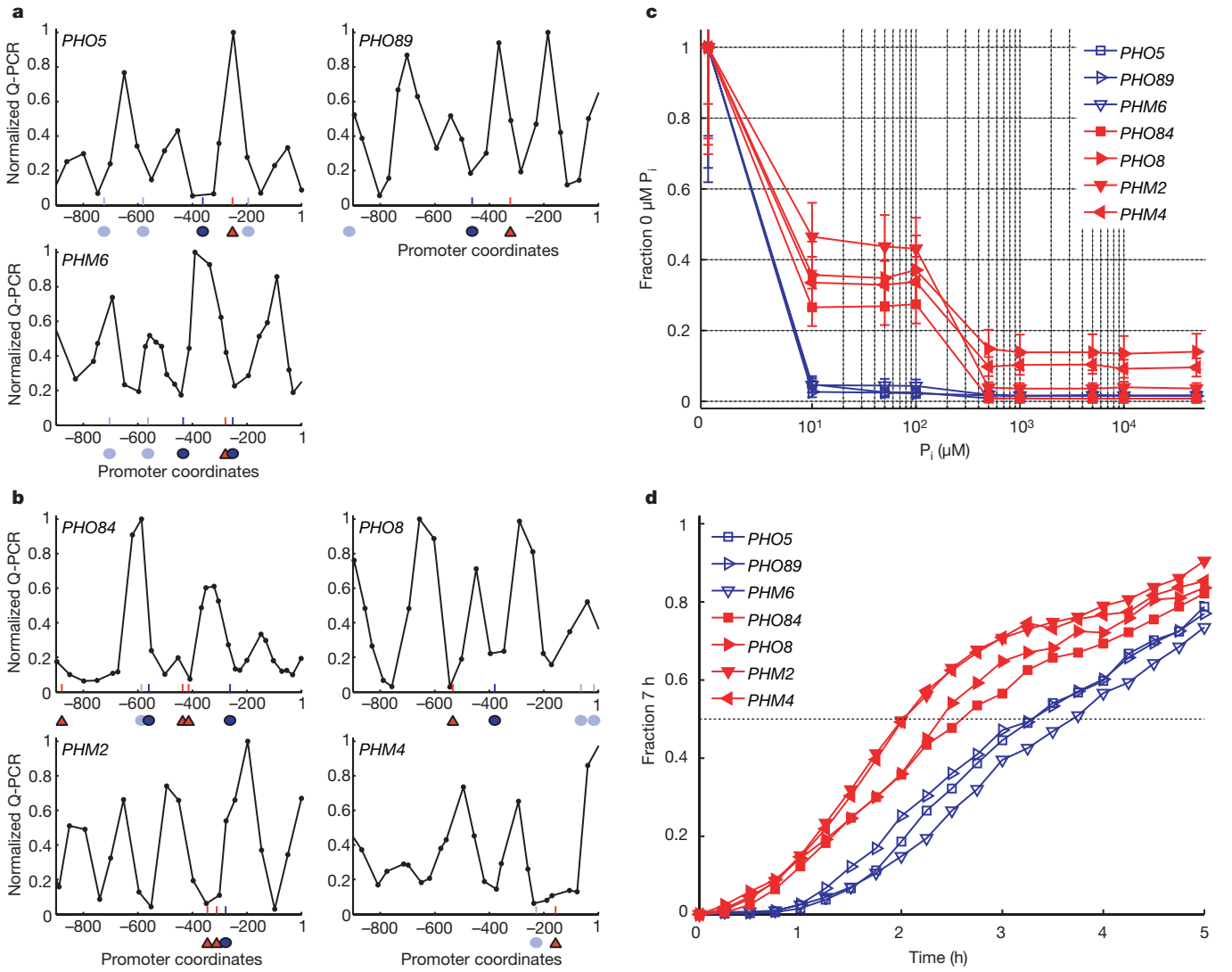


Figure 2 | Promoter architecture and quantitative expression behaviour of representative *PHO* genes. **a**, **b**, Promoter architecture schematized by superimposing nucleosome positions measured in repressing (10 mM P_i) conditions onto Pho4-binding sites identified through bioinformatic analysis (Supplementary Fig. 4). Red triangles represent evolutionarily conserved high-affinity motifs (CACGTG consensus), dark-blue ovals represent evolutionarily conserved low-affinity motifs (deviations from the high-affinity motif), light-blue ovals represent low-affinity motifs that are not evolutionarily conserved (Supplementary Fig. 5), and the *x* axis units reference promoter coordinates with respect to translation start (ATG = 1). In **a** are *PHO* promoters with an accessible low-affinity Pho4 site; in **b** are promoters with at least one accessible high-affinity Pho4 site. **c**, Steady-state transcriptional response of *PHO* target genes to P_i. Error bars are interquartile ranges (see Fig. 1b). **d**, Induction kinetics in P_i starvation.

conditions at *PHO* promoters containing accessible high-affinity sites than at promoters with exposed low-affinity sites (Fig. 4a). Tuning Pho4 activity such that only promoters with exposed high-affinity sites are substantially induced in intermediate P_i is physiologically relevant: the products of first-response genes like *PHO84* and *PHM4* allow the cell to take up environmental P_i and mobilize internal reserves before a second-order energy commitment is made in starvation conditions to upregulate *PHO5* and other P_i scavenging components[14]. Chromatin renders the signal required for promoter induction sensitive to fewer sites (that is, those exposed), while allowing nucleosomal sites to influence maximum transcriptional output once the promoter becomes open (Fig. 4b). The threshold of activation and dynamic range thus become functionally decoupled, which would not be the case if all binding sites were exposed.

These observations indicate that the interplay between chromatin structure and binding-site affinity may allow different eukaryotic promoters regulated by the same factor to interpret and respond to

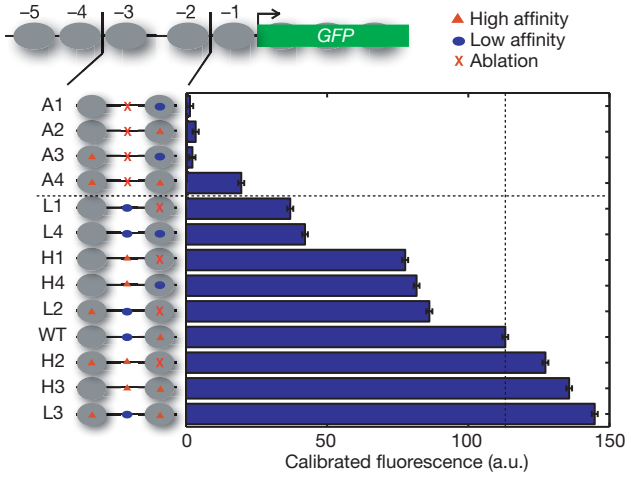


Figure 3 | Maximum expression of *PHO5* promoter variants. Maximal induction levels of *PHO5* promoter variants measured from strains containing a deletion of the *PHO80* gene. The dotted vertical line references the expression output of the wild-type promoter. Data points represent mean ± s.d. from triplicate measurements; a.u., arbitrary units.

©2008 Nature Publishing Group

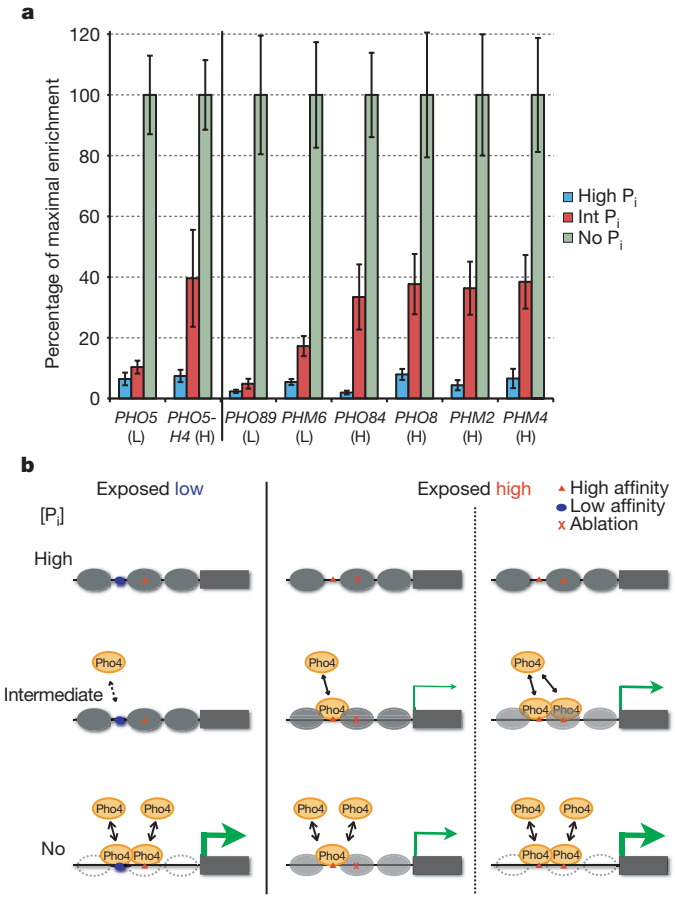


Figure 4 | Pho4 binding *in vivo* to *PHO* promoters, and model of threshold-dynamic range decoupling. a, Chromatin immunoprecipitation of Pho4 to *PHO* target genes and the H4 *PHO5* promoter variant (see Fig. 1a) shows differential occupancy in intermediate P_i conditions at promoters with an exposed high- (H) versus low (L)-affinity site. Data points represent mean ± s.d. from at least three independent experiments. **b**, Schematic depicting a possible mechanism that decouples promoter induction threshold from dynamic range. In high P_i (top row), nucleosomes are fully assembled; in intermediate P_i (middle row), substantial Pho4 occupancy occurs only at promoters with exposed high-affinity sites, resulting in chromatin remodelling and transcription commensurate to the total Pho4 recruited; and in P_i starvation (bottom row), saturating Pho4 activity results in remodelling and maximum expression at all promoters. Nucleosome occupancy is indicated by the opacity of grey ovals, Pho4 by yellow ovals, and the amount of transcription by the thickness of the green arrows. See text for description.

cellular signals uniquely. One implication of this finding is that future models of gene expression will require knowledge both of motifs and nucleosome positions. Additionally, the behaviour of *PHO* genes—largely a function of a single transcription factor—will likely prove simple compared with combinatorially controlled genes. But even with the nonlinearities created by multiple transcription factors, it will be fascinating to ascertain if similar architectural principles of promoter chromatin are conserved in other pathways and organisms as a mechanism for quantitatively configuring gene expression.

METHODS SUMMARY

Promoter variants. Pho4-binding sites or motif ablations were introduced into haploid *S. cerevisiae* at the chromosomal *PHO5* locus. Additionally, the *PHO5* coding region was completely substituted with the yeast-enhanced green fluorescent protein (*yeGFP1*)[30] reporter using targeted gene replacement. To assess maximal induction levels, a derivative series of strains was created harbouring both the *PHO5* promoter mutations and a deletion of the *PHO80* gene.

Heterozygous transcriptional reporters. For representative *PHO* genes (except *PHO84*, where *PHO84pr–yeGFP1* was placed ectopically), one allele in diploid *S. cerevisiae* was replaced with *yeGFP1* at the chromosomal locus.

Growth conditions. For dose–response profiles, strains were grown in 0–50 mM extracellular P_i for 18 h to reach steady-state expression. For induction kinetics, strains grown overnight in 10 mM P_i were subjected to time-course measurements after transfer into 0 mM P_i. For measurements of maximum induction levels, *pho80Δ* strains were grown 18 h in 10 mM P_i to optical density (OD_{600}) 0.05–0.1.

Flow cytometry. Fluorescence distributions were read using a Becton, Dickinson LSR II with 488 nm laser and calibrated to fluorescent polystyrene microspheres to control for experimental variation. Autofluorescence was assessed in all experiments by inclusion of a strain lacking yeGFP1. Data analysis was done using MATLAB.

Promoter chromatin maps. Mononucleosome-length DNA purified from micrococcal nuclease treated spheroplasts was assayed by quantitative polymerase chain reaction using primer sets that tile approximately 1 kilobase of each promoter region.

Chromatin immunoprecipitation. Cells were grown in high (10 mM), intermediate (100 μM) or no (0 μM) P_i medium for 2.5 h, and fixed in 1% formaldehyde. Pho4 fold enrichment over *POL1* in segments including the nucleosome-free region of each promoter was measured by quantitative polymerase chain reaction, and further re-normalized to the maximal fold enrichment observed in no P_i.

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

Received 24 October 2007; accepted 22 February 2008.
Published online 16 April 2008.

1. Kornberg, R. D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285–294 (1999).
2. Narlikar, G. J., Fan, H. Y. & Kingston, R. E. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108**, 475–487 (2002).
3. Khorasanizadeh, S. The nucleosome: from genomic organization to genomic regulation. *Cell* **116**, 259–272 (2004).
4. Yuan, G. C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
5. Sekinger, E. A., Moqtaderi, Z. & Struhl, K. Intrinsic histone–DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* **18**, 735–748 (2005).
6. Buck, M. J. & Lieb, J. D. A chromatin-mediated mechanism for specification of conditional transcription factor targets. *Nature Genet.* **38**, 1446–1451 (2006).
7. Liu, X., Lee, C. K., Granek, J. A., Clarke, N. D. & Lieb, J. D. Whole-genome comparison of Leu3 binding *in vitro* and *in vivo* reveals the importance of nucleosome occupancy in target site selection. *Genome Res.* **16**, 1517–1528 (2006).
8. Workman, J. L. Nucleosome displacement in transcription. *Genes Dev.* **20**, 2009–2017 (2006).
9. Miller, J. A. & Widom, J. Collaborative competition mechanism for gene activation *in vivo*. *Mol. Cell. Biol.* **23**, 1623–1632 (2003).
10. Kaffman, A., Herskowitz, I., Tjian, R. & O'Shea, E. K. Phosphorylation of the transcription factor PHO4 by a cyclin-CDK complex, PHO80–PHO85. *Science* **263**, 1153–1156 (1994).
11. O'Neill, E. M., Kaffman, A., Jolly, E. R. & O'Shea, E. K. Regulation of PHO4 nuclear localization by the PHO80–PHO85 cyclin–CDK complex. *Science* **271**, 209–212 (1996).
12. Komeili, A. & O'Shea, E. K. Roles of phosphorylation sites in regulating activity of the transcription factor Pho4. *Science* **284**, 977–980 (1999).
13. Magbanua, J. P., Fujisawa, K., Ogawa, N. & Oshima, Y. The homeodomain protein Pho2p binds at an A/T-rich segment flanking the binding site of the basic-helix–loop-helix protein Pho4p in the yeast PHO promoters. *Yeast* **13**, 1299–1308 (1997).
14. Springer, M., Wykoff, D. D., Miller, N. & O'Shea, E. K. Partially phosphorylated Pho4 activates transcription of a subset of phosphate-responsive genes. *PLoS Biol.* **1**, E28 (2003).
15. Ogawa, N. *et al.* Structure and distribution of specific *cis*-elements for transcriptional regulation of PHO84 in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **249**, 406–416 (1995).
16. Vogel, K., Horz, W. & Hinnen, A. The two positively acting regulatory proteins PHO2 and PHO4 physically interact with PHO5 upstream activation regions. *Mol. Cell. Biol.* **9**, 2050–2057 (1989).
17. Ogawa, N., Hayashi, N., Saito, H., Noguchi, K. & Yamashita, Y. in *Phosphate in Microorganisms: Cellular and Molecular Biology* (eds Torriani-Gorini, A., Yagil, E. & Silver, S.) 56–62 (American Society for Microbiology, Washington DC, 1994).
18. Almer, A., Rudolph, H., Hinnen, A. & Horz, W. Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. *EMBO J.* **5**, 2689–2696 (1986).
19. Barbaric, S., Munsterkotter, M., Goding, C. & Horz, W. Cooperative Pho2-Pho4 interactions at the PHO5 promoter are critical for binding of Pho4 to UASp1 and for efficient transactivation by Pho4 at UASp2. *Mol. Cell. Biol.* **18**, 2629–2639 (1998).

©2008 Nature Publishing Group

20. Korber, P., Luckenbach, T., Blaschke, D. & Horz, W. Evidence for histone eviction in trans upon induction of the yeast PHO5 promoter. *Mol. Cell. Biol.* **24,** 10965–10974 (2004).
21. Fisher, F. & Goding, C. R. Single amino acid substitutions alter helix–loop–helix protein specificity for bases flanking the core CANNTG motif. *EMBO J.* **11,** 4103–4109 (1992).
22. Maerkl, S. J. & Quake, S. R. A systems approach to measuring the binding energy landscapes of transcription factors. *Science* **315,** 233–237 (2007).
23. Svaren, J., Schmitz, J. & Horz, W. The transactivation domain of Pho4 is required for nucleosome disruption at the PHO5 promoter. *EMBO J.* **13,** 4856–4862 (1994).
24. Venter, U., Svaren, J., Schmitz, J., Schmid, A. & Horz, W. A nucleosome precludes binding of the transcription factor Pho4 *in vivo* to a critical target site in the PHO5 promoter. *EMBO J.* **13,** 4848–4855 (1994).
25. Thomas, M. R. & O'Shea, E. K. An intracellular phosphate buffer filters transient fluctuations in extracellular phosphate levels. *Proc. Natl Acad. Sci. USA* **102,** 9565–9570 (2005).
26. Boeger, H., Griesenbeck, J., Strattan, J. S. & Kornberg, R. D. Nucleosomes unfold completely at a transcriptionally active promoter. *Mol. Cell* **11,** 1587–1598 (2003).
27. Adkins, M. W., Howar, S. R. & Tyler, J. K. Chromatin disassembly mediated by the histone chaperone Asf1 is essential for transcriptional activation of the yeast PHO5 and PHO8 genes. *Mol. Cell* **14,** 657–666 (2004).
28. Adkins, M. W. & Tyler, J. K. Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. *Mol. Cell* **21,** 405–416 (2006).
29. Dhasarathy, A. & Kladde, M. P. Promoter occupancy is a major determinant of chromatin remodeling enzyme requirements. *Mol. Cell. Biol.* **25,** 2698–2707 (2005).
30. Cormack, B. P. *et al.* Yeast-enhanced green fluorescent protein (yEGFP)a reporter of gene expression in *Candida albicans*. *Microbiology* **143,** 303–311 (1997).

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

Acknowledgements We thank: S. J. Maerkl and S. R. Quake for sharing Pho4-binding data; J. S. Weissman, H. Li, R. Losick, B. Stern and present and former members of the O'Shea laboratory for discussion and commentary on the manuscript; B. S. Margolin for strain EY1995 and assistance with cell sorting; and B. E. Shakhnovich for assistance with motif analysis. Support was provided by the National Institutes of Health, the Howard Hughes Medical Institute, the David and Lucile Packard Foundation (E.K.O.), and the Burroughs Wellcome Fund (F.H.L.).

Author Contributions F.H.L., D.J.S. and E.K.O. designed the experiments. D.J.S. made initial promoter variants and preliminary measurements of expression behaviour and nucleosome positions. F.H.L. conducted further strain construction, kinetic and steady-state expression measurements, flow cytometry, nucleosome mapping, chromatin immunoprecipitation and computational analysis. F.H.L. and E.K.O. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.K.O. (erin_oshea@harvard.edu).

©2008 Nature Publishing Group

METHODS

Identification of Pho4 motifs. Analysis by position-specific scoring matrix was used to identify systematically Pho4-binding sites among our selection of chromatin-mapped *PHO* promoters. From the 256 $\Delta\Delta G_{full}$ values measured by Maerkl and Quake of Pho4 binding to their NNNNGTG DNA oligonucleotide library[22], the corresponding probabilities of binding were computed to produce an initial position-specific scoring matrix for the four base positions representing the E-box 5′ half-site and flanking base (that is, permuted positions only). To reflect the homodimeric nature of Pho4 binding to DNA[31] and the optimality of palindromic motifs[22], these base probabilities were reverse-complemented to fill the 3′ half-site plus flanking position to produce a full octameric position-specific scoring matrix.

Based on the resulting binding-probability profiles of *PHO5*, *PHO84* and *PHO8* (Supplementary Fig. 4a, d, e, respectively), the Pho4-promoter interactions of which have been characterized by *in vitro* DNase I protection assays[15,17,24,32], a noise threshold for likely Pho4 regulatory elements was determined by selecting the weakest probability matching an experimentally verified footprint. To distinguish high- from low-affinity motifs, a second threshold was established by comparing the binding profiles of *PHO5* and *PHO84* with *in vitro* competition footprinting experiments from the literature[15,17], and selecting the most stringent probability necessary to recapitulate validated high-affinity sites. Finally, putative Pho4 sites were plotted on multiple alignments of *Saccharomyces* orthologues to assess evolutionary conservation (Supplementary Fig. 5) on the assumption that conserved sites indicate a higher likelihood of function[33,34].

Flow cytometry analysis. To control for day-to-day instrument variation, a five-point calibration sample of fluorescent polystyrene microspheres was read at the start and end of each experiment. A weighted least-squares fit to the calibration points established a standard curve spanning an approximate 1,000-fold dynamic range that allowed for accurate quantitative comparison across different experiments. Signal area data from the fluorescein isothiocyanate channel (FITC-A) for 10,000 cells were collected from all samples and interpolated off these standard curves before further analysis. Any reference to 'calibrated fluorescence' in this study refers to these standardized units.

Although samples were sonicated before flow cytometry to reduce cell aggregation, simple rectangular gating in the forward-scatter and side-scatter channels was performed to further reduce variability. An automated approach was implemented whereby data points outside the 0.1 and 0.85 quantiles in either forward scatter and side scatter were eliminated. This worked well empirically to eliminate debris and any remaining cell aggregates.

Although autofluorescence was minimized in all synthetic media experiments by the addition of supplementary adenine and tryptophan, a wild-type strain lacking yeGFP1 was included in all experiments to assess the level of fluorescence specific to gene expression. Instead of doing subtraction with fitted distribution parameters estimated from the yeGFP1$^-$ and transcriptional reporter samples, a randomly permuted vector subtraction of autofluorescence from the transcriptional reporter distributions was performed instead. This numerical approach was deemed more unbiased as assumptions of normality or distribution symmetry would not have to be made, and may have been particularly suited in situations where population behaviour was highly skewed (for example, during induction in P_i starvation conditions).

Two approaches were taken to estimate times of half-maximal induction (Supplementary Tables 1 and 2). The first was a simple linear interpolation of the time needed to reach a level halfway between the median expression observed at 0 and 7 h (Figs 1c and 2d). However, as interpolated half-times are highly sensitive to maximal induction levels, a second approach was taken for comparison where the unscaled kinetic data were subjected to a nonlinear least squares fit of a Hill equation with three freely varying parameters: n (Hill coefficient), Y_{max} (asymptotic maximal value) and T_{half} (half maximal time). Because Y_{max} is a freely varying parameter, T_{half} is not constrained to an arbitrary maximum and is free to adopt the value that best approximates the continuous underlying behaviour. In general, the fits gave normally distributed residuals centred around zero and low root mean squared errors (fit standard errors), indicating that the Hill model described the kinetic data well.

31. Shimizu, T. *et al.* Crystal structure of PHO4 bHLH domain–DNA complex: flanking base recognition. *EMBO J.* **16,** 4689–4697 (1997).
32. Barbaric, S., Fascher, K. D. & Horz, W. Activation of the weakly regulated PHO8 promoter in *S. cerevisiae*: chromatin transition and binding sites for the positive regulatory protein PHO4. *Nucleic Acids Res.* **20,** 1031–1038 (1992).
33. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423,** 241–254 (2003).
34. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12,** 996–1006 (2002).

©2008 Nature Publishing Group

naturejobs

JOBS OF THE WEEK

Last month's annual meeting of the US National Postdoctoral Association (NPA) in Boston saw a lively debate over the role and purpose that might be served by a set of core competencies for postdocs. The discussion questioned, for example, whether such competencies should be tantamount to regulations, as is seen in medicine and law.

As part of a first draft, an NPA study group identified six possible competencies that might define a postdoc: scientific knowledge, research skills, communication skills, professionalism, leadership and management skills, and responsible conduct of research. Not surprisingly, some in the audience bristled at the prospect of mandatory competencies, arguing that, unlike in medicine or law, there's no central body that accredits researchers, nor should there be. Others noted that most US graduate schools already encourage or require these principles, making them redundant.

But such competencies could be a ready barometer for industrial employers, and might round off the skill sets of the growing number of postdocs heading into 'non-traditional' careers. For postdocs from overseas working in the United States, with their varied graduate-school backgrounds and training, the competencies might also help even out their skill sets. And some at the meeting welcomed the idea of principles that might offer an alternative to publication count as a way of measuring success. One commentator, however, cautioned that faculty members would need to see the competencies as part of the training experience, rather than as just additional paperwork.

Strongly regulated competencies are a bad idea, especially given the multitude of career paths postdocs can take. But just having an endorsed set of carefully crafted core competencies could further distinguish and define the postdoc's role and importance, help individual postdocs to identify training areas in need of improvement, and smooth the transition to the ideal post-postdoc professional experience.

Gene Russo is editor of *Naturejobs*.

CONTACTS
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:** Dennis Chu

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

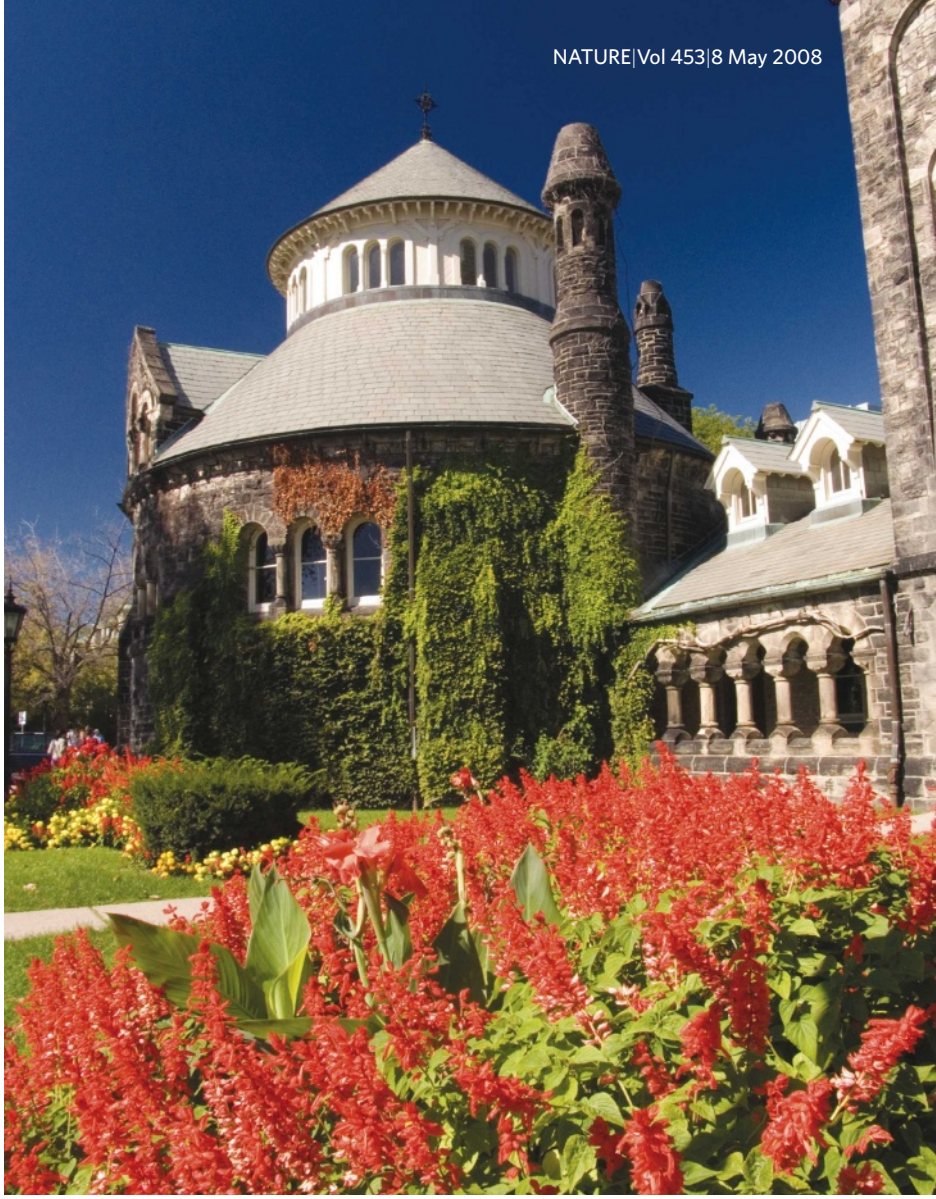
US Sales Manager: Peter Bless

India
Vikas Chawla (+91 1242881057)
e-mail: v.chawla@nature.com

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 (+81 3 3267 8765)
e-mail: a.watanabe@natureasia.com
Business Development Manager, Greater China/Singapore:
Gloria To (+852 2811 7191)
e-mail: g.to@natureasia.com

©2008 Nature Publishing Group



R. T. NOWITZ/CORBIS

TORONTO RISING

Specialist research centres are springing up in Canada's biggest city, nourished by government funds that also attract high-calibre scientists. **Kurt Kleiner** reports.

When the University of Toronto managed to lure chemical geneticist Guri Giaever away from Stanford University two years ago, part of the inducement was a new, bigger lab, and part was a prestigious government-funded research chair. But the biggest factor in the move, Giaever says, was the colleagues with whom she would be working. "In terms of what I'm doing, I would pretty much say hands down that Toronto is the best place in the world," she says.

Canadian scientists and administrators welcome such adulation. With the much bigger and richer United States to the south, Canada has often been preoccupied with a brain drain, as the brightest minds sought greater rewards at one of its neighbour's institutions. Increasingly, though, the country's biggest city, Toronto, is celebrating a 'brain gain' as it succeeds in attracting top researchers, often to work at brand new research centres. Federal and provincial efforts that began a decade ago are helping to attract high-calibre researchers and putting them in charge of long-term, 'big science' projects, according to researchers and business development officials. The new policies are an attempt to build on Toronto's impressive existing research infrastructure.

Billion-dollar budget

Most of this basic research is concentrated in Toronto's city centre. Within two kilometres of the intersection of University Avenue and College Street, on the University of Toronto campus, there are nine research hospitals, roughly 5,000 principal investigators, and research budgets totalling about Can$1 billion (US$990 million) a year. Since 2005, 93,000 square metres of research space have been added in this zone, with twice as much more planned.

The main research engine is the University of Toronto, along with its affiliated research hospitals, including the Hospital for Sick Children, St Michael's, Sunnybrook and Mt Sinai. Also downtown is the Centre for Addiction and Mental Health, which employs 100 research scientists and is building an 110,000-square-metre site at the cost of Can$380 million. Other universities in the area include Ryerson University, York University to the north, and McMaster University, about an hour away in Hamilton. McMaster is building a 158,000-square-metre innovation park, which will include space for the Canadian government's Materials Technology Laboratory, and a new Engineering Technology Centre.

Giaever was recruited in 2006 to set up the HIP-HOP (HaploInsufficiency Profiling/Homozygous deletion Profiling) Chemical Genomics Lab, which explores yeast gene function and conducts drug screening. Her lab is in the Donnelly Centre for Cellular and Biomolecular Research at the University of Toronto, which was completed two years ago at the university's downtown hub. As part of her package, she received a federally funded Canada Research Chair. The Canada Research Chair programme began in 2000, and spends Can$300 million a year to attract and retain research professors across the country.

Another of the new specialized research centres is the Ontario Institute of Cancer Research (OICR) at the University of Toronto, started in 2005 and funded by the provincial government with Can$350 million over

B. RAHN/A-FRAME

©2008 Nature Publishing Group

five years. The multidisciplinary translational centre will eventually employ 50 principal investigators.

The OICR's scientific director, Thomas Hudson, says that he is not only in the enviable position of recruiting researchers for a well-funded centre, but he has benefited from the Toronto area's wealth of potential candidates and its favourable reputation among researchers in many parts of the world. "There are 5,000 principal investigators within 15 minutes' walk of here," says Hudson, who helped found the International HapMap Project and is former director of the McGill University and Genome Quebec Innovation Centre. "There are few centres like that in North America."

Just across the street, Aled Edwards runs the Structural Genomics Consortium, a collaboration between the University of Toronto, the University of Oxford, UK, and Sweden's Karolinska Institute.

"We've gained this remarkable ability to do huge projects," says Edwards. The University of Toronto has long conducted good research, he says. But the new funding commitment has freed many researchers from having to chase two- and three-year grants. Edwards claims that the research culture in Toronto tends to give younger scientists more responsibility early in their careers compared with many places in Europe and some in the United States. "If you want to launch your academic career, Toronto is a fantastic place to do it," Edwards says.

Postdoc Ian Weaver agrees. Weaver, a developmental biologist from England, works on stem-cell research at the Hospital for Sick Children in Toronto. He praises the city's strong stem-cell science. (Scientists James Till and Ernest McCulloch were the first to prove the existence of stem cells in 1963 while at Princess Margaret Hospital's Ontario Cancer Institute.)

Although the cost of living is relatively high, notes Weaver, it still pales in comparison to that of cities such as London, Boston and San Francisco. Weaver himself hopes to stay in Toronto, and has applied for a tenure-track position at the University of Toronto.

Recent postdoc arrival Roman Iakoubov, an MD–PhD from Germany, says he finds the University of Toronto has more funding and better equipment than he is used to. He is impressed with the quality of his fellow researchers and with the city's multicultural mix. The main problem, he says, is one that postdocs everywhere face — competition for tenure-track positions is intense.

Falling short

Toronto's growing pharmaceutical company presence complements its academic successes, although industry in the region has under-performed so far. Vaccine specialist Sanofi Pasteur is the latest entrant; in April it announced plans for a Can$100-million research facility on its campus in north Toronto. Work on the 15,000-square-metre facility will begin later this year, and finish by 2010. Thirty research jobs are expected. Toronto was attractive for its existing vaccine-research infrastructure, diverse work force, and reasonable cost of living, according to Sanofi communications manager Mark Beazley.

Other companies with a major presence in Toronto include Apotex, GlaxoSmithKline, AstraZeneca and Eli Lilly. The pharmaceutical industry as a whole employs about 10,000 people in the Toronto region, mostly in sales, manufacturing and management. Province-wide, pharmaceutical companies spend about Can$500 million a year on research and development, according to Rx&D, the association of Canadian research-based pharmaceutical companies.

A. VANEK D. LEHTO

Gleaming: the new Donnelly Centre.

Where Toronto has had less success so far is in commercializing its academic research. Ontario has more biotechnology companies than any US state with the exception of Massachusetts and California. But judged against the amount spent on basic research in Toronto, the region generates only about half the commercialization opportunities it should, compared with successful biotech clusters such as Boston, says David Shindler, executive director of Biodiscovery Toronto, an organization that commercializes research.

"When you look at University Avenue and the billion dollars spent there annually, you're sort of saying, why aren't we the size of San Diego? Where are all the companies?" says Grant Tipler, chair of the Biotechnology Initiative, a non-profit organization committed to promoting the growth of biotechnology in Toronto and the surrounding region. He says there are a number of reasons Toronto has lagged — a research culture that values basic research more than entrepreneurship; lack of government funding for applied research; and a shortage of venture capital for early stage companies.

Therapeutic prospects

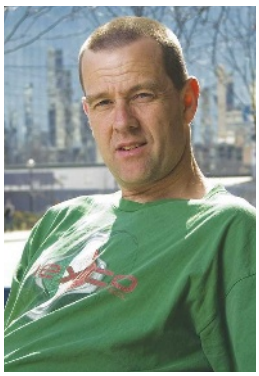
But there are some promising prospects. Among the region's emerging biotech companies is Amorfix Life Sciences, spun out of Neil Cashman's lab at the University of Toronto, concentrating on diagnosis and treatment of neurodegenerative diseases. Others include Transition Therapeutics, a drug-discovery company, and Arius Research, which develops therapeutic antibodies.

And Toronto has a number of initiatives attempting to bridge the gap between basic research and commercialization. Most visible is the Can$230-million downtown MaRS Centre, a non-profit organization that provides offices, lab space and business services for start-up biotech companies. It houses a number of biotech funding and commercialization organizations, including Biodiscovery Toronto and the Ontario Institute for Cancer Research.

The MaRS Centre has been at its current location — the old Toronto General Hospital building — since 2005. It plans to add new buildings with an additional 230,000 square metres of space by 2010. A programme called MaRS Innovation announced it had received Can$15 million in government funds in February to invest in commercialization ventures.

Tipler says the MaRS Centre has helped to provide a meeting place for discussion and ideas, one the advantages of big biotech clusters. He says that he runs into most of the important players in Toronto biotech simply by sitting in the atrium of the building. "Do we have a lot of the ingredients we need? Yes," says Tipler. But he acknowledges the challenges of forging a major Canadian biotech industry. "It may take a while," he says, "but we'll get there." ■

Kurt Kleiner is a science writer based in Toronto.



R. FAUBERT/CIHR

P. GAUTREAU

Settled in: Aled Edwards (top) and Guri Giaever recommend Toronto.

©2008 Nature Publishing Group

NATURE|Vol 453|8 May 2008

MOVERS

Karin Lochte, director, Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany

2000–07: Director, Biological Oceanography Research Unit, Leibniz Institute of Marine Sciences, University of Kiel, Kiel, Germany
1995–2000: Director, Biological Oceanography Research Department, Leibniz Institute for Baltic Sea Research Warnemünde, Rostock University, Rostock, Germany

Karin Lochte thought that she would be content teaching biology and chemistry. But a marine-science training course made her realize she preferred generating new knowledge as a scientist. Now a sought-after expert on the ocean's role in global climate change, Lochte says her most recent move will demand that she continue to inform a contentious policy debate with robust scientific findings.

As a postdoc at the Institute of Marine Science, University of Kiel, Lochte examined carbon turnover in the deep sea. To understand carbon-cycle dynamics relevant to climate change, it was important to understand how carbon is biologically exported from the sea surface to the sea floor, effectively exiting the carbon cycle. "With this work, I unintentionally drifted into the climate debate," she says.

Her work then took a southward turn. As a research scientist at the Alfred Wegener Institute for Polar and Marine Research, Lochte went to Antarctica to study how bacteria cooperate with phytoplankton in sea ice. She moved on to the University of Bremen in Germany, and then accepted a professorship in biological oceanography at the University of Rostock, in the former East Germany, to see first-hand the changing former communist region — and help establish a competitive scientific-research institute.

The move drastically altered her research. She began working on the nitrogen cycle in polluted coastal waters rather than carbon cycle in the open ocean. At the same time she was asked to sit on international scientific panels.

Ultimately, Lochte decided to return to her first love: open-ocean science. She focused on iron as an important 'fertilizer' of the ocean that can help to soak up carbon dioxide from the atmosphere while at the Leibniz Institute of Marine Sciences in Kiel. Happy there, Lochte admits she had to be coaxed into her current position at the Wegener institute. But, she says, it's the perfect place to strengthen much-needed research in the Arctic, a region experiencing more rapid changes than any other ecosystem in the world. She laments how a lack of funding for the ships and infrastructure needed in polar regions is crippling marine research — a trend exacerbated by soaring oil prices.

Former colleague Carol Turley at the Plymouth Marine Laboratory, UK, says that, with contentious issues such as climate change, it is important to have leaders such as Lochte with integrity as well as an appreciation of the whole picture. "Karin won't spin the facts," says Turley. ■

Virginia Gewin

NETWORKS & SUPPORT

Animal assets in academia

With pet owners willing to pay big money for special operations, many veterinary surgeons are taking up lucrative specialist private practices. Recruiting graduates into academia has become difficult. This could be detrimental, as graduates are well-suited to address animal-health crises, such as infectious diseases, as well as basic biomedical questions.

"We are losing a core of faculty that have taught students in the past," says Michael Kotlikoff, dean of the College of Veterinary Medicine at Cornell University in Ithaca, New York. In response, Cornell has created a two-year clinical-fellowship programme as a route for academics. Until now, veterinary students with academic inclinations have had only a handful of joint doctor of veterinary medicine (DVM)–PhD programmes to apply to. Those who realize their academic calling during a DVM must follow up with a PhD. Kotlikoff was inspired by programmes for academia-bound physicians eager to combine basic research skills with clinical training. "Nobel prizewinner Harold Varmus followed this route — obviously with great success," says Kotlikoff.

Sophy Jesty, one of the first three fellows on the Cornell programme, jumped at the chance of a two-year paid fellowship: $60,000 salary and $15,000 in research funds. "I was never that interested in stepping out of clinics long enough to earn a PhD," she says. Jesty, a trained cardiologist, says that this experience will probably steer her towards academia, ideally as a clinical professor spanning the gap between basic science and the clinic.

Last year, the University of Pennsylvania's School of Veterinary Medicine set up a similar programme, focused on translational research in infectious diseases. It plans to offer fellowships in three other areas — comparative oncology, regenerative medicine and stem-cell biology — as funding becomes available. "We're looking for people who want to cure, not manage, disease," says Joan Hendricks, the school's dean.

Kotlikoff and Hendricks say that veterinary medicine offers unique biomedical insights, particularly into naturally occurring genetic diseases that are also found in humans. And its cutting-edge resources could benefit other fields. Cornell's biobank of canine DNA contains a huge suite of tissue samples, and blood tests will aid canine genetics and other research.

Kotlikoff and Hendricks hope other schools will adopt similar approaches, and convince veterinary scientists this is a viable career path. ■

Virginia Gewin

POSTDOC JOURNAL

Going with your gut

The hummus in Israel is made from freshly cooked chickpeas yielding a creamy, delectably rich spread that complements fresh pitta and Syrian olives. I find conversations flow easily with hummus — whether debating who has the best hummus in town or the best interpretation for a set of experiments. It's a wonder how simple things can be interpreted so differently. Yet, just as the quest for the ultimate hummus can lead to new destinations; so too a simple band on a gel can direct me to one experiment and my colleague to another.

Differences in interpretation that may initially seem trivial could have a crucial impact on the direction and take-home message of a project. I think part of becoming a seasoned scientist is learning how to build confidence in one's ability to interpret data independently and to defend those interpretations.

I experienced this recently while preparing a manuscript. I debated with my colleagues on how to interpret a key phenotype in a pepper mutant. In the end I went with my gut — which, I am learning, is a wise move for a good scientist, and for a hummus aficionado. Now, when I discuss data with friends, it is at my chosen restaurant. Who would have thought my ability to choose between hummus with a touch of cumin or a 'shpritz' of lemon would help me defend my experimental interpretations? ■

Zachary Lippman is a postdoctoral fellow at the Hebrew University of Jerusalem's faculty of agriculture.

©2008 Nature Publishing Group

C001_nat_08may 2/5/08 07:01 Page 1

University of Toronto, the research powerhouse above America:

- Publishes more science research than any public university in North America (Thomson ISI)
- Third most highly cited science faculty in North America (Thomson ISI)
- One of only 8 universities worldwide in the top 20 across humanities, social sciences, technology, natural sciences, and life sciences (Times Higher Education Supplement, 2007)

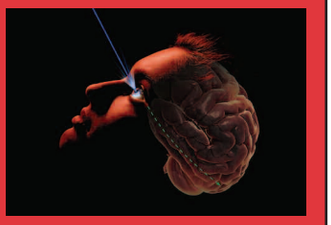
UNIVERSITY OF
TORONTO

www.jobs.utoronto.ca/faculty

CANADA'S ANSWERS TO THE WORLD'S QUESTIONS.

NW131481A



REAL-WORLD SOLUTIONS TO TODAY'S CHALLENGES

Faculty and graduate students at York University tackle real-world scientific, technological, and socio-cultural issues facing Canada and the world.

Researchers at York are leaders in a range of scientific fields including biomolecular interactions, geomatics engineering, health technologies, space science, environmental sciences and digital media. York is also home to world-class research centres like the Centre for Vision Research, the Centre for Research in Mass Spectrometry, and the Centre for Research in Earth and Space Science.

York supports unique collaborations by connecting its researchers to hospital clinicians and industry practitioners, and by engaging in innovative regional consortia such as the National Centre for Medical Device Development and the Consortium on New Media, Creative and Entertainment R&D in the Toronto Region.

&

science.yorku.ca

YORK UNIVERSITÉ UNIVERSITY
redefine THE POSSIBLE.

NW131892R

SGC Post-Doctoral Positions

Structural Genomics Consortium Toronto

The Structural Genomics Consortium (SGC) is a charitable enterprise, with a mandate to produce medically relevant human proteins and proteins from human parasites, determine the three-dimensional structures and interactions with natural and synthetic binding partners and place the associated information in the public domain without restriction. The SGC develops and implements rapid, parallel methodologies for protein production, structural biology, and chemical biology. We have purified thousands of human proteins implicated in cancer, cellular signaling, epigenetic signaling, ubiquitin signaling, diabetes and metabolic disease, osteoporosis, and drug metabolism, as well as proteins from malaria parasites. We have determined hundreds of protein structures including protein-protein, protein-peptide and protein small molecule complexes providing insight into the mechanism of interaction, catalysis, signaling and inhibition. For more details see **www.TheSGC.com.**

The Toronto site of the SGC, housed within the MaRS Discovery District **http://www.marsdd.com/MaRS-Home.html**, seeks Post-Doctoral Fellows and visiting scientists to work on human and malaria protein families with strong medical and/or pharmaceutical relevance. Experience in one or more of the following areas is desirable; protein crystallography, enzymology, protein expression and purification, protein biophysical characterization, computational chemistry or parasitology. For information on our visiting scientist program see **http://www.thesgc.com/Visiting%20Scientists/**

Interested candidates should submit their CVs and a brief statement of research interests to **HR.sgc@utoronto.ca**

The University of Toronto is strongly committed to diversity within its community and especially welcomes applications from visible minority group members, women, Aboriginal persons, persons with disabilities, members of sexual minority groups, and others who may contribute to the further diversification of ideas. All qualified candidates are encouraged to apply; however, Canadians and permanent residents will be given priority.

NW131477R

Heart and Stroke Richard Lewar Centre of Excellence in Cardiovascular Research

Towards better outcomes in cardiovascular disease

The HSRLCE is a unit of Canada's premier University of Toronto. The HSRLCE brings together the best of cardiovascular sciences at the University and its affiliated hospitals. With a focus on translational research, the HSRLCE provides ample opportunities for scientific interactions and collaborations, uniting clinicians and scientists within the Centre and throughout Toronto.

We aim to provide promising young investigators the opportunity to interact with multi-disciplinary teams of basic and translational scientists. The Centre offers several studentships, fellowships, and other support for qualified individuals involved in cardiovascular research at U of T or at any of it's affiliated hospitals and research institutes. In addition, the HSRLCE hosts regular scientific and educational events, include a program of Distinguished Visiting Professors, a strategic training initiative for physicians and biomedical researchers, and a training program for graduate students and post-doctoral fellows.

The Centre believes that trainees and young investigators are the lifeblood of research and we invest accordingly. For more information about the HSRLCE at the UofT, please visit our website at www.hsrlce.com

UNIVERSITY OF TORONTO
FACULTY OF MEDICINE

N W131864R

C003_nat_08may 2/5/08 13:07 Page 3

Scientist Position in Inflammation Research
Departments of Medicine and Microbiology and Immunology,
Schulich School of Medicine & Dentistry
The University of Western Ontario

The Departments of Medicine and Microbiology & Immunology are seeking a faculty position, to be appointed jointly, at the rank of Assistant Professor in the general area of Inflammation. However, priority will be given to candidates with research interests and expertise in the areas of infectious diseases, autoimmunity, and transplantation, which will complement current areas of research strength in both departments. Exceptional candidates considered for appointment at a higher rank. Candidates with PhD, MD., or MD./PhD degrees are encouraged to apply, and those with an MD or equivalent must be eligible for licensure in the Province of Ontario. Individuals with a Ph.D. or equivalent will be appointed in a probationary (tenure-track) position. The position guarantees 75% protected time for research activities. An outstanding start-up package includes high quality laboratory space in the Department of Microbiology and Immunology, excellent benefits and competitive salaries commensurate with academic qualifications. The candidate is expected to establish an independent, externally funded research program and collaborations with others at the University and its affiliated institutes. The successful candidate will hold an academic position in the appropriate Division within the Department of Medicine and will be expected to participate in the teaching programs of the Departments of Medicine and Microbiology & Immunology at both undergraduate and graduate levels, as appropriate.

To look at the full advertisement please see the ad on Nature's website: http://Naturejobs.com OR the Department of Microbiology and Immunology website at: http://www.uwo.ca/mni

NW131986R

www.cam.ac.uk/jobs/
A world of opportunities

UNIVERSITY OF CAMBRIDGE

Lecturer in Cell and Molecular Biology

Department of Pathology
£34,793 - £44,074 pa

We wish to appoint a University Lecturer in Cell and Molecular Biology. You should have research interests that can bring to bear a genetics, genomics or molecular biological perspective to basic questions of tissue differentiation, regeneration, repair and remodelling. You should have a publication record of high quality and the ability to win external grant funding. You will also contribute to undergraduate teaching in the areas of genetics, genomics and cell biology. The post has tenure, subject to satisfactory performance in the first 5 years, and is remunerated on the above scale according to experience.

The Department of Pathology comprises 300 staff, with strong research programmes in Cell Biology, Cancer, Infection and Immunology and plentiful opportunities for development of cross-disciplinary projects. It was awarded 5* status in the last RAE. Information on the Department can be found at www.path.cam.ac.uk.

Further particulars for this post can be obtained by emailing jobs@path.cam.ac.uk.

Applications should include a full CV with list of publications, a one-page statement of research interests and future plans, together with the contact details of three academic referees. Please address applications to Professor Andrew H Wyllie, FRS and send by email to jobs@path.cam.ac.uk or post to the Administration Office, Department of Pathology, Tennis Court Road, Cambridge, CB2 1QP. Please quote the job reference (PK03359) on all correspondence.

Closing date: 22 May 2008.

The University is committed to Equality of Opportunity.

U131931R

Appointments to BBSRC Strategy Panels

The Biotechnology and Biological Sciences Research Council (BBSRC) funds basic and strategic biological research, supports postgraduate training, promotes knowledge transfer and public engagement with science.

We are looking to appoint high calibre, committed individuals from **academic** and **industrial** sectors to join BBSRC's Strategy Panels. The Panels play a leading role in the development and implementation of the Council's policies and priorities. We are currently seeking to fill a number of vacancies on the following:

Healthy Organism
Integrative and Systems Biology
Bioscience for Society
Sustainable Agriculture
Tools and Resources
Bioscience for Industry
Studentships and Fellowships

It would be helpful for applicants to have some knowledge of BBSRC and to be aware of the Council's priorities. However, it is more important to have a broad understanding of UK science funding, and the issues and opportunities relating to UK bioscience in an international context. In addition, applicants should have expertise in one or more of the major strategic areas covered by the Panel to which they are applying. Members will also be expected to be able to show proven ability and experience to exercise judgement across a broad spectrum of strategic and policy issues and make difficult decisions on a number of competing perspectives.

Previous applicants who meet the above criteria, but who were not selected for an appointment from earlier calls for applications are encouraged to reapply as the membership needs of the Panels change each year.

Appointments will commence in January 2009 for one year initially, with anticipated extension for a further two years.

Further details can be found on the BBSRC website at **www.bbsrc.ac.uk**

The closing date for applications is Friday 13 June 2008.

The BBSRC welcomes applications from all sections of the community irrespective of race, age, transgender, ethnicity, religion, sexual orientation, disability or gender.

bbsrc
biotechnology and biological sciences research council

U131933R

Cifc_nat_08may 2/5/08 07:37 Page 2

Research grows at SickKids

Join Canada's most research-intensive hospital, covering the breadth of health research, from the molecular to the population level. You'll find more than 2,000 of the world's best researchers, trainees and other staff in an environment rich in fostering growth, development and understanding.

We are always looking for talented postdoctoral fellows and trainees in these research programs:

- Cell Biology
- Child Health Evaluative Sciences
- Developmental & Stem Cell Biology
- Genetics & Genome Biology
- Molecular Structure & Function
- Neurosciences & Mental Health
- Physiology & Experimental Medicine

In the next year we will be recruiting faculty in these child health research areas:

- Cellular & Molecular Imaging
- Epigenetics & Population Genetics
- Health Policy
- Neurobehaviour & Systems Neuroscience
- Stem Cell & Regenerative Medicine

For more information, please visit **sickkids.ca/research**

Affiliated with the University of Toronto

NW131557R

The icosahedral anaster

A slight detour?

John P. Boyd

Once, Samaha was flesh and blood, but now he lives only as a pattern of electrons, circulating in a tiny cube barely two centimetres on a side, plus back-ups. He had expected to sleep the long years to Sagartha, but the alarm awakened him into the virtual-reality immersion of a starship cabin.

It was an illusion, and it was the truth. He was in a starship only three metres long, hurtling between the stars. The virtual instruments would control the starship as truly as if they were real. And the planet that had triggered the alarm was real, too.

JACEY

Interstellar space was filled with planets. Once, it had been thought that solar systems simply condensed in place. In reality, a young star system was a battleground of collisions and near-misses. A hundred Plutos absorbed in collisions. Another hundred flung by near-misses or triple interactions into the dark between the suns.

Iceballs, most of them, with sometimes a little rock. Why awaken him?

The AI appeared on-screen as a middle-aged woman with spiky white hair. "The planet is unnatural. Perhaps a detour is necessary."

"The void is full of anasters. Worlds without stars. Why wake me?"

The screen filled with an image that reminded him of a 20-sided die he had used as a child. A light year from the nearest star, the planet should be as black as the Deep itself. But this was an infrared image, and the planet was burning up with heat.

"An icosahedron!" It was absurd. The world was half as large as Earth and almost as dense. The 12 vertices, each a vast 5-sided pyramidal mountain, should have slid back into the great depression at the centre of each triangular face. No granite, no obsidian, could hold up a mountain many hundreds of kilometres high. An icosahedron was a very bad approximation to the sphere; the vertices were nearly 30% farther from the centre than were the centres of each face.

And yet the mountains held; the rock did not flow downhill in great rivers of plastic deformation. The 20 triangles remained triangular. Disconcerted, he woke up the captain.

Ramon was not at all pleased. His virtual body sat in the co-pilot's chair, glowering. "I expected to wake up in Sagartha, dammit! I feel like a man who had already started to walk the pearly stairs to heaven and then wham! Back in the world of pain."

"Don't be so melodramatic. You have no body to feel pain. And — haven't you any curiosity?"

"Yes! Half the human race, and most of the rest of the galaxy, live the Life Electric in Sagartha. A whole galaxy of sophonts, bitstreams in a world free of all constraints of time and space. Immortal!" Ramon pounded the virtual desk. "I want it. I want it bad. And now … Don't you understand, Sam? Everything is there. The religious dreams of a better world, evil-free, where a man can fly! It's all true, but it's not God's gift. It's the gift of the Janx."

That ancient race, the first to emerge as a lasting civilization, had discorporated at least a hundred million years ago, building an immense virtual-reality structure a few thousand light years from the centre of the galaxy. In time, perhaps a little lonely, they had invited other races to share. Some of the recently sentient inhabitants of the third planet around an obscure yellow star were still in carbon-and-water, back on Earth and its colonies, but most races had 'gone electric' when humans were mute banana-eaters walking on their knuckles.

"But aren't you curious? This planet is physically impossible. I want to investigate …"

"No! No! No! This ship is carrying more than a million psyches. Do they want to be called back from heaven for a bit of science? Sagartha is the summation of every mystery and desire of a hundred races for half a billion years. This funny ball of rock is so insignificant … You're watch captain. Not boss."

Samaha hung his head.

Like a lot of other humans still in carbon, he had found a pen pal among the Janx. His 'angel' was named Yth, and had tried very hard to answer all his questions. But over the very low bandwidth interstellar link, Yth had to speak in words, translated perhaps imperfectly into English, of things that could really only be explained by total immersion. And so, in time, like all the other consciousnesses crammed into the tiny hold, Samaha had chosen to flash his body into ash, and send his psyche to Sagartha.

Finally Samaha said: "Thank you, Ramon." The other waved dismissively and then was simply not there.

Samaha did not sleep. Instead, he commanded the AI to collect data during the fly-by. Were there cities on this fantastic, impossible world? Great pentagonal towers lit only by mighty star-lamps within? Fugitives fleeing from … what?

He would not pass near enough to see cities. At closest approach, he did learn that each face contained a large, triangular ocean, liquefied by heat bubbling up from some thoroughly artificial heat source within. He suspected that the mass was distributed non-uniformly, minimizing the strain on the impossible 5-sided mountains. He suspected or fantasized much more. But he would never know.

With paradise waiting, who would want to stay in flesh a minute longer? And yet his Christian and Muslim friends had loved the grass, and the waves slowly rolling in from the ocean, as much as he. Some passions of one's first life were precious even in the face of a promised but unimaginable paradise.

He would never know. ■

John Boyd has been a professor of atmospheric and oceanic science at Michigan for 30 years. He has published two books, 170 scientific articles and a dozen science-fiction stories.

FUTURES

©2008 Nature Publishing Group