nature

www.nature.com/nature

Vol 462 | Issue no. 7273 | 3 December 2009

Climatologists under pressure

Stolen e-mails have revealed no scientific conspiracy, but do highlight ways in which climate researchers could be better supported in the face of public scrutiny.

The e-mail archives stolen last month from the Climatic Research Unit at the University of East Anglia (UEA), UK, have been greeted by the climate-change-denialist fringe as a propaganda windfall (see page 551). To these denialists, the scientists' scathing remarks about certain controversial palaeoclimate reconstructions qualify as the proverbial 'smoking gun': proof that mainstream climate researchers have systematically conspired to suppress evidence contradicting their doctrine that humans are warming the globe.

This paranoid interpretation would be laughable were it not for the fact that obstructionist politicians in the US Senate will probably use it next year as an excuse to stiffen their opposition to the country's much needed climate bill. Nothing in the e-mails undermines the scientific case that global warming is real — or that human activities are almost certainly the cause. That case is supported by multiple, robust lines of evidence, including several that are completely independent of the climate reconstructions debated in the e-mails.

First, Earth's cryosphere is changing as one would expect in a warming climate. These changes include glacier retreat, thinning and areal reduction of Arctic sea ice, reductions in permafrost and accelerated loss of mass from the Greenland and Antarctic ice sheets. Second, the global sea level is rising. The rise is caused in part by water pouring in from melting glaciers and ice sheets, but also by thermal expansion as the oceans warm. Third, decades of biological data on blooming dates and the like suggest that spring is arriving earlier each year.

Denialists often maintain that these changes are just a symptom of natural climate variability. But when climate modellers test this assertion by running their simulations with greenhouse gases such as carbon dioxide held fixed, the results bear little resemblance to the observed warming. The strong implication is that increased greenhouse-gas emissions have played an important part in recent warming, meaning that curbing the world's voracious appetite for carbon is essential (see pages 568 and 570).

Mail trail

A fair reading of the e-mails reveals nothing to support the denialists' conspiracy theories. In one of the more controversial exchanges, UEA scientists sharply criticized the quality of two papers that question the uniqueness of recent global warming (S. McIntyre and R. McKitrick *Energy Environ.* **14,** 751–771; 2003 and W. Soon and S. Baliunas *Clim. Res.* **23,** 89–110; 2003) and vowed to keep at least the first paper out of the upcoming Fourth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). Whatever the e-mail authors may have said to one another in (supposed) privacy, however, what matters is how they acted. And the fact is that, in the end, neither they nor the IPCC suppressed anything: when the assessment report was published in 2007 it referenced and discussed both papers.

If there are benefits to the e-mail theft, one is to highlight yet again the harassment that denialists inflict on some climate-change researchers, often in the form of endless, time-consuming demands for information under the US and UK Freedom of Information Acts. Governments and institutions need to provide tangible assistance for researchers facing such a burden.

The e-mail theft also highlights how difficult it can be for climate researchers to follow the canons of scientific openness, which require them to make public the data on which they base their conclusions. This is best done via open online archives, such as the ones maintained by the IPCC (www.ipcc-data.org) and the US National Climatic Data Center (www.ncdc.noaa.gov/oa/ncdc.html).

Tricky business

But for much crucial information the reality is very different. Researchers are barred from publicly releasing meteorological data from many countries owing to contractual restrictions. Moreover, in countries such as Germany, France and the United Kingdom, the national meteorological services will provide data sets only when researchers specifically request them, and only after a significant delay. The lack of standard formats can also make it hard to compare and integrate data from different sources. Every aspect of this situation needs to change: if the current episode does not spur meteorological services to improve researchers' ease of access, governments should force them to do so.

The stolen e-mails have prompted queries about whether *Nature* will investigate some of the researchers' own papers. One e-mail talked of displaying the data using a 'trick' — slang for a clever (and legitimate) technique, but a word that denialists have used to accuse the researchers of fabricating their results. It is *Nature*'s policy to investigate such matters if there are substantive reasons for concern, but nothing we have seen so far in the e-mails qualifies.

"The theft highlights the harassment that denialists inflict on some climate-change researchers."

The UEA responded too slowly to the eruption of coverage in the media, but deserves credit for now being publicly supportive of the integrity of its scientists while also holding an independent investigation of its researchers' compliance with Britain's freedom of information requirements (see http://go.nature.com/zRBXRP).

In the end, what the UEA e-mails really show is that scientists are human beings — and that unrelenting opposition to their work can goad them to the limits of tolerance, and tempt them to act in ways that undermine scientific values. Yet it is precisely in such circumstances that researchers should strive to act and communicate professionally, and make their data and methods available to others, lest they provide their worst critics with ammunition. After all, the pressures the UEA e-mailers experienced may be nothing compared with what will emerge as the United States debates a climate bill next year, and denialists use every means at their disposal to undermine trust in scientists and science. ■

RESEARCH HIGHLIGHTS

IMMUNOLOGY

Timely defence

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0906361106 (2009)

Many activities of the immune system follow rhythmic daily cycles. Now researchers have found that some immune cells have their own circadian clocks.

Achim Kramer of the Charité Medical University in Berlin and his colleagues took immune cells and tissues from mice at regular intervals throughout the day. They found that macrophages — cells that form part of the immune system's first line of defence against bacterial infections — from the spleen, lymph nodes and abdominal area express circadian clock genes. In addition, they showed that about 8% of macrophage genes are expressed rhythmically.

The authors also report that the secretion of immune modulators by spleen macrophages in response to bacterial toxins follows circadian rhythms.

BIOLOGY

Beetle-juice antifreeze

Proc. Natl Acad. Sci. USA **106,** 20210–20215 (2009)

Many animals survive extreme cold by producing 'antifreeze' compounds that inhibit ice growth. The compounds described so far have all been proteins.

Kent Walters at the University of Notre Dame in Indiana and his colleagues have now characterized the first animal antifreeze that contains little or no protein.

They isolated the compound from the darkling beetle *Upis ceramboides* (pictured below), which can withstand temperatures as low as –60° C, allowing it to live in harsh climates like that of Alaska. Analysis showed it to comprise a xylomannan saccharide with a fatty acid component.

S. SHARMA

Chess obeys the law

Phys. Rev. Lett. **103,** 218701 (2009)

A. BARRIENTOS/AP PHOTO

In a large sample of texts, 'the' is the most frequently used word. Coming in second is 'of', which is used about half as often. 'And' — in third place — is used about one-third as often, and so on. This curious relationship is known as Zipf's law after the linguist George Kingsley Zipf.

Various phenomena follow formulations of Zipf's law, and chess can now be added to the list. Bernd Blasius of the University of Oldenburg in Germany and Ralf Tönjes of Ochanomizu University in Tokyo analysed the first 40 moves of more than a million chess games recorded in an online database. They found that the frequency of the most common moves followed the law. They add that board games could help physicists to develop new statistical tools.

ENVIRONMENTAL CHEMISTRY

Plucking pollutants

Environ. Sci. Technol. doi:10.1021/es902407g (2009)

DNA fragments offer an effective way to extract arsenic from contaminated groundwater, a team in South Korea reports.

Jiho Min at Chonbuk National University, Yang-Hoon Kim at Chungbuk University and their colleagues designed an aptamer — a short, single strand of DNA that can bind to a specific molecule — for the purpose. The aptamer was able to efficiently remove arsenic from samples of groundwater collected in different areas of Vietnam, where arsenic levels are often higher than the US Environmental Protection Agency's recommended maximum.

Aptamer devices could be made with cheap materials such as silicon, the authors say.

GENETICS

One on one

Science **326,** 1231–1235 (2009)

Human cells, with their two sets of chromosomes, do not lend themselves to large-scale genetic screens as simple model organisms such as yeast have so profitably done.

Thijn Brummelkamp at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, and his colleagues have devised a way around the problem. Using a cell line with only one copy of most human chromosomes, they inactivated various genes using a method called insertional mutagenesis. The researchers then screened cells that were resistant to particular pathogens to see which genes invaders might rely on to attack.

Using the technique, the team identified two host genes used by the influenza H1N1 virus to infect cells, as well as genes exploited by other bacterial toxins to kill host cells. The authors say the method could help in developing new antiviral therapies.

NEUROSCIENCE

Brain's immune connection

Neuron **64,** 463-470 (2009)

Connections between neurons strengthen or break during brain development. Unexpectedly, key cell-surface proteins involved in immunity seem to regulate some of this plasticity.

Carla Shatz of Stanford University in Palo Alto, California, and her colleagues found that two members of the family of major histocompatibility complex class I (MHC I) proteins limit the 'tuning up' of circuitry involved in visual processing. Mice in which the genes for these two proteins had been deleted performed better in a visual task involving the blocking of one eye than did normal mice. The MHC I proteins curb the retuning of circuitry that enables the functioning eye to compensate for the blocked one, the authors say.

ECOLOGY

Diverse recovery

J. Ecol. doi:10.1111/j.1365-2745.2009.01603.x (2009)

Protecting biodiversity improves the chances that an ecosystem will contain a species that allows it to recover after an extreme environmental event, scientists have concluded.

Jasper van Ruijven and Frank Berendse at Wageningen University and Research Centre in the Netherlands followed the progress of more than 100 small experimental plots planted with either individual species or varying mixtures of eight common species, including several grasses, before and after a natural drought. The drought occurred six years after planting.

They found that plots with greater biodiversity did not show improved resistance to the drought — but they were able to recover more efficiently. They attribute most of this effect to one species, *Anthoxanthum odoratum*. Recovery was independent of pre-drought biomass.

GENETICS

Immune impediment

Nature Genet. **41,** 1341–1344 (2009)

Transplanted bone marrow cells commonly attack the recipient's cells, even though key proteins on the surface of the donor's and recipient's cells match. To find out what might be the cause of this 'graft-versus-host disease', Steven McCarroll of Harvard Medical School in Boston, Massachusetts, and his colleagues analysed gene deletions in the genomes of 1,345 pairs of patients and immune-matched siblings from whom bone-marrow transplants were made.

They found that the immune attack was more likely to occur when the donor — but not the recipient — had deletions in both copies of the gene *UGT2B17*. The donor's immune cells seem to respond to the gene's protein as 'foreign' in the recipient.

ASTRONOMY

A black hole draws near

Astrophys. J. **706,** L230–L234 (2009)

Astronomical distances are best measured by trigonometric parallax — using the annual shift in star position caused by Earth's motion around the Sun to derive distance. Now James Miller-Jones of the National Radio Astronomy Observatory in Charlottesville, Virginia, and his colleagues have used parallax to measure the first accurate distance to a nearby binary star system containing a black hole.

The authors took radio measurements at three-month intervals for a year and combined them with archival data. They reveal that the star–hole system, known as V404 Cygni, is just 2,390 parsecs from Earth — nearly half the distance previously thought. The group believes that the previous work underestimated the interstellar dust along the line of sight to the star system. The authors suggest that future parallax measurements will improve the understanding of how black holes form and behave.

NEUROSCIENCE

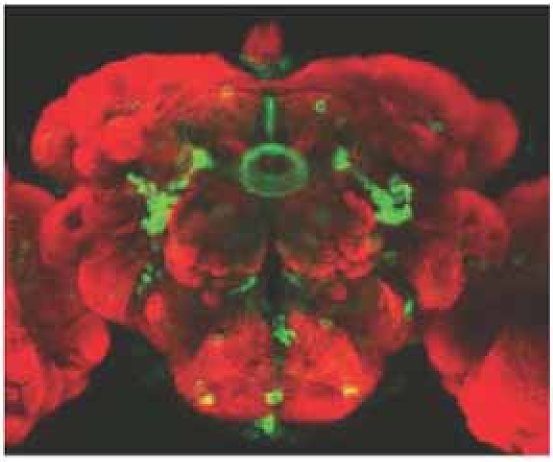
Rude awakening

Neuron **64,** 522–536 (2009)

Fruitflies recruit distinct neural circuits when undergoing different forms of arousal — either waking from sleep, or being disturbed by puffs of air.

David Anderson at the California Institute of Technology in Pasadena and his colleagues found that flies with loss-of-function mutations in the dopamine receptor were more easily startled by air puffs than were flies without the mutation, but seemed less easily roused from sleep, as they slept longer.

When the researchers restored normal functionality to the dopamine receptor in a brain area called the central complex (pictured left, dopamine receptors labelled green), the puff-induced arousal dropped to normal levels but sleep arousal remained unchanged.



ELSEVIER

JOURNAL CLUB

Peter Baumann
Howard Hughes Medical Institute, Stowers Institute for Medical Research, Kansas City, Missouri

A molecular biologist explores ways to revolutionize agriculture.

The complete absence of sex in a few species has long fascinated biologists, but their research is driven by more than just curiosity. Hybrid plants are the mainstay of agriculture, but require ongoing breeding and selection to maintain their desirable traits. Apomixis, or asexual reproduction by seeds, is rare among commercially important crops, but engineering plants capable of this could produce stable crops with valuable traits.

Three Herculean tasks are involved: alteration of meiosis (the cell division that normally reduces the number of chromosomes in the sex cells, or gametes) to maintain the full maternal genome; fertilization-independent development of the embryo; and formation of the endosperm tissue that nourishes the embryo.

Raphaël Mercier of the French National Institute for Agricultural Research in Versailles and his team have taken a step towards achieving this goal. Using a combination of three mutants, they engineered a mustard weed that produces gametes carrying the complete maternal genome (I. d'Erfurth *et al. PLoS Biol.* **7,** e1000124; 2009). Their breakthrough came while characterizing a mutation in the aptly named *omission of second division (osd1)* gene, which causes the reproductive cells to skip the second meiotic division. By combining an *osd1* mutant with mutations that modify two other steps in meiosis, the team made meiosis similar to mitosis — cell division that occurs in non-reproductive cells.

Conservation of the genes involved across crop species fosters hopes that the strategy can be applied to many of them. The problem of endosperm formation will have to be overcome, and unfertilized seeds will need to be coaxed into development. The available tool kit of mutants affecting these processes makes me optimistic that these challenges will be overcome. However, convincing consumers that heavily engineered plants can secure future food supplies may require more than scientific ingenuity.

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

NEWS BRIEFING

M. VATSYAYANA/AFP/GETTY IMAGES

TRAGIC ANNIVERSARY

Local authorities have delayed plans to open the site of the Union Carbide factory in Bhopal, India, that was responsible for a chemical leak that killed thousands of people 25 years ago this week. The state government of Madyha Pradesh said in early November that it would allow visitors to tour the plant in a bid to reassure people that it was now safe, amid protests from victims' groups (above). But the state's Gas Relief and Rehabilitation Minister Babulal Gaur said last week that the opening would be pushed back until after municipal elections in December, in line with rules forbidding major policy announcements in the run-up to the polls.

POLICY

Controversial cuts: China will cut its carbon intensity — the amount of carbon dioxide emitted relative to its economic output — by 40–45% from 2005 levels by 2020, China's State Council decreed on 26 November. The announcement came the day after US President Barack Obama pledged to reduce emissions by 17% by 2020 compared with 2005 levels. See page 550 for more.

Sceptic victory: The Australian Liberal Party, the main opposition to Prime Minister Kevin Rudd's governing Labor Party, elected climate-change sceptic Tony Abbott as its leader on 1 December. The move has jeopardized Rudd's attempt to pass legislation on an emissions-trading scheme before the United Nations Climate Change Conference in Copenhagen next week. Abbott replaces Malcolm Turnbull, who backed amended cap-and-trade legislation, triggering sceptics in his party to try to oust him last week.

Enrichment move: President Mahmoud Ahmadinejad has ratified an order calling on the Atomic Energy Organization of Iran to draw up plans for ten gas-centrifuge enrichment plants similar to the one already being constructed in Natanz. The directive comes just days after the International Atomic Energy Agency in Vienna urged the Islamic republic to halt its enrichment activities.

HIV decline: The death toll from AIDS has topped 25 million people, but new HIV infections are dropping sharply, according to a report released on 24 November by the Joint United Nations Programme on HIV/AIDS and the World Health Organization (WHO). In 2008, some 2.7 million people became infected with the virus, but the rate of infection has dropped by 17% over the past eight years. Meanwhile, the WHO recommended on 30 November that the widely used antiretroviral drug stavudine be phased out because of its "long-term, irreversible" side effects, which include nerve damage.

Science stimulus: The Spanish government announced on 26 November which universities would benefit from the inaugural round of a €150-million (US$226-million) annual programme to bolster teaching and research in the country. See page 552 for more.

Radiation contamination: The Nuclear Power Corporation of India has ordered a probe into what it believes to be sabotage that exposed 55 workers at its Kaiga plant in southwestern India to radiation last week. The workers were treated for poisoning after drinking from a water cooler contaminated with tritium — a radioactive isotope of hydrogen. Anil Kakodkar, chairman of the Atomic Energy Commission of India, says the tritium may have been introduced into the cooler in a "malevolent act" by a disgruntled employee.

SOUND BITES

"I hope that I will have the key to the door so that I do not let the buggers out until they have done a deal."

John Prescott, the Council of Europe's rapporteur on climate change, presses home the urgency of getting a climate deal in Copenhagen this month.

Saving science: Japanese researchers have launched protests against budget cuts proposed by the country's new government. The move coincides with the end of hearings on 27 November to help government-appointed working groups recommend where the axe should fall. See page 557 for more.

RESEARCH

Record breaker: Just ten days after it was restarted, the Large Hadron Collider (LHC) has become the world's highest-energy particle accelerator. On 30 November, the giant machine, located at CERN, Europe's premier high-energy physics laboratory near Geneva, Switzerland, accelerated two beams of protons to energies of 1.18 teraelectronvolts (TeV). That beats the previous record of 0.98 TeV set by the Tevatron at Fermi National Accelerator Laboratory in Batavia, Illinois. The LHC hopes to reach energies of 3.5 TeV by early 2010.

Old bird: NASA's decade-old Quick Scatterometer or QuikSCAT satellite can no longer fulfil its main mission of measuring global wind speeds and direction after the failure of its key instrument. The satellite's radar dish antenna stopped spinning on 23 November. Data from the satellite, which was designed to last only two years, were used by weather forecasters worldwide.

Antarctic melt: Global sea levels could rise by 1.4 metres by 2100 — around twice as much as predicted by the Intergovernmental Panel on Climate Change in 2007 — owing to the melting of ice sheets in western Antarctica, according to an international review of the continent's climate published on 1 December. The Scientific Committee on Antarctic Research predicts that Antarctica will warm by around 3 °C by 2100 in part because the hole in the ozone layer above the continent, which has shielded the continent from global warming, is healing. See go.nature.com/9BAguz for a copy of the report.

NUMBER CRUNCH

€20 bn

Amount of GDP that will be lost each year by the European Union if global temperatures rise by 2.5 °C

Source: European Commission Joint Research Centre

Healthy carbon cuts: Reining in greenhouse-gas emissions could improve public health, says an international task force publishing in *The Lancet*. The team, led by epidemiologist Andrew Haines at the London School of Hygiene and Tropical Medicine, found that an Indian programme to replace 150 million indoor biomass-burning stoves with low-emissions stoves could, for example, save 12,500 disability-adjusted life-years as well as the equivalent of 0.1–0.2 megatonnes of carbon dioxide per million people in a year (P. Wilkinson *et al. Lancet* doi:10.1016/S0140-6736(09)61713-X; 2009). See go.nature.com/4BcJ6v for more.

● PEOPLE

Early exit: Leszek Borysiewicz (pictured below), chief executive of Britain's Medical Research Council (MRC), will step down in October 2010 to become vice-chancellor of the University of Cambridge, UK. He will leave a full year before his term was due to expire, prompting concerns over the council's future. See page 553 for more.

A. KATRIN PURKISS/REX FEATURES

European reshuffle: Ireland's Máire Geoghegan-Quinn is to be the European Union's next commissioner for research and innovation. Geoghegan-Quinn, appointed on 27 November, is currently a member of the European Court of Auditors. If approved by the European Parliament, she will take up office in January and succeed Slovenia's Janez Potočnik, who will take over the directorate for the environment. The new climate directorate will be headed by Danish environment minister Connie Hedegaard, who will preside over the Copenhagen climate summit this month. Germany's Günther Oettinger has been named as energy commissioner.

THE WEEK AHEAD

5–9 DECEMBER

The American Society for Cell Biology meets in San Diego, California.

➧ www.ascb.org

7–18 DECEMBER

Governments attempt to reach a deal on greenhouse-gas emissions at the United Nations Climate Change Conference in Copenhagen.

➧ http://en.cop15.dk and www.nature.com/roadtocopenhagen

9 DECEMBER

NASA's Wide-field Infrared Survey Explorer is scheduled to launch from Vandenberg Air Force Base in California.

➧ go.nature.com/EDqMqP

Ethics appointments: US President Barack Obama has created a panel to make policy recommendations on bioethical issues. Political theorist Amy Gutmann, president of the University of Pennsylvania in Philadelphia, will chair the Presidential Commission for the Study of Bioethical Issues. Its vice-chair will be materials scientist James Wagner, president of Emory University in Atlanta, Georgia. The commission is expected to produce more policy guidance and be wider ranging than President George W. Bush's bioethics advisory council. See page 553 for more.

BUSINESS WATCH

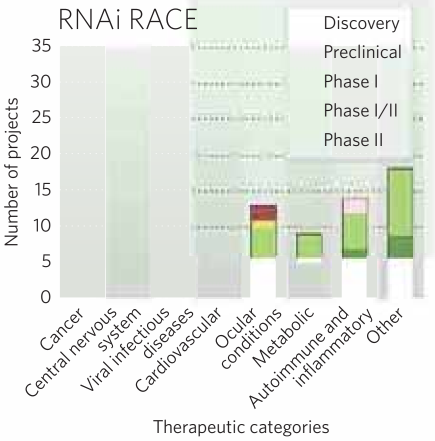
Companies are racing to produce treatments that exploit RNA interference (RNAi), a natural process that silences gene expression in cells. PharmaVision, a biomedical consultancy based in Chichester, UK, estimates that the market for RNAi therapies could be worth more than US$2.9 billion by 2020. But no RNAi therapy has yet completed phase III clinical trials.

Results from a phase Ib trial of an RNAi therapy developed by a consortium including TransDerm of Santa Cruz, California, were published on 24 November (S. A. Leachman *et al. Mol. Ther.* doi:10.1038/mt.2009.273; 2009). A patient treated for pachyonychia congenita, a rare inherited skin disorder, saw some benefit after a series of RNAi injections into lesions on the sole of one foot. But the injections are too painful to treat the disease in the long term and many cancers, for example, cannot be treated by direct injection of an RNAi product.

Delivery is not the only problem facing RNAi. In May, a treatment for wet age-related macular degeneration was ditched by Allergan, a biotech company based in Irvine, California, after the therapy failed to improve patients' vision.

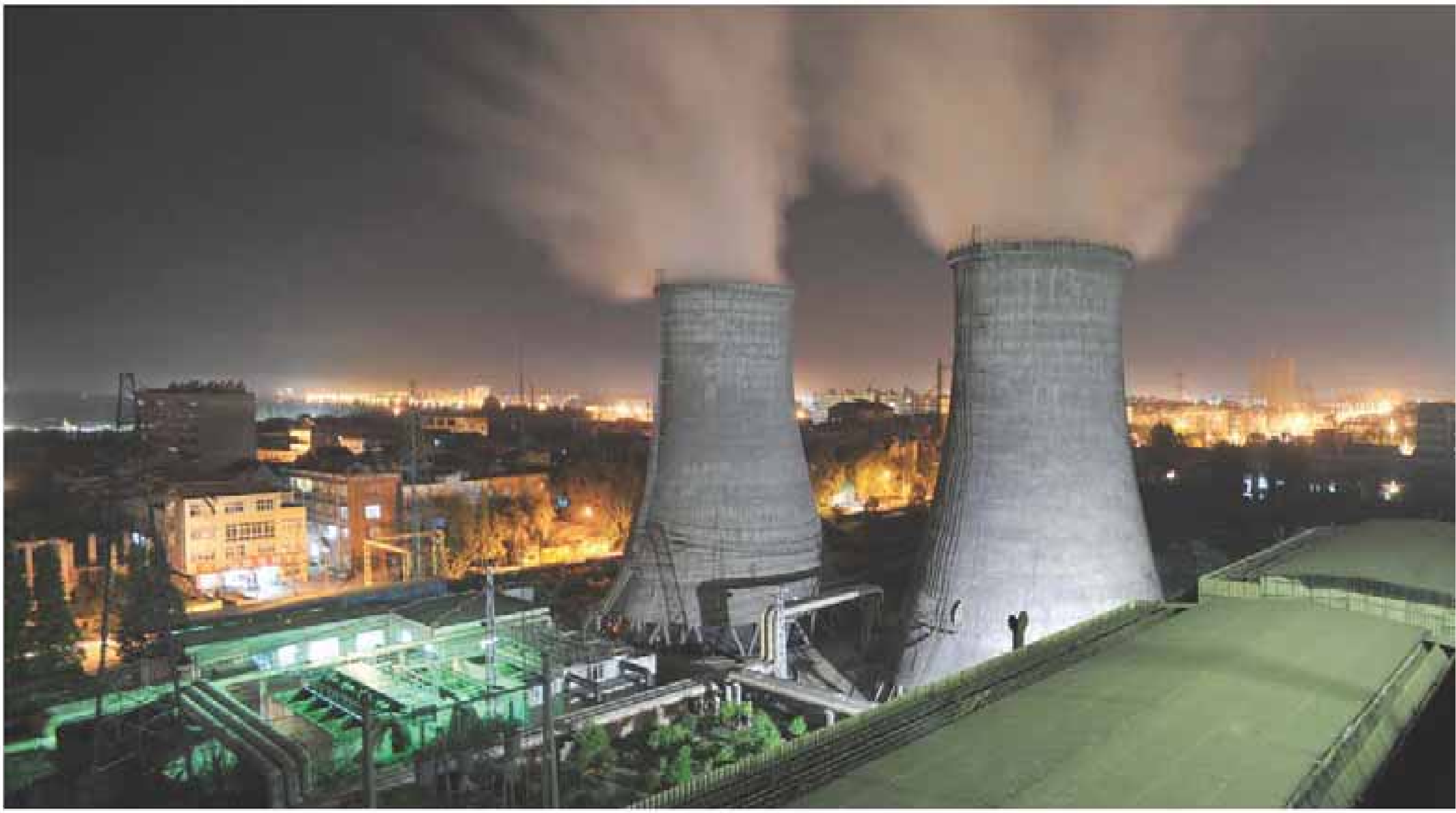
With concern growing over the side effects of RNAi-based treatments, firms must now "conclusively prove the mechanism of action of their products", says PharmaVision consultant Cheryl Barton.

RNAi RACE



SOURCE: PHARMAVISION

NEWS


REUTERS

China's climate target: is it achievable?

The world's top emitter pledges cuts, but "substantial societal reforms" are needed to make them.

Climate analysts are praising China's promise to slash the country's emissions — even as they wonder if the target is achievable or ambitious enough.

Last week, China's State Council announced that the country will cut its carbon intensity — carbon emissions per unit of gross domestic product (GDP) — by 40–45% from 2005 levels by 2020. "It is a very welcome decision," says Fatih Birol, chief economist at the International Energy Agency in Paris. "If the target is met, it would have significant implications for China and the rest of the world."

Yet some think that the target is not far-reaching enough given China's booming economy and its track record of improving energy efficiency. The country reduced its energy intensity — energy consumption per unit of GDP — by 47% between 1990 and 2005, and looks likely to cut it by another 20% from 2005 levels by the end of next year. Carbon intensity can drop faster than energy intensity if clean-energy sources are brought into the mix.

Xie Zhenhua, deputy director of China's National Development and Reform Commission, says that China has picked low-hanging fruit by closing energy-inefficient factories and power plants. In China, industry accounts for an unusually large share — 50% — of energy consumption. "The further we go, the more challenging and costly it will get," he says.

If China sticks to current policies, it will reduce its carbon intensity by about 30% by 2020, says Zou Ji, an environmental economist at Renmin University in Beijing. "To get extra mileage and reach the 40–45% target, China will have to instigate substantial social and economic reforms across the board," he says.

Indeed, the China Council for International Cooperation on Environment and Development (CCICED), a joint Chinese and international advisory board to the state government, recently laid out a road map to a low-carbon economy. It includes recommendations in such wide-ranging areas as energy pricing, industrial development, technological innovation, tax systems, land use and urban planning. "The daunting challenge that China faces cannot be underestimated," says Zou. "The concern is not only whether China is willing to make that step forward but whether its development state will allow a smooth economic transition."

Emissions peak

The new pledge will be included in China's next five-year plan along with policies to help it shift towards a low-carbon economy.

The target puts China on a path for emissions to peak around 2030, says Knut Alfsen, head of research at the Center for International Climate and Environmental Research in Oslo, and an author on the CCICED report. That peak, he says, "will take place at a level where emissions per capita are only half of what we have in the developed world today".

China's announcement came the day after US President Barack Obama pledged to cut his country's emissions by 17% from 2005 levels by 2020. It is the first time that the world's top two emitters have offered specific targets at the same time for controlling their emissions.

"It's very important for the two countries to put numbers on the table," says Jim Watson, a policy researcher at the University of Sussex in Brighton, UK.

In recent weeks other developing countries have made ambitious pledges. South Korea has promised a 30% cut below a business-as-usual scenario, and Brazil at least a 36% cut by 2020. Both would be modest cuts compared with 2005 emissions.

India followed China's announcement by saying it would "be willing to sign on to an ambitious global target for emissions reductions or limiting temperature increase" — but with the catch that "this must be accompanied by an equitable burden-sharing paradigm". China and India, along with Brazil, South Africa and Sudan, last weekend reiterated developing countries' insistence that developed nations help bear the cost of climate change, including facilitating technology transfer (see page 555).

So far, the European Union has pledged the most aggressive emissions cuts in the developed world, of 20% from 1990 levels by 2020, to be increased to 30% below if rich non-EU nations follow suit. The US target announced last week would be equivalent to a 3% reduction from 1990 levels.

Obama must also work with the Democrat-dominated Congress to pass climate legislation that would make its targets binding. The 17% cut he announced last week is in agreement with a bill passed by the House of Representatives earlier this year. The Senate is expected to vote on its own version of climate legislation early in 2010. The Environmental Protection Agency has the authority to regulate carbon dioxide emissions if Congress does not act.

Meanwhile, the Australian parliament has been trying to pass a climate ruling that would cut emissions by up to 25% from 2000 levels by 2020. But on 1 December, the main opposition party elected a new leader who has vowed to oppose the bill, throwing its future into jeopardy. Australia has the highest emissions per capita of any developed nation.

China, if it sticks to its plans, may end up leading the way for developing countries. Compared with current levels, the new target would avoid 1 gigatonne of carbon dioxide emissions — equivalent to a quarter of what the world would need to do to limit global temperature rise to 2 °C over pre-industrial times. "China would champion the fight against global warming," says Birol. ■

Jane Qiu

For more, see News, page 555, and www.nature.com/roadtocopenhagen.

'TEMPLE OF THE MIND' UNLOCKED
Glutamate receptor mapped in detail
go.nature.com/xUG9va

Battle lines drawn over e-mail leak

As the blogosphere continues to buzz with discussion about e-mails leaked from the Climatic Research Unit (CRU) at the University of East Anglia (UEA) in Norwich, UK, climatologists are insisting that the controversy will not discredit their science, or hamper a global climate deal.

CRU confirmed on 20 November that more than 1,000 e-mails and documents had been copied from its servers and distributed on the Internet (see *Nature* **462,** 397; 2009). Since then, climate sceptics have seized on the material, citing the contents of selected e-mails as evidence that the case for anthropogenic global warming has been over-stated, and US Senator James Inhofe (Republican, Oklahoma) has promised an investigation into the affair.

Yet climate experts say the broader impact of the leak will be minimal. "Any suggestions that these e-mails will affect public and policy-makers' understanding of climate science give far too much credence to blog chatter and boastful spin," says Peter Frumhoff, director of science and policy at the Union of Concerned Scientists in Cambridge, Massachusetts.

Some, however, are pointing out that certain e-mails highlight a tendency for scientists to respond to critics either by retreating into an ivory tower, or by attempting to quiet dissenting voices. In an open letter posted on climateaudit.org, Judith Curry, a climatologist at Georgia Institute of Technology, Atlanta, wrote last week: "Scientists need to consider carefully skeptical arguments and either rebut them or learn from them. Trying to suppress them or discredit the skeptical researcher or blogger is not an ethical strategy and one that will backfire in the long run."

The UEA has launched an independent inquiry into both the security breach and whether CRU has dealt appropriately with the deluge of requests for raw climate data it has received under the UK Freedom of Information Act (see *Nature* **460,** 787; 2009). It has also pointed out that more than 95% of the raw data used in CRU climate models has been publicly available for several years.

Quirin Schiermeier

See Editorial, page 545.

What the climate experts say

Thomas Stocker, University of Berne
"Science and science institutions should be transparent, but they are not a 24-hour help service for climate sceptics who lack fundamental scientific and technical skills."

Svend Soeyland, of environment group Bellona Foundation, Washington DC
"Only openness will make the buzz go away. If only the vaguest impression lingers on that studies have been cooked up or that facts have been hidden it will feed conspiracy theories for ages."

Eric Rignot, University of California, Irvine
"Given the overwhelming scientific evidence for climate change, we should deal less and less with climate sceptics. Otherwise we should also deal with folks who think Elvis Presley is still alive, that Earth is less than 6,000 years old and that we cannot possibly have descended from monkeys."

Guy Brasseur, National Center for Atmospheric Research in Boulder, Colorado
"It is important that scientists make their studies completely transparent, but the least ethical way to accuse others is to highlight a sentence and ignore the context in which this sentence has been written."

Mike Hulme, University of East Anglia, UK
"It is possible that climate science has become too partisan, too centralized. The tribalism that some of the leaked e-mails display is something more usually associated with social organization within pre-modern cultures; it is not attractive when we find it at work inside science."

Rajendra Pachauri, chair of the Intergovernmental Panel on Climate Change
"I doubt that negotiations in Copenhagen will be influenced by this unfortunate incident."

CLOCKWISE FROM TOP LEFT: UNIV. BERN; F. BIMMER/AP; U. DAHL/TECHNISCHE UNIVERSITÄT BERLIN; M. TREZZINI, KEYSTONE/AP; JPL/NASA; J. STRAUBE/BELLONA

Spanish awards rekindle old rivalries

Infrastructure programme steers substantial resources to major cities, upsetting some regional centres.

An ambitious effort to develop Spanish universities into campuses that are among Europe's best has stoked some long-standing regional rivalries.

On 26 November, the government announced which universities would benefit from the inaugural round of an annual programme called the Campus of International Excellence, administered jointly by the Ministry of Science and Innovation and the Ministry of Education. The €150-million (US$226-million) scheme is designed to steer resources, in the form of government seed money and loans, to the strategic infrastructure projects with the most potential to aid teaching and research.

But with €73 million of this year's financing going to just five proposals from institutions in Madrid and Barcelona, the awards have already provoked complaints of regional bias in a country that has a long history of wrangling over the way that resources tend to be channelled towards its metropolitan centres.

Following the Ministry of the President's announcement of the competition on 23 July, an initial selection round saw 51 entrants share €53 million in seed funding to develop their proposals. The 18 winners were then divided into three tiers: the five Campuses of International Excellence; four Regional Campuses of Excellence, which share €30.1 million; and nine Promising Projects, mostly awarded between €2 million and €4 million (see graphic).

Regional universities dominate this lowest tier. "We're discouraged," says agricultural engineer Juan Julía Igual, rector of the Polytechnic University of Valencia, which won a total of €8 million for its plan to create a campus based around health and sustainability research. "The goal is positive, but the way it proceeded is unacceptable."


A University of Barcelona project has won the status of 'Campus of International Excellence'.

P. APRAHAMIAN/CORBIS

Julía Igual is particularly critical of the labelling system used to rank the proposals. "Classifying institutions as Promising Projects makes us sound almost like the junior football leagues," he says. "Valencia Polytechnic appears ahead of its Madrid and Barcelona equivalents in international rankings, so I find it hard to believe that they were the only Spanish campuses of international excellence."

Call for transparency

Julía Igual says that in the future, the government should use a more transparent competition system that is based on international ranking models, and it should raise the overall budget. "The funding agencies need to recognize the same reality the international rankings do." However, Ministry of Education spokesman Fernando Herranz counters that "the objective of this plan was not to create another ranking". And responding to the charges of regional bias, he adds that "with more than 50 applicants, not everybody can win".

María José Alonso, vice-rector for research at the University of Santiago de Compostela, is involved in the university's €150-million 'Campus of Life' campaign. The project was awarded €7.5 million, she thinks, because it focuses on the institute's strongest disciplines and draws in its neighbours and other agencies. "It was obvious we weren't going to win it all from this ministry call," she says. Despite the modest size of the award, she says, "we understand it is a stimulus and a stamp of approval from the central government. It will help us negotiate with the regional government for even more support," she says.

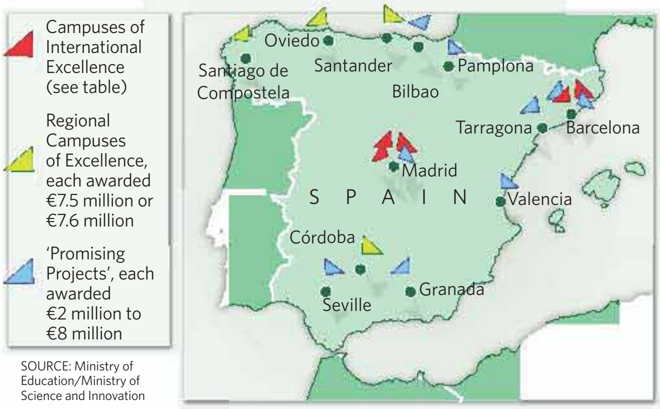
The goal of creating centres of excellence was partly inspired by Germany's Excellence Initiative, a competition between universities to win elite status that recently had its €2.7-billion budget for the period 2012–17 confirmed (see *Nature* **462,** 24; 2009).

But the programmes have different goals, says Matthias Kleiner, who oversees Germany's excellence initiative and is president of the nation's Research Foundation (DFG). Kleiner, who was a member of the international committee asked to evaluate the Spanish proposals, points out that the German initiative focuses on research, whereas teaching is a very important part of the Spanish one. ■

Lucas Laursen

BUILDING EXCELLENCE

Spanish universities in Barcelona and Madrid have taken the lion's share of the funding in the government's plan to create Campuses of International Excellence.



SOURCE: Ministry of Education/Ministry of Science and Innovation

PREMIER PROJECTS

Project	Universities	Financing (€ million)
Barcelona Knowledge Campus	University of Barcelona & Polytechnic University of Catalonia	21
University City in Moncloa: Campus of International Excellence of the Community of Madrid	Complutense University of Madrid & Polytechnic University of Madrid	21
Charles III Campus	Charles III University of Madrid	10
UAB Campus of International Excellence: Betting on Knowledge and Innovation	Autonomous University of Barcelona (UAB)	10
Campus of International Excellence UAM-CSIC	Autonomous University of Madrid (UAM)	11

US bioethics commission promises policy action

Five months after abruptly dismantling the bioethics advisory council left by his predecessor, US President Barack Obama last week created a new bioethics commission that will move beyond the issues that consumed previous panels, such as stem cells and cloning. Based within the Department of Health and Human Services, the Presidential Commission for the Study of Bioethical Issues is explicitly charged with recommending legislative and regulatory action and promises to have more influence on policy.

UNIV. PENNSYLVANIA

Amy Gutmann is to head the US bioethics commission.

Bioethical, social and legal questions relating to genomics and behavioural research are all on the commission's agenda. So are issues of intellectual property, scientific integrity and conflicts of interest in research.

The contrast with the previous bioethics council established by President George W. Bush is stark. Bioethicist George Annas of Boston University, Massachusetts, has described that council, which existed in two incarnations, as having a "narrow, embryo-centric agenda", focusing largely on the research implications of questions such as the moral status of the embryo and when life begins (see *Nature* 431, 19–20; 2004).

In another break with the past, Obama has chosen not to appoint bioethicists to lead the commission. Instead, it will be chaired by political theorist Amy Gutmann, president of the University of Pennsylvania in Philadelphia, and its vice-chair will be materials scientist James Wagner, president of Emory University in Atlanta, Georgia.

Gutmann's work deals with deliberative democracy, and using reasoned argument to depolarize politics. Wagner served at the Food and Drug Administration's Center for Devices and Radiological Health for a decade, and now, as Emory's president, stresses that ethical engagement is integral to the university's strategic vision.

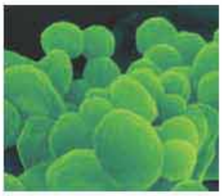
"The appointments of Gutmann and Wagner reinforce the expectation that this commission will seek to provide practical, actionable guidance to the administration and the country," says Ruth Faden, executive director of the Berman Institute of Bioethics at the Johns Hopkins University in Baltimore, Maryland. "This is a wise way to structure the leadership of the commission."

The remaining members of the 13-strong commission are expected to include bioethicists specializing in medicine and law, along with experts chosen from the fields of science, engineering, theology and philosophy. Between one and three of those members will be appointed from the government's executive branch. "These appointments, and the council's place in the executive branch of the government, suggests that it will be more than just a talking shop, with perhaps a significant influence over practice," says political theorist Michael Gottsegen of Brown University in Providence, Rhode Island.

Annas believes that the commission may not be sufficiently independent of government. "Bioethics advisory commissions should be totally free-standing, and not linked to the government and presidential terms, in order to avoid doing 'Republican' or 'Democratic' bioethics," he says.

The commission's wider scope will also force some tough choices in deciding priorities, says Annas. "Doctors' [involvement in] force-feeding prisoners at Guantanamo, doctors and torture, and international human-research rules are pressing issues of our day which demand our attention," he says. Among the other issues he thinks the commission should juggle are new reproductive technologies, an overhaul of informed-consent procedures and — perhaps most immediate — fairer ways to apportion health care. ■

Vicki Brower



BLUEPRINT FOR A BACTERIUM

How *Mycoplasma pneumoniae* does a lot with very little

go.nature.com/woue6a

D. KUNKEL/PHOTOLIBRARY

British medical research chief quits midterm

The early departure of the head of Britain's Medical Research Council (MRC) has prompted concern for the future of the funding body.

Leszek Borysiewicz announced on 26 November that he will quit as chief executive of the MRC on 1 October 2010 — a year before his four-year term was due to expire — to become vice-chancellor at the University of Cambridge, UK. "It's a thrilling and exciting opportunity for me and one I feel I couldn't resist," says the 58-year-old.

The appointment has generally drawn acclaim for Borysiewicz from Britain's biomedical establishment. Richard Henderson, a researcher at the Cambridge-based MRC Laboratory of Molecular Biology who also sits on the MRC Council, believes that Borysiewicz's background and his political acumen will serve him well in his new role.

But for some, there is also anxiety over the future of the MRC. "I think that Borys has done an excellent job," says Colin Blakemore, a neuroscientist at the University of Oxford, UK, and Borysiewicz's predecessor at the MRC. But Blakemore adds that he is "deeply worried about what this might mean for the MRC, especially for the support of basic biomedical research".

Borysiewicz has overseen a major increase in spending, managing a budget that reached £704.2 million (US$1.2 billion) this year. His scientific background, a mix of basic and applied bioscience, has been credited with helping the MRC to increase its emphasis on translational medicine without losing its strength in basic research.

But the future seems less clear. Some believe that the UK government's Department of Health may seek a larger stake in the MRC, pushing it further towards biomedical research and away from fundamental science. There are even worries that the MRC may be absorbed in the Department of Heath, or broken up.

With a general election looming next summer, Borysiewicz's departure "could make the MRC vulnerable at a very critical time", says Blakemore. "It will need a strong new leader, respected by both basic and clinical researchers."

But Borysiewicz says he sees little cause for concern. "The MRC is stronger now than it has been for a very long time," he says. ■

Geoff Brumfiel

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

Technology transfer on the table

Climate summit will seek ways to help developing nations build a low-carbon energy infrastructure.

Cutting global carbon emissions and slowing climate change will require a massive dissemination of clean-energy technology from rich nations to the developing world. And although negotiators remain deadlocked over goals for cutting emissions (see page 550), they are converging on a framework for speeding up the spread of the necessary technologies.

A. DAVE/REUTERS

To tackle climate change, poor countries may need green technologies, such as these wind turbines in India.

Several proposals for this will be on the table when delegates from 192 countries gather in Copenhagen for the climate summit next week, but two elements seem to be gathering momentum in the run-up discussions. The first is the idea of a centralized technology-transfer body under the United Nations climate convention. The second is a network of regional centres, or some kind of technology corps, to help poor nations implement sustainable-development plans.

Both proposals sidestep the issue of access to patented technology, the focus of a long-running dispute between rich and poor nations. Instead, they would help poor countries address an array of mundane but in many ways more pernicious issues, such as energy infrastructure, government policy and workforce development, that hinder their ability to absorb new technologies. Blueprints for a solar thermal power station, for instance, aren't much use without qualified engineers to build and run it and power lines to carry the electricity — challenges even for industrialized nations.

> "We need a proliferation of efforts and institutions, because we don't know what is going to work."

"You have to come back to the basic question about how technology is flowing to the developing world, and it's primarily flowing through transactions within the business community," says Björn Stigson, president of the World Business Council for Sustainable Development in Geneva, Switzerland. Government policies, local economics and workforce issues can all affect those deals, he says. "The biggest bottleneck is availability of human resources."

Exactly how this new framework would function, what kind of authority it would have and how much money it would command remain to be worked out. But poor and rich countries alike support the general idea, says Kunihiko Shimada, a Japanese delegate who has stepped aside as a negotiator in order to co-chair the technology-transfer group at the UN climate talks.

Shimada acknowledges that negotiators in Copenhagen could still get bogged down in the debate over access to the patented technologies found in everything from the latest solar panels to low-emission coal-plant prototypes. Developing countries have to varying degrees called for compulsory licensing, which would force companies to put certain patents in the public domain, or for a fund to purchase patented technologies, which would then be put in the public sphere. Industrialized countries are resisting such proposals, declaring that intellectual property is crucial to driving innovation.

In fact, little is known about how patents affect technological diffusion in the energy industry, when any given 'technology', such as a wind turbine or a clean coal plant, might contain dozens or hundreds of patents, many of which originated in other industries. "We are flying blind," says Bernice Lee, a researcher who has been studying the issue at Chatham House, a think tank in London.

Lee recently headed a Chatham House analysis of nearly 57,000 patents in six energy sectors, which found that the 30 most-cited patents in each sector took two to three decades to hit the mass market. That lag time will need to be halved by 2025 if the world is to meet its climate goals, according to the report. In order to disperse crucial technologies more quickly and widely, Chatham House recommends expanding global demonstration programmes for high-risk sectors such as carbon capture and storage, coordinating technology standards, and accelerating international collaboration on research and development.

All of these functions could be promoted through a technology-transfer body under the UN climate convention, although few think that will be enough to get the job done. "We need a proliferation of efforts and institutions, because we don't know what is going to work," says Michael Levi, a climate expert at the Council on Foreign Relations in New York City. "What we can't afford is to focus on a single mechanism and then find out over the course of a decade that it doesn't work very well."

The developing world, for its part, can't simply wait for the rich countries to equip it for a low-carbon future, says Shane Tomlinson, a researcher with the London-based advocacy group E3G, which promotes sustainable development. One proposal under discussion would require countries to create their own sustainable-development strategies, perhaps in concert with plans for adapting to climate change, to be eligible for international aid. "It's a balanced approach between top–down strategic prioritization and bottom–up low-carbon development plans that is really key to getting the marketplace right," Tomlinson says. "We really do need both." ■

Jeff Tollefson

For more, see www.nature.com/roadtocopenhagen.

Vol 462|3 December 2009

K. CAMPBELL/GETTY

Japan budget threat sparks backlash

Nobel laureates and leading researchers rally to protest at proposed spending cuts.

Japanese scientists have mobilized a huge national and international protest against the budget-slashing policies of their new government.

The criticisms came as government-appointed working groups of roughly 20 people — with few scientists among them — reached the final week of hearings that are recommending budget cuts for 220 government-funded projects, including many major research initiatives. The recommendations are part of the government's effort to reduce next year's national budget by ¥3 trillion (US$33.7 billion).

The proposed cuts would hit, among others, the SPring-8 synchrotron in Harima and a project to build the world's fastest supercomputer. But they also call for reductions in the grants that form the lifeline for many scientists (see *Nature* **462,** 258–259; 2009).

A hastily arranged symposium at the University of Tokyo on 25 November saw four Japanese Nobel laureates and a Fields medalist speak about the potential damage in front of some 1,000 attendees (see go.nature.com/oV9dZ2). At the end of the meeting, the audience erupted into applause to show its support for a proposed statement from the distinguished scientists calling for the government to "take into account the opinions of scientists and academics when deciding budgets for universities and allocations for research grants".

KYODO

Eminent Japanese scientists gathered to air concerns that the country could lose its competitive edge.

In the past week, the normally staid world of Japanese science policy has seen a rash of such statements. On 24 November, presidents from Japan's top nine national and private universities issued a declaration saying the government's policies "are moving in the opposite direction from the rest of the world" and called for funding of young-researcher grants and university-operating costs to be maintained.

> "The weakening of our science and technology means the decay of our country."

The next day, the heads of nine university-related centres that focus on computer and information technology issued a statement calling for support for the threatened supercomputer project. Leaders of 17 Centers of Excellence at the University of Tokyo, along with its president, have released a statement appealing for the maintenance of their budgets. Presidents from Japan's top nine national and private universities and 31 leaders of various Global Centers of Excellence have also weighed in.

Shiro Ishii, an emeritus professor of law at the University of Tokyo who organized the symposium, says that the sudden outcry reminds him of one of the most turbulent periods in Japanese history: the 1960 'anpo' demonstrations, when faculty members and students protested against the US–Japan security treaty.

On 26 November, the four Nobel laureates — chemist Ryoji Noyori, immunologist Susumu Tonegawa and physicists Reona (Leo) Esaki and Makoto Kobayashi — joined a delegation of scientists who met with Japan's Prime Minister Yukio Hatoyama to press their case. During the meeting, Noyori "pointed out that world-class infrastructures such as supercomputers, accelerators and bioresources are absolutely essential to academia and industry to carry out the highest-standard science and technology research", he later told *Nature*.

In a statement issued after the meeting, all eight science Nobelists living in Japan said: "For Japan, a country poor in resources, the weakening of our science and technology means the decay of our country."

"I will firmly support science," responded Hatoyama. "I will take the opinions of the Nobelists into consideration and decide our future direction."

In the last few days of the working-group hearings, which ended on 27 November, the groups gave full support to a prototype Japanese–European fusion reactor, part of the international ITER atomic-fusion project, in line with the Hatoyama administration's enthusiasm for developing new energy sources and green technology.

But a mid-size rocket being developed by a consortium including the Japan Aerospace Exploration Agency faces a suspension of its budget. The working groups also recommended unspecified cuts to the ¥97-billion budget for major university-affiliated facilities, such as the Super-Kamiokande neutrino observatory in Hida and Japan's Subaru telescope at the Mauna Kea Observatory in Hawaii.

Hitoshi Murayama, director of the Institute for the Physics and Mathematics of the Universe at the University of Tokyo, has been particularly active in rallying support from international colleagues. "I've asked distinguished scientists from around the world to write e-mails to the vice-ministers of education and, as far as I know, about a hundred people responded."

George Smoot, a Nobel-prizewinning astrophysicist at the Lawrence Berkeley National Laboratory in Berkeley, California, has been quoted in the Japanese media as saying that the cuts "will send a strong message to the worldwide science community that Japan is no longer a trustworthy partner".

Marine geophysicist Greg Moore of the University of Hawaii laments the proposed 10–20% cut to Japan's ocean-drilling programme, which supports *Chikyu*, the world's largest research-drilling vessel. "I wonder how the Japanese government can justify wasting this technological jewel by further cutting support for its use," he wrote to the science ministry. "The worldwide marine geoscience community is rapidly losing confidence in Japan's ability to lead the Integrated Ocean Drilling Program," he added.

The government is now considering the working groups' recommendations and will unveil its budget by the end of the year. ■

David Cyranoski

Dirty pigs beat disease

D. BURTON/NATUREPL.COM

Outdoor pigs had more 'friendly' gut bacteria than indoor pigs.

Living like a pig could be good for you, according to research showing that dirty piglets pick up 'friendly' bacteria that help them to develop robust immune systems later in life.

The results provide support for the hygiene hypothesis, which suggests that a lack of exposure to microbes in early life can affect development of the immune system and increase susceptibility to certain disorders, such as allergies and inflammatory bowel disease.

Denise Kelly, a gut immunologist at the University of Aberdeen, UK, who worked on the study, says that the results provide the first direct link between early exposure to microbes, immune health and gene expression (I. E. Mulder *et al. BMC Biol.* **7,** 79; 2009). Until now, she says, that link had been circumstantial. "There has been a lot of hearsay around gut microbiota and how it influences immune function and susceptibility to diseases and allergies."

Although many researchers now accept the hygiene hypothesis, says Jean-François Bach, an immunologist at the University of Paris Descartes, there are still questions about how it works, including what role the microbiota has in the gut and how infection helps to protect against disease. "This paper shows that the first days of life are very important," he adds.

The researchers took 54 piglets and divided them equally between an outdoor environment, an indoor environment and isolated conditions in which they were fed antibiotics daily. The scientists then killed piglets on day 5 (neonatal stage), day 28 (weaning age) and day 56 (nearing maturity) to study their gut tissue and faeces.

The study found that 90% of bacteria in the guts of the outdoor piglets came from the phylum Firmicutes. Most of these were Lactobacillaceae, a family of bacteria known for their ability to limit intestinal pathogens such as *Escherichia coli* and *Salmonella* species. By contrast, Firmicutes made up less than 70% of the gut flora in indoor pigs and slightly more than 50% of that in isolated pigs. Pigs from these cleaner environments also had much smaller proportions of bacteria from the Lactobacillaceae.

Kelly's team also found that the differences in gut microbes affected the expression of genes associated with the piglets' immune systems. Animals raised in the isolated environment expressed more genes involved in inflammatory immune responses and cholesterol synthesis, whereas genes linked with infection-fighting T cells were expressed in the outdoor-bred pigs.

Glenn Gibson, a food microbiologist at the University of Reading, UK, says that previous studies have suggested that immune responses are linked to organisms in the gut. "This study takes a step forwards by tallying the gene-expression response into this," he says. However, he adds, because the study was carried out in pigs, there is no way to be certain that the results are relevant to humans.

Jonathan Rhodes, a gastroenterologist at the Royal Liverpool University Hospital in Liverpool, UK, points out that people with chronic inflammation of the digestive tract, known as Crohn's disease, have reduced numbers of Firmicutes, as did the cleaner piglets. But, he adds, patients with Crohn's also have reduced overall bacterial diversity, similar to the outdoor pigs, suggesting that the results might not extrapolate directly to human disease.

Kelly argues, however, that the comparable organ sizes of humans and pigs, and the similarities between the microorganisms found in their guts, makes pigs good model animals for such studies. ■

Natasha Gilbert

Correction

The News Feature 'Reflecting the past' (*Nature* **462,** 30–32; 2009) suggested that wolves from reintroduction programmes might find their way to the Oostvaardersplassen reserve. In fact, the closest wolves are expanding from eastern Europe without human help. In addition, what was described as cattle 'wallows' are better termed 'scratching pits', as the cattle use their hooves to make them rather than rolling on their backs.

The neutrino and the whale

An underwater effort to detect subatomic particles has ended up detecting sperm whales instead. **Nicola Nosengo** reports on a partnership between marine biologists and particle physicists.

To the dock workers and sailors at the port of Catania, in Eastern Sicily, it all looked very suspicious. About once a month during 2005 and 2006, two strangers would walk out to a large wooden cabin at the end of a pier, unlock the door, and remove a small box. Then they would lock up again and disappear until the next month.


G. RICCOBENE

Riccobene: a particle physicist.

The locals had to question what the two men were up to. But when asked, the strangers reassured them that there was nothing to worry about. They were scientists. And the boxes they were retrieving were computer hard drives containing hours of sound data relayed by an underwater cable from microphones — or, more accurately, hydrophones — placed on the Mediterranean sea floor 28 kilometres offshore. Giorgio Riccobene, a particle physicist at the Southern Laboratories of the Italian National Institute for Nuclear Physics (INFN) in Catania, was hoping to show that the hydrophones could be used to detect subatomic particles called neutrinos that had come from deep space. Giovanni Pavan, a marine biologist from the University of Pavia in Northern Italy, was there to help Riccobene deal with background noise in the recordings.

But what Riccobene and Pavan discovered as they listened to their data will bring them back to the port next year with their roles reversed. Then, the physicist will be helping the biologist, and their quarry will not be neutrinos, but sperm whales.


G. PAVAN

Pavan: a marine biologist.

The road to this unexpected destination began nearly a decade ago with Riccobene's involvement in the Neutrino Mediterranean Observatory (NEMO), a collaboration of around 100 researchers from the INFN and other Italian institutes who are hoping to study neutrinos in the ocean. Cosmological neutrinos are constantly streaming through Earth, carrying invaluable information about distant sources such as supernovae. But these fundamental particles have no electric charge and have masses close to zero; they interact with matter so rarely that studying them requires gigantic detectors — the bigger, the better. Hence the NEMO design calls for thousands of optical detectors distributed over 2 cubic kilometres of water, 3,500 metres under the sea at a site off Capo Passero in southern Sicily. The idea is that an incoming neutrino will very occasionally interact with a water molecule, producing a pulse of light that the detectors will capture.

Riccobene was working on a way to enhance the detection. "Theoretically, higher-energy neutrinos should also produce detectable sound waves," he says. "As sound travels better than light in water, an acoustic detector could multiply chances to capture neutrino events." No one knew if this would work. But as the NEMO design includes hydrophones anyway — they are needed to position the optical detectors— Riccobene was asked in 2002 to supervise a feasibility study called the Ocean Noise Detection Experiment (ONDE), which would be located at the project's 2,000-metre-deep test site east of Catania.

Noise control

To educate himself, Riccobene went to Paris for a workshop about acoustic neutrino detection, and immediately noticed something missing from the talks. "Background noise was not even mentioned," he recalls. "Everyone was taking for granted that at great depths it would be very low, but there were no published data." Riccobene went back to Catania, just in time to discover that a local environmental group was hosting a talk by Pavan, who had pioneered the digital recording of sea-mammal sounds in the early 1980s, and who was acknowledged as one of the world's leading experts in the field. He was obviously the right man to answer Riccobene's question: how high would background noise be at a depth of 2,000 metres?

With little data to rely on, Pavan had no simple answer. "Systems to record at great depths were simply not available until a few years ago," he says. About all he could say for sure was that deep waters were not nearly as silent as the neutrino physicists were assuming.

"At first I was appalled," Riccobene says. The noise levels Pavan estimated were well above the expected level of a neutrino event. That did not necessarily make neutrino detection unfeasible, he says. But it did mean that the NEMO team couldn't hope to isolate the neutrino signals until it had an accurate survey of the background noise it would have to filter out.

Riccobene invited Pavan to join the ONDE team on a long-term monitoring project of the Sicilian seabed soundscape — the first ever attempted at such depths. Pavan had no funds to support his participation, but accepted anyway. Riccobene would give him access to depths he could never reach otherwise, allowing him to study the largely unknown acoustic environment of the deep sea. Pavan particularly hoped to measure the level of sound pollution there, as it is a potential cause of stranding for many deep-diving whales — whose vocalizations he also expected to hear in the recordings.

By January 2005, Riccobene and his team had positioned four high-sensitivity hydrophones at the NEMO test site and had laid an optical data cable back to that cabin on the pier in Catania. Soon after that Riccobene and Pavan were obtaining data. And in April 2005, Pavan began listening to the first recordings.

As he predicted, Pavan could hear low, uniform background noise, mostly caused by natural water movement and ship traffic, plus an occasional burst of identifiable sounds: the propeller of a large ship, a sonar impulse, even some explosions. But what captured his attention were short, regularly repeating sequences of 'clicks' — the signature sounds made by sperm whales compressing air through their respiratory system. "They probably use them to estimate depth and to locate prey, measuring their echoes more or less like bats do," Pavan says. Hearing clicks every now and then was not surprising: they are among the loudest sounds produced by any animal, and can travel up to 20 kilometres in water. What was surprising was that the clicks kept appearing in the recordings month after

Equipment designed to detect subatomic particles called neutrinos has picked up sperm-whale communication.

W. POELZER/PHOTOLIBRARY.COM

month. "Sperm whales are considered very rare in this area," Pavan says. "Published data hint at a very sparse population." But these studies are usually based on sightings or sound recordings taken near the sea surface. The ONDE recordings came from the deep waters where sperm whales dive for food and spend most of their time, suggesting that the creatures could be much more abundant in this area than previously thought.

Riccobene and Pavan kept recording until November 2006, when the hydrophones were removed and replaced by a prototype of the optical detection system, which was connected to the same cable. By that point the two scientists had collected more than 600 hours of recordings, and although they hadn't detected any neutrinos, they were satisfied. "My interest at the time was in proving that deep underwater recording was feasible, and could produce data of good quality," Riccobene says. "I did not expect to detect neutrinos at this stage." At the same time, Pavan now realized he could get much more from the data than a simple list of background noise sources.

Yet hearing sperm whales is one thing. Counting them and compiling reliable statistics is another. In principle, the acoustic properties of a click can give an idea of the animal's size and sex. And if the click's reflection from the sea surface also shows up in the recording, it is possible to estimate the animal's position and its ascending or descending trajectory. "But either you develop a software algorithm to do this, or you do a long and awful hand count," Pavan says — and the latter is exactly what he and his colleagues have been doing since 2006, as they had no funding to pay for such algorithms.

Still, Pavan says, a statistical picture has slowly begun to emerge. Sperm whales appear in half of the recorded days, something existing population estimates cannot account for. Pavan and his colleagues have also been able to detect seasonal patterns and hints of social behaviour in the recordings — data they presented at a conference in Pavia in September. "Whales are more frequent in spring and autumn, with more animals moving at the same time, probably belonging to the same school," Pavan says. Sometimes the recordings contain 'codas', brief sequences of clicks with a peculiar pattern. These sounds are thought to have a social function, as they are emitted only when males, normally solitary, gather around female groups. Codas seem to work as local dialects, with different populations using different patterns. "The most frequent pattern in the Mediterranean is the 3 + 1 type," says Pavan, "a rapid succession of three clicks and then an additional one." But the recordings show that a 2 + 1 type is more frequent than expected, which could hint at whales in transit from outside the Mediterranean basin.

Deep-down disturbances

Meanwhile, Riccobene, by now an enthusiast of bioacoustics almost as much as of neutrino physics, has managed to set up a continuation of the study by involving the European Seas Observatory Network (ESONET), a European collaboration that is developing a network of deep-sea monitoring stations in the Mediterranean. Although ESONET's focus is on geophysics and climatology, it has agreed to finance a new incarnation of the ONDE platform called LIDO (Listening Into the Deep Ocean). LIDO will last for three years, starting next March when an array of four hydrophones will once again be deployed off Catania to listen for whales. This time, continuous recording sessions will last for up to one year, and algorithms will be used to choose which data to keep and which to disregard. Detailed comparisons of recordings from the four hydrophones, placed on the vertices of a square, will allow researchers to determine each detected animal's size, speed and direction, and give a more precise estimate of sperm-whale populations and seasonal habits. In the future, the project envisages the deployment of similar stations at other locations, including the Gulf of Cádiz, close to the Strait of Gibraltar. "One of our key problems is to understand whether Mediterranean sperm whales are a closed population, or whether there is an exchange with the oceanic population," Pavan says.

Riccobene and his colleagues are also continuing their work on the final, large-scale NEMO observatory. If all goes as the developers hope, it will be deployed before the end of the next decade off Capo Passero. The larger project will include an acoustic system, at least for positioning purposes if not for detection. Riccobene now knows there is plenty of background noise in deep waters, but he has decided that he needs more data to work out whether he can detect neutrinos acoustically, and that depends on winning more funds. And NEMO's data will be shared with biologists, seismologists and any other interested researchers. Pavan will surely be among them, eager to study not just sperm whales but also fin and beaked whales, whose low frequencies are best heard at greater depths.

"Sea floors at such depths are among the least-known regions of the planet, and there is a lot down there for many researchers," Riccobene says. "Once we put a broadband connection 3,500 metres under the sea, I have a feeling people will queue up for the data." ■

Nicola Nosengo is a freelance science writer based in Rome.

Flies on film

A unique collaboration is bringing automated screening to the study of fly behaviour and could change the way that machines see humans. **Lizzie Buchen** reports.

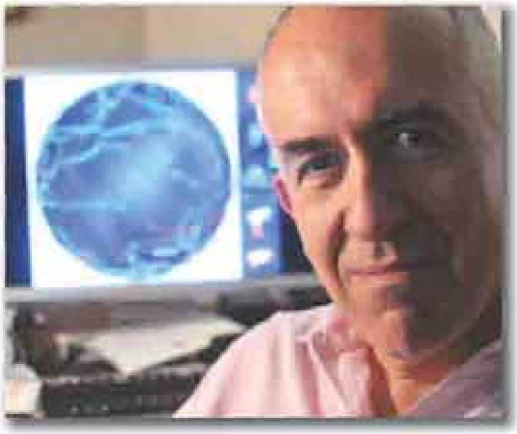
At full speed, the altercation would have looked like nothing — a brief contact, fractions of a second long, between two flies. But slowed to 1/20th of normal speed it has all the flash and dazzle of an elaborate professional wrestling move. Biologist David Anderson calls the grainy, black and white video on his computer the "fly lucha libre".

One fly, a male, rears up and clamps down on his neighbour. Flipping backwards he whips his opponent into the air, executing two somersaults in the process. The victim, helpless, flaps its wings as it crashes to the floor, then rights itself and flies away. Take-downs like these are just one of dozens of behaviours that Anderson studies at California Institute of Technology (Caltech) in Pasadena (see the video from Martin Heisenberg at the University of Würzburg in Germany at go.nature.com/o8sRLs).

"It's this whole world you'd have no clue existed," says Anderson's collaborator, Pietro Perona, whose forays into this world have been relatively recent. For 20 years the computational-vision scientist, also at Caltech, has been trying to develop machines capable of detecting and interpreting complex behaviours in humans. It's an ambitious goal, but he has found unlikely allies in Anderson and behavioural neuroscientist Michael Dickinson. Since about 2005, the three have collaborated to create tools that combine the ease of manipulating the fly nervous system with ways to automatically track complex social behaviours such as aggression and courtship. The hope is to understand their neural underpinnings, which may inform studies in humans.

This year, the group demonstrated two systems that can do this[1,2]. Anderson is now using them to screen thousands of lines of mutant flies to understand the genes and neural circuits that control behaviours, and Dickinson is exploring the social dynamics of flies in large groups — something that has been nearly impossible to approach in the lab.

"This is a very tough problem," says Joel Levine, a neurogeneticist at the University of Toronto in Mississauga, Ontario, who has been using one of the software programs, called Ctrax, since November 2008. "I've been trying for a very long time to get people in my lab to write this kind of software, and no one could do it."



Pietro Perona uses machine vision to track flies.

B. PAZ

Because the programs are freely available, fly researchers from around the world are using the new technology and taking off in their own directions. And through the flies, Perona is inching closer to his goal of building machines that can interpret human behaviour.

The tools promise to be a boon for the field of fly behaviour, and especially its students, technicians and postdocs who perform a thankless job: watching videos frame by frame, tallying every lunge, timing every chase, classifying every wing threat and repeating.

"It's mind-numbing to sit in front of videos and score them by hand," Anderson says over lunch with Perona and Dickinson at the Caltech faculty restaurant. It's not just a matter of saving time, Perona adds. Accuracy counts, yet observing flies can be a subjective exercise. "Suppose we use the clipboard and stopwatch method. We have a postdoc at Caltech who does it, and another postdoc in Norway who does it." If results don't agree, getting the answer is next to impossible: "Those postdocs are gone, they've got jobs somewhere else, so we're left in this situation of irreconcilably different opinions — opinions that never become fact because there is nothing to ground them on."

Big brother

Perona had been looking to develop machines that can understand peoples' actions and intentions by watching their movements. Machines with such capabilities would have an array of applications. Imagine workplace safety cameras that can warn when a machinist is dozing or a security camera at an airport that flags suspicious behaviour.

To build these systems, machine-vision scientists train computers using a common language: labelled data. Annotate thousands of examples of pedestrians stepping on to a crossing, for example, and one can produce an algorithm that could identify that behaviour in novel situations.

But simple, specific behaviours are of limited use when the goal is to recognize behaviours that would qualify as 'suspicious' or 'dangerous'. Perona is approaching this

E. HOOPFER

Pro fighters: fruitfly behaviour is more complex than many assume.

problem by trying to break down behaviour into its component parts.

Pouring a glass of wine, for example, involves gripping the bottle, lifting it, moving it towards a glass and finally tipping it down and up. Perona has dubbed such elementary motions ‘movemes’[3], and they form the basis of his behavioural hierarchy. Movemes combine into an action, such as pouring wine. An action that is prolonged in time is an activity, such as having dinner. But testing such a theory, he says, requires enormous sets of labelled videos — videos that are hard to come by.

“You just can’t get enough data,” Perona says as he counts off the reasons on his fingers. It’s illegal to tape without permission, difficult to get approval and then harder, with permission granted, to get natural, candid activity. “Also humans are boring,” Perona says. “How much time do you have to spend in pubs before you finally see a fight?”

Perona realized that he could learn a lesson from the life sciences. “When biologists encounter complicated issues such as schizophrenia, what do they do?” he asks. “They use a model organism. Something that lets them do very intrusive, invasive experiments and get answers much more quickly.”

A few years ago, he started asking his colleagues at Caltech about their model organisms, searching for one with behaviours that were interesting, but not too complex. Meanwhile, Anderson had just come back from a sabbatical where a lab was using an off-the-shelf tool to measure how much flies were moving around.

“It worked, but it was clunky,” Anderson remembers. Dickinson suggested that he should build something himself. So Anderson approached Perona with descriptions of the flies’ intricate behaviours. Perona was intrigued. “I didn’t know flies do more than just go buzz and push themselves through the air,” he says.

High ambitions

Anderson wanted a program that could automatically detect specific behaviours during aggression and courtship — counting and timing wing threats, chases, wrestling bouts and copulations. He wanted to be able to run many pairs of flies at the same time and have the machine spit out when, where and in what order every action occurred.

Dickinson, who has devoted most of his career to the aerodynamics and biomechanics of fly flight, wanted to understand how flies behave in groups of dozens to hundreds. To do this, he would need to keep track of each fly’s movement and maintain their identity over long periods of time — something no one had done before. There were a couple of programs that could follow flies, but they had an ‘occlusion problem’. Whenever two insects touched or entered into the same pixel, an experimenter would have to manually tell the machine how to separate the flies out.

Both had tall orders for Perona. “But that’s the great thing about collaborations,” Anderson says. “Something that to biologists looks nearly impossible may be in the realm of achievability for someone like Pietro.”

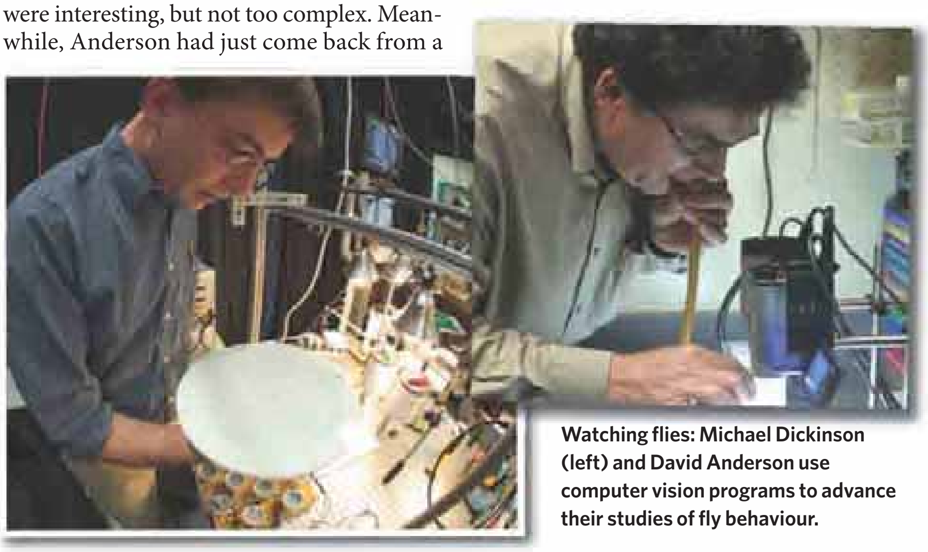
Anderson’s specialized behaviour detector, and Dickinson’s exploratory tracker required very different software, but the tasks seemed simple enough. Perona shared one postdoc with Dickinson — Kristin Branson — and another with Anderson — Heiko Dankert. Perona estimated that each program would take a matter of months. “Of course it ended up taking three years, but that’s fine,” Perona says. Much of the time was spent making it understandable to biologists and also robust enough to work reliably in the hands of others, what Dickinson calls the “fender and body work”.

Meanwhile, Dickinson was tinkering away, trying to create an environment for Branson’s software — Ctrax — to analyse. The product was a circular arena just large enough to hold a football, surrounded by eight halogen lights.

Inside, up to 50 flies walk around; their wings clipped to prevent them from taking off and confusing the tracking system. A camera hangs directly over the centre of the arena, capturing the movements and relaying them to a nearby computer. On the monitor, each fly leaves a thin, brightly coloured trail as it explores the arena. After a few minutes the screen is so clogged with lines that it looks like a ball of multicoloured yarn.

In essence the program doesn’t do much more than follow flies around. The idea was to create an all-purpose tracker and quantifier that would let the experimenter decide which behaviours were interesting — and to tell the machine not only what to look for but how to learn what to look for.

“This is one of the most exciting aspects



P. PERONA

B. PAZ

Watching flies: Michael Dickinson (left) and David Anderson use computer vision programs to advance their studies of fly behaviour.

Automating the analysis of fly behaviour could lead to machines that identify potential safety risks.

P. DOLLÁR, P. PERONA

B. PAZ

of it, this discovery phase of research," says Dickinson. Pietro describes one intriguing behaviour they identified. "So there was one fly going slowly, and another fly was overtaking it," Pietro says, demonstrating the motions with his fingers. "As the faster fly was overtaking the slow fly, the slow fly stops. Now it could have been a fluke that the fly decided to stop, but so then Michael goes forward and he finds another example of the same thing happening. And so, now we can very quickly program the software to pull out every encounter of flies going like this." All they have are hunches as to why the flies do it, says Perona.

But once an experimenter recognizes an interesting sequence of movements or interactions, he or she can propose hypotheses and answer them with a new analysis — no new data required. The program will then generate quantitative descriptions of each behaviour for each individual fly. These 'ethograms' allow the user to discover and quantify subtle behavioural differences between populations of flies, such as males versus females or flies with genetic mutations, and even between individuals within a population — fly personalities.

Dickinson's group has already made some novel observations using the ethograms. Male flies, for example, are not shy — they walk close to one another, inspecting, even nudging each other from time to time. But females need their personal space. They obey certain rules of etiquette when they meet each other and when they walk around.

His group is also starting to explore how social interactions change in more complex environments, placing cones and obstacles on the landscape. He has found that flies tend to climb to the top of the cones, and chase each other off in 'king of the mountain' fashion. With these new observations come new hypotheses. Dickinson says the behaviour might have something to do with maintaining a good vantage point for identifying food sources.

In a lab a few hundred metres away, Anderson is using his program, called the Caltech Automated Drosophila Aggression-Courtship Behavioral Repertoire Analysis (CADABRA), to take a quantitative look at behaviours he already knows are there — wrestling, tackling and other manoeuvres that flies attempt during aggressive bouts, and the serenades and dances that lead to mating.

He is now developing a screen in which he has genetically activated and silenced different populations of neurons. These selective disruptions will allow him to dissect the neural circuitry of each behaviour — not just aggression as a whole, but the frequency of each component of the behaviour, and even the frequency at which one behaviour leads to another.

In his set-up, a pair of flies sits in a shallow well about as wide as a golf ball. A plate about a centimetre high holds 12 such wells. Hanging above, a camera films all 12 interactions in parallel. The video can then be run, frame by frame, through CADABRA, which then produces statistics on every behaviour.

CADABRA's ethograms, for example, illustrate the frequency of each lunge and chase, as well as the frequency of transitions — such as how often a chase is followed by a lunge versus a flirtatious wing extension.

"To do this by hand would take 270 person hours," Anderson says. "We basically did it in 20 minutes."

Flight club

Levine learned about Ctrax when he and Dickinson were lecturing at the Marine Biological Laboratory in Woods Hole, Massachusetts, in the summer of 2009. He has now configured the system so that it can follow individual flies in a group over a period of days, without any interference.

"So little is known about group dynamics," he says. "Who interacts with whom? Why do some flies mate more than others? How do social interactions affect the circadian rhythms? This has been very difficult to study. And now along comes Michael with this software, and, finally, we can ask these questions."

Leslie Vosshall, a neuroscientist at Rockefeller University in New York, is using Ctrax to study courtship. She is using a modified set-up, designed in Dickinson's lab, in which the flies wander around in a dish-like arena covered by a plastic lid. This allows them to move about with their wings intact — a crucial modification, because of the involvement of wing signals and wing songs in courtship.

And Michael Reiser — Dickinson's former graduate student, now at the Howard Hughes Medical Institute's Janelia Farm Research Campus in Virginia — is taking advantage of Ctrax's flexibility for his 'Fly Olympiad'. He is putting mutant flies through a barrage of behavioural tests — from visual reflexes to walking behaviour to odour sensation.

Perona says he's pleased that his work has started to bear fruit for biologists. And Anderson's and Dickinson's projects are producing the videos of labelled behaviours of which he had been starved.

The translation of fly behaviours into ethograms is also giving Perona his first opportunity to test his hypothesis of behavioural hierarchies — in which a continuous, extended activity such as courtship can be broken down into actions, which can further be decomposed into elementary but meaningful motions.

He hopes the concept will form the basis of an overall computational theory of behaviour. But although the simplicity of fly behaviour allows Perona to make conceptual progress, it can only take him so far. "Am I closer to building a machine that can figure out what people are doing? Yes and no. Flies have a much simpler repertoire of behaviour than humans and they don't waste time." Watching fly behaviours should allow him to go back to human data with more solid ways of testing his hypotheses. But Perona's goals have also been evolving. He says he has become increasingly interested in fly behaviour for its own rewards. "If it all ended up with me having contributed something to biology, it would still be fantastic," he says. ■

Lizzie Buchen is an intern with *Nature* based in Washington DC.

1. Dankert, H., Wang, L., Hoopfer, E. D., Anderson, D. J. & Perona, P. *Nature Meth.* **6,** 297–303 (2009).
2. Branson, K., Robie, A. A., Bender, J., Perona, P. & Dickinson, M. H. *Nature Meth.* **6,** 451–457 (2009).
3. Del Vecchio, D., Murray, R. M. & Perona, P. *Automatica* **39,** 2085–2098 (2003).

A tale of two sciences

An innovative approach to reducing toxic-chemical use scrambles to stay alive as big science prospers, says **Daniel Sarewitz**.

The main laboratory of the Toxics Use Reduction Institute (TURI) at the University of Massachusetts Lowell looks more like a 1940s factory machine shop than a state-of-the-art research facility. Yet despite its low-tech feel, TURI could revolutionize the way the United States deals with chemicals in the environment.

Regulation of toxic chemicals is supposed to be based on science, mainly using epidemiological and animal-model approaches to assessing risk. Although Europe recently enacted a comprehensive regulatory framework for chemicals (called REACH, for Registration, Evaluation, Authorisation and Restriction of Chemical substances), the United States remains gridlocked in an adversarial system that pits those with an interest in using a particular chemical — industry and its allies, for the most part — against those who want to get rid of it — environmental groups and their allies, for the most part. Technical experts line up on either side, and the often fuzzy demarcation between science and politics seems to vanish. The result is a morass of litigation, politics, science and uncertainty, in which debates over how to regulate some chemicals drag on for decades while the backlog of unevaluated substances grows unabated.

TURI offers a different path. The idea is this: because chemicals are valued for their functionality, the sensible procedure is not to ban or restrict toxic compounds, but to replace them with safer ones. At the core of this approach is a different role for science: not to reduce uncertainties about risk or to explain causal mechanisms, but to offer solutions by showing that a function served by a toxic chemical can be equally well served by a less toxic one.

For example, the US Environmental Protection Agency regulates the solvent trichloroethylene, but its health effects remain disputed and it is still widely used. The standard scientific approach would be to do more research on how trichloroethylene behaves in groundwater and in humans to reduce risk uncertainties before tightening the regulatory noose.

Instead, TURI found alternatives to the compound, such as non-chlorinated solvents with no known health risks, and water-based, ultrasonic cleaning processes. TURI researchers tested the substitutes for effectiveness and developed cost–benefit estimates. They worked with small firms to understand barriers to adoption, and cooperated with state agencies and professional organizations to demonstrate the replacements. The result: a 90% reduction in trichloroethylene use over two years.

WORLD VIEW

TURI thus turns adversarial regulation on its head by making firms that use toxic chemicals into constituents for safer chemicals. It evades endless debates over uncertainty by focusing on finding solutions rather than diagnosing problems. This type of research does not generate many high-prestige publications or huge federal grants, but between 1990 and 2005 TURI helped Massachusetts firms to reduce toxic chemical use by 40%, and chemical waste by 71%.

The heart of TURI is clever science policy. In 1989, efforts in Massachusetts to pass legislation regulating toxic chemicals stalled because of industry opposition to chemical bans. The impasse was broken when Ken Geiser (then a professor at Tufts University, but soon to move to Lowell as TURI's first director) proposed that, rather than imposing bans, the law should require firms to develop plans for reducing their use of toxic chemicals. Geiser attributes this insight to his education in architecture: "Architects understand the role of a plan in diffusing information and focusing on solutions."

Geiser's approach was adopted in the 1989 Massachusetts Toxic Use Reductions Act, which requires firms to report their use of toxic chemicals and to develop plans for reducing this use. TURI was set up to provide technical and organizational support. Being university-based, the institute was able to maintain political independence, technical legitimacy and organizational flexibility. Its funding came from a levy on the 500 or so firms in the state that use toxic chemicals. TURI was therefore insulated from political whims and economic pressures; the state legislature merely allocated to it the money paid by the firms to the state.

Until this year. As the Massachusetts legislature confronted the state's budget crisis, lawmakers grabbed at every straw they could find. On 30 June, the final day for completing the fiscal 2010 budget, legislators swept TURI's modest $1.5-million allocation into the state coffers. In July, university leadership gave the institute four months to find alternative support or close down.

Meanwhile, it was the best of times for big science, as the US government pumped more than $20 billion in stimulus funds into federal research agencies. In early August — at about the same time that lay-off notices went out to TURI employees — presidential science adviser John Holdren and Office of Management and Budget director Peter Orszag sent a memo to federal agency heads emphasizing the need to "develop outcome-oriented goals for their science and technology activities ... and target investments toward high-performing programs".

Yet shoving billions into existing institutions will produce more of what society already has, regardless of whether that's what it actually needs. So, on 28 October, the National Institutes of Health announced a $30-million stimulus grant to study the health effects of bisphenol A, a chemical used in plastics production. No doubt many peer-reviewed publications will result — but substitutes for plastics containing bisphenol A are widely available, and children's products containing it are already being phased out.

TURI, by the way, was saved — at least for a year. As imperial science lapped up its windfall billions, university officials worked to secure stimulus dollars for TURI from a US Department of Education programme aimed at preserving jobs in state educational institutions. On 16 September, two weeks before the plug was to be pulled, TURI announced that it would receive $1.3 million for fiscal 2010, enough to keep most of its operation alive. Geiser is now organizing firms that pay money into the programme to put pressure on legislators to restore funding in future years. The fact that these companies support a programme to reduce toxic-chemical use shows how radical TURI is, and how far beyond mainstream science institutions it has ventured.

US science policy is based on the idea that more money is the best route to more social benefit. TURI teaches us otherwise. ■

Daniel Sarewitz is co-director of the Consortium for Science, Policy and Outcomes at Arizona State University, based in Washington DC. His column appears in the first issue of each month.
e-mail: dsarewitz@gmail.com

CORRESPONDENCE

Carbon emissions: the poorest forest dwellers could suffer

Debate on the carbon-credit system known as REDD ('reducing emissions from deforestation and forest degradation') has focused on technical and methodological obstacles and on sourcing carbon finance. The impact of the system on the world's 350 million tropical forest dwellers calls for closer scrutiny.

Without careful planning, REDD stands to create large numbers of 'carbon refugees' as governments curb financially unrewarding deforesting activities such as those of small-scale agriculturalists and fuel-wood harvesters, who mostly pay no taxes on what they produce. Forest dwellers could become excluded from their means of subsistence to preserve carbon.

A similar situation has occurred during previous attempts to conserve tropical forests. Last year I worked in Liberia's forests bordering Ivory Coast, and heard of park guards in the Tai National Forest, a well-protected Ivorian biodiversity conservation area, shooting local hunters dead. I met Ivorian subsistence hunters, excluded from their ancestral lands, relocating to Liberia to maintain their livelihoods. The journal *Conservation & Society* is investigating the possible displacement of thousands of people in Africa by biodiversity conservation projects.

The Centre for International Forestry Research has shown that forest-based sources of income generated by local communities are often rendered illegal by forest law. Crackdowns tend to target the poor, rather than the criminal networks behind the estimated 50% of global tropical-timber exports that are illegal. A REDD-inspired redoubling of current efforts at law enforcement would further victimize forest-dependent peoples.

Forest dwellers should be seen as an important part of the solution to deforestation. Evidence from 80 forest commons in 10 countries shows that community ownership, larger forest areas and a high degree of community autonomy in decision-making are all associated with both high carbon storage and livelihood benefits. Conversely, local users with insecure property rights extract resources at unsustainable rates (A. Chhatre and A. Agrawal *Proc. Natl Acad. Sci. USA* **106,** 17667–17670; 2009).

Extending legal collective property rights to forest users over large areas, combined with forest-encroachment monitoring by independent scientists and local agencies, could reduce deforestation without human rights violations. This plan may substantially reduce deforestation by cutting off the supply of 'empty' land for outsiders to deforest.

There is good will on the ground for REDD to work, with safeguards. With transfers of US$10 billion a year under discussion, the REDD agreement should ensure that at least 50% of carbon payments go directly to forest dwellers, and that their property rights are assured. Otherwise, some of the world's most marginalized people will end up paying a high price for reducing carbon emissions.

Simon L. Lewis Earth and Biosphere Institute, School of Geography, University of Leeds, Leeds LS2 9JT, UK
e-mail: s.l.lewis@leeds.ac.uk

Carbon emissions: dry forests may be easier to manage

You discuss in an Editorial (*Nature* **462,** 11; 2009) the promise of the emissions trading scheme REDD, whereby tropical countries will be rewarded for increased sequestration by forests. But it is important for negotiations to focus on the realities rather than on the ideal.

REDD countermeasures to tropical deforestation will affect food supplies and employment and will increase prices of forest products. They are likely to be contested by the powerful political forces that control logging, ranching, plantations and agricultural expansion in rainforests. It may therefore be premature to expect deforestation to be significantly reversed in the short term under REDD.

It may be easier to manage the politics and economics of emissions from degradation (that is, the thinning out rather than clearance of forest) in the world's dry forests and savanna woodlands. This type of degradation results primarily from the exploitation of forest by local communities as part of their livelihood. It has been tackled successfully in Nepal, India and Tanzania, for example, under programmes that promote community forest management.

Dry forests do not have the international status of the majestic Amazonian and Congo forests, and the dry-forest degradation option is relatively neglected in REDD debates. Although the carbon content of dry forests is considerably lower per hectare, more of their area is degraded because they are more densely populated. Carbon losses may be more easily adressed because the commercial value of dry forests is lower, and their use is not so contested.

REDD should give more attention to dry forests. It should strengthen local communities' rights to manage the forests and to hold tenure there. Accounting the carbon savings would depend on proper estimates of local dry-forest degradation rates (which are virtually unknown at present) and on devising ways to monitor the gradual increase in carbon stock resulting from community forest management.

Margaret Skutsch, Michael K. McCall CIGA, Universidad Nacional Autónoma de México, Morelia, CP 58190, México
e-mail: mskutsch@ciga.unam.mx
Jon C. Lovett CSTM, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands

King Canute and the wisdom of forest conservation

Copenhagen is where the world's nations are meeting this month to attend the all-important climate summit. It was also the capital of King Canute's empire and, by a quirk of history, it was Canute who drafted the first forest-conservation legislation almost 1,000 years ago, as king of England (John Manwood *A Treatise of the Lawes of the Forest* Societie of Stationers, London; 1615).

Manwood's book proclaims "Carta de Foresta, of King Canutus … in the yeare of our Lord 1016" followed by the list of Canute's 34 Forest Laws. These were aimed at conserving forests at all costs, and especially their wildlife, with heavy penalties for offenders.

Law 28, for example, includes "No man may lay his hands upon the Kings demesne Woods without licence of the Verderor: for if he does, hee shal be aiudged guilty of the breach of the Kings free chase Royal."

Perhaps each nation attending the Copenhagen conference might agree to create its own new forest — thereby preserving endangered species, combating global warming and honouring Canute's apocryphal command to the seas to stop rising.

Roger Short Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Victoria 3010, Australia
e-mail: r.short@unimelb.edu.au

For more about limiting emissions, as the Copenhagen conference approaches, see pages 550, 555, 568 and 570 and http://go.nature.com/sRCuKV.

Contributions may be submitted to correspondence@nature.com. Please see the Guide to Authors at go.nature.com/cMCHno. Science publishing issues are regularly featured at http://blogs.nature.com/nautilus.

OPINION

No quick switch to low-carbon energy

In the first of two pieces on reducing greenhouse-gas emissions, **Gert Jan Kramer** and **Martin Haigh** analyse historic growth in energy systems to explain why deploying alternative technologies will be a long haul.

To combat climate change, the world's entire energy system needs a major overhaul before the middle of the century. But can we build new energy supplies that quickly? Some argue that with the right incentives we can see similar rates of change in the energy system as have been seen in information technology. So most of the debate focuses on how much the transition will cost and who will foot the bill. Here, we argue that cost is less important than the rate at which existing low-carbon energy technologies can be physically deployed. Because the scale of the energy system is so huge, it takes time to build the human and industrial capacity to achieve substantial deployment.

There have been high-profile proposals to 'repower' the world in a decade, loosely based on the way innovative consumer goods such as mobile phones or iPods conquer their markets[1,2]. Unlike with consumer goods, we believe that there are robust empirical 'laws' that limit the build rate of new and existing energy technologies and thereby the potential to deliver much of the hoped-for transformation by 2050 (ref. 3). To accelerate deployment, policy-makers need to tailor their policies to specific technologies in ways that recognize the stage of development.

In the twentieth century, it took 30 years for energy technologies that were available in principle to grow exponentially and become widely available (Fig. 1). This reaching 'materiality' can be defined as delivering about 1% of the world's energy mix. After that, the growth becomes linear until the technology captures its final market share. This pattern is remarkably consistent across energy technologies and the two growth phases can be seen as the 'laws of energy deployment' (see 'The laws of energy-technology deployment'). Policy-makers concerned about carbon dioxide emissions will want to accelerate the first phase, making energy technologies 'material' within one decade instead of three. But we see two fundamental reasons why the exponential growth in the early, pre-material phase will be hard to beat.

First, scale-up means learning by doing, which takes time in the energy industry. Where energy technology relies on conversion processes — as with next-generation nuclear energy, biofuels or carbon capture and storage (CCS) — historically it has taken three years to build a demonstration plant, one year to start it up and two to five years to overcome setbacks and reach satisfactory operability. So it can take a decade to reach the point where one is confident enough to build the first full-scale commercial plant. It can take another decade to build a dozen.

Where energy technology relies on conversion devices — wind energy, for example — scale-up is equally time-consuming. From 1993 to 2007, worldwide electricity from wind grew at more than 25% a year, in agreement with the first law. Almost two-thirds of this growth comes from more-powerful turbines. In the mid-1980s, 50-kilowatt wind turbines delivered an annual total of 1,000 terajoules (TJ). It would have been impossible to deliver 2007 levels of wind energy (about 600,000 TJ globally) with those turbines. To deploy today's powerful turbines (1–5 megawatts) at industrial scale, and at reasonable cost, required a multi-decadal development effort.

Second, industrial capacity is more important than money. In rough terms, it takes US$100 million–200 million to deliver a project at the bottom end of the energy scale (equivalent to 1,000 TJ per year). It takes a few hundred billion to bring the same technology to materiality[4]. When that technology is new, it takes time to build the human and industrial capacity to do that. You cannot just spend $1 trillion overnight in a $30-billion industry, which is where photovoltaics — solar power — is today.

SUMMARY

- There are physical limits to the rate at which new technologies can be deployed
- Governments need to design policies targeted at specific technologies to accelerate deployment
- More action is required on demand side to increase efficiency and curtail consumption

After reaching materiality, growth curves have historically levelled off (Fig. 1). This is our second law. Unlike consumer goods that may become obsolete in a few years, the capital goods of the energy system have a lifetime of 25–50 years. That means only 2–4% of existing technology needs replacing in a given year. These replacement rates are hard to increase because the economic barrier to replacing old technology is extremely high: industry will only consider early retirement of the existing capital stock if the total cost of the new technology (capital and operating costs) falls below the operating cost of the old.

Law 1

When technologies are new, they go through a few decades of exponential growth, which in the twentieth century was characterized by scale-up at a rate of one order of magnitude a decade (corresponding to 26% annual growth). Exponential growth proceeds until the energy source becomes 'material' — typically around 1% of world energy.

Law 2

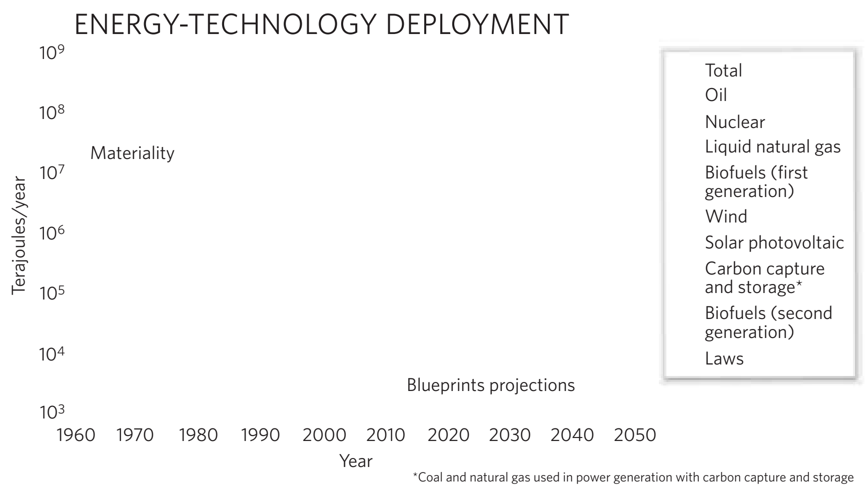
After 'materiality', growth changes to linear as the technology settles at a market share. These deployment curves are remarkably similar across different technologies.

Photovoltaics supply just 0.01% of world energy today. But suppose they supply 10% of the global energy demand by 2050. If solar panels last roughly 20 years, that is a turnover rate close to 5%. As such, the long-term industrial capacity needed to build solar panels and install them will be just 0.5% of world energy demand per year. Together, the replacement rates of old and new stocks explain why the growth curves become linear.

Burn out

The sheer scale and inertia of the energy system may explain why some conclude that the energy challenge requires a response comparable to industrial war efforts[5]. Alas, such arguments completely ignore the second law of deployment. In addition, war-scale efforts typically burn out within a decade, leaving a massive bill for posterity.

The empirical laws we describe here are not laws of nature. They are societal laws best explained using a prudent investor perspective — which applies both to private investors and to government financing. Ever since the rise of coal and oil, every major deployment of new energy technology (nuclear, wind, biofuels and even natural gas) has occurred with some



*Coal and natural gas used in power generation with carbon capture and storage

HISTORIC DATA: OECD/IEA/PREDICTIONS: SHELL INTERNATIONAL

Figure 1 | Global production of primary energy sources. When a technology produces 1,000 terajoules a year (equivalent to 500 barrels of oil a day), the technology is 'available'. It can take 30 years to reach materiality (1% of world energy mix). Projections after 2007 taken from Shell's Blueprints scenario[3].

form of government support. The challenge in the decades ahead is to match, perhaps even outperform, the historic 'laws' by designing energy policies directed at decarbonizing the energy industry.

So what might be possible if policy were aimed at delivering a low-carbon energy supply? Shell tried to answer this question with its 2008 energy scenarios, one of which (Blueprints) has optimistic projections for new energy deployment[3]. In the Blueprints scenario, most new energy types reach materiality by 2030 (photovoltaics by 2020) and their subsequent deployment is on aggregate faster than the historic laws (Fig. 1). CCS is fully available and there is significant carbon pricing and trading. By 2050, total energy demand is one-third lower than business-as-usual projections, mostly through enhanced efficiency and adoption of electric vehicles. Renewable energy supplies one-quarter of the total demand for energy, but none of the individual energy sources exceed 10%.

We believe that the Blueprints scenario is the best we can reasonably hope to achieve for new energy deployment, yet in it, by 2050 two-thirds of the world energy supply still comes from fossil fuels and CO_2 concentrations stabilize at around 550 p.p.m.. At this level, the world is still at considerable risk of dangerous climate change, especially when contributions from other greenhouse gases are taken into account.

How does Blueprints achieve such fast deployment? Partly with a mix of policies that change as a technology moves along the deployment curve. When an energy option is only available in principle, there is little use in subsidizing its deployment through market incentives. Instead, government support for R&D and pilot projects is key. Neither a carbon trading price nor a fuel subsidy will be enough to stimulate commercial investments in demonstration plants for second-generation biofuels or CCS. As technology families they must be picked as winners by the government, even if the market can be left to the job of choosing specific technologies.

Nurturing development

In the case of CCS, governments are hoping to beat the first law by supporting two-dozen pilot projects[6], each at a much larger scale than today's 30-MW project. But CCS will continue to need preferred treatment beyond 2020 to bring it to materiality — through full recognition of CCS in carbon-trading markets and directed government support.

Indeed, as technologies move up the deployment scale, the nature of the support changes from pilot projects to market interventions, such as feed-in tariffs to cover the difference between energy-generation costs and wholesale energy prices. These can be effective, but subsidies need to be technology-specific. Already, feed-in tariffs for photovoltaics and wind differ, allowing photovoltaics to compete with wind. We hope that legislators will treat second-generation biofuels differently from the first generation when the time comes.

Once the threshold of materiality is crossed, the technology costs do become more important. Unit costs will need to fall sufficiently so that any remaining subsidies are small. The real challenge at this stage moves to planning infrastructure and land use.

For example, the dispersed nature of renewable energy means higher land-use requirements, which may act as a significant brake[7]. Another limiting factor is the need for enabling technologies. Utility-scale energy storage, for example, will be needed to smooth the supply from intermittent renewable-energy sources such as wind and solar. Intermittent resources can destabilize an electricity grid if they supply more than 20% of the power unless there is storage available. Similarly, significant CCS deployment will require CO_2 pipelines and storage, eventually at a scale comparable to today's natural-gas infrastructure. Governments must plan for this well in advance, otherwise such factors will limit the later market share of new technologies, and therefore their contribution to a low-carbon world by 2050.

Even with all these policies in place, the CO_2 concentrations achieved in the Blueprints scenario fall short of environmental ambitions. An even tougher goal of stabilizing CO_2 concentrations at 450 p.p.m. — as climate science recommends — would require a largely decarbonized energy sector by 2050. Our best chance of beating the deployment laws requires efforts on multiple fronts, as Blueprints shows, but going beyond those optimistic projections remains an even more significant challenge.

One implication of the deployment laws is that more action is required on the demand side to increase efficiency and curtail consumption. The good news is that demand-side solutions are subject to different laws. In principle, everyone in the developed world could use less energy tomorrow. The bad news is that it has proven exceedingly difficult to restrain our appetite for more energy. No climate actions are easy and none of them is quick. ■

Gert Jan Kramer is principal scientist at Shell Global Solutions International, Grasweg 31, 1031 HW Amsterdam, The Netherlands and **Martin Haigh** is energy adviser at Shell International, Carel van Bylandtlaan 16, 2596 HR The Hague, The Netherlands.
e-mail: gertjan.kramer@gmail.com

1. Friedman, T. L. *Hot, Flat, and Crowded: Why We Need a Green Revolution — and How it Can Renew America* (Farrar, Straus and Giroux, 2008)
2. Gore, A. The Climate for Change *New York Times* (9 November 2009).
3. Shell Energy Scenarios to 2050 (Shell International, 2008); available at http://www.shell.com/scenarios
4. *Energy Technology Perspectives 2008*, Chapter 6 (International Energy Agency, 2008).
5. Jacobson, M. Z. & Delucchi, M. A. *Sci. Am.* **301,** 58–65 (November, 2009).
6. *Science* **325,** 1644–1645 (2009).
7. MacKay, D. J. C. *Sustainable Energy — Without the Hot Air* (UIT Cambridge, 2008).

The authors declare competing financial interests: details accompany the article online at go.nature.com/m4cjLa.

See Opinion, page 570. Discuss this article at go.nature.com/9eLudp and see online at www.nature.com/roadtocopenhagen.

OPINION

Let the global technology race begin

In the second of two pieces on decarbonization, **Isabel Galiana** and **Christopher Green** argue that fostering a technology revolution, not setting emissions targets, is the key to stabilizing the climate.

S. LEEN/NATL GEOGRAPHIC/GETTY

The fixation on near-term targets for reducing greenhouse-gas emissions at the climate meeting in Copenhagen has resulted in insufficient attention towards the technological means of achieving them. Let us not forget that the true objective of climate-change policy is to limit global temperature increases. To achieve that requires a revolution in energy technology and the courage to dispense with target-focused climate policies. The race to solve the climate problem will be won by Aesop's tortoise — not the hare.

Stabilizing the climate is a huge technological challenge[1–3] and the solution of ready-to-deploy, scalable low-carbon technologies is far from being a reality[4,5]. We calculate that if global emissions are to be reduced by at least 80% by 2100 — a suggested goal if the rise in global temperature is to be limited to 2 °C — while maintaining global economic growth at 2.2% per year, two things are required. First, the energy intensity of the economy will need to be reduced to one-third of its 2000 level and, second, consumption of carbon-free energy will need to be almost three times greater than the total energy consumed globally in 2000, 85% of which was supplied by fossil fuels. To achieve this goal by 2100, energy-technology research and development (R&D) will be essential to decarbonize the global economy, through huge scale-ups of existing low-carbon technologies as well as breakthroughs.

SUMMARY

- Reducing carbon emissions requires an energy-technology revolution that has not yet started
- A slowly rising carbon price could provide long-term financing for energy research and development
- A global technology race asks parties to use scientific capabilities rather than to sacrifice future development

To describe the required trade-offs of any climate policy, analysts use the Kaya identity $C = P \times (GDP/P) \times (E/GDP) \times (C/E)$, which relates carbon emissions, C, to its four driving factors: population (P); per capita gross domestic product (GDP/P); energy intensity of the economy (E/GDP); and emissions per unit of energy (C/E). Conventional climate policy considers only the emissions, C, and the political will needed to achieve reductions, but ignores the driving factors. Policy-makers are understandably reluctant to use population or economic growth to reduce greenhouse-gas emissions; hence policy should focus on the technological drivers. A useful way of looking at these is by combining E/GDP and C/E to yield the economy's carbon intensity (C/GDP).

In recent decades, although global GDP has grown at about 3% per year and global carbon intensity has declined by about 1.4% per year, emissions have grown well in excess of 1% per year. In view of this, the proposal by the Group of 8 rich nations (G8) to cut global emissions in half by 2050, consistent with limiting global long-term temperature increase to 2 °C — and to do this without slowing economic development — would require a tripling of the average annual rate of decline in carbon intensity for the next 40 years. This accelerated decline in carbon intensity requires a revolution in energy technology that has not yet started.

Can a technology-led approach avoid dangerous climate change? We proposed such a policy[6] as part of the 2009 Copenhagen Consensus on Climate, in which a panel of leading economists ranked 15 policy responses to global warming. Our analyses show that cumulative emissions consistent with minimizing the rise in global temperature (climate stabilization) can be achieved by investing US$100 billion a year for the rest of the century in global energy R&D, testing, demonstration and infrastructure.

For two of the three technology paths we investigated, cumulative carbon emissions would be kept at levels that could limit long-term warming to 2 °C (ref. 7), despite the greatest emissions reductions occurring after 2050. Moreover, all of the paths we considered passed cost–benefit tests, usually by very large margins. A technology-led approach can stabilize the climate with higher probability and much lower cost than the emissions-target approach.

How would the technology-led approach work? First, governments would replace emissions targets with credible long-term global commitments to invest in energy R&D. To finance this, we propose a low carbon price of $5 per tonne of emitted carbon dioxide, which would raise almost $150 billion per year globally and $30 billion in the United States alone.

Second, the low carbon fee should be allowed to rise gradually over time, doubling, say, every 10 years. This would send a 'forward price signal' to deploy new or improved low-carbon technologies as they become scalable and cost-effective. Third, R&D funds should be isolated as far as possible from political interference by placing them in dedicated trust funds that are administered by independent committees drawn from the public and private sectors. Allocation of funds would be left to experts, akin to The Bill & Melinda Gates Foundation. Energy-technology competitions could be open to individual enterprises, nations or international coalitions. Countries that decide not to participate in R&D could use the funds raised to purchase successfully developed technologies from those that do participate.

Together, these three elements are catalysts for the development and deployment of an increasing flow of successful low-carbon energy innovations over the coming decades. These would span the technology spectrum: basic R&D in breakthrough technologies; 'enabling' R&D that

Increasing the supply of low-carbon energy requires policy actions now.

allows scale-up of existing technologies (such as utility-scale storage for intermittent solar and wind energy); testing and demonstration projects; and energy-related infrastructure, such as 'smart grids' that help to manage intermittent energy sources. As these innovations diffuse, they will accelerate the decline in carbon intensity and reduce emissions to negligible levels by 2100 while maintaining economic growth.

Emission-target pitfalls

Our technology-led proposal inverts the usual relationship between carbon pricing and technology, whereby carbon pricing is naively expected to induce fundamental technological innovation. Experience shows that, even if there is successful innovation, the market is unlikely to undertake risky basic R&D when the benefits are difficult to appropriate. For example, public funds supported the basic scientific knowledge underlying radar, nuclear power, the Internet and the understanding of DNA — none of which are easy to patent. Also, governments cannot credibly convince current investors that future governments will set carbon prices high enough to compensate for upfront and risky R&D investments[8]. Moreover, potential pay-offs are decades rather than years in the future and are thus discounted away by myopic investors.

What if aggressive emissions-reduction targets were adopted and enforced without paying attention to the capabilities of existing technologies to deliver good low-carbon substitutes? The Kaya identity tells us that attempting to reduce emissions sharply would substantially reduce economic growth. For example, meeting the G8 goal of slashing global carbon emissions by 50% by 2050 — which would require developed countries to reduce emissions by up to 80% — could result in economic costs as high as 10% or more of their GDP.

Another problem with conventional policy is that high carbon prices fall heavily on energy-intensive industries and activities, which are often important to domestic production and employment. In populous, rapidly developing countries, the construction of new high-rise buildings and roads requires cement, steel, glass, copper and aluminium, all of which are highly energy-intensive. Understandably, most consumer-voters will be unwilling to pay the high carbon prices required by near-term emissions-reduction targets[9] in the absence of reliable and inexpensive clean-energy sources. To be feasible, a carbon price must start low and rise slowly.

Proponents of a high carbon price (whether through a carbon market or tax) too easily assume that it will induce rapid technological change. But implementing high carbon prices now on the assumption of future technological pay-offs, without direct technological investment, is risky and expensive. Thus, we must address the technology challenge now. Still, supporters of conventional policy may argue that a technology-led approach lacks accountability for emissions reductions. Although we cannot predict exactly when emissions will be reduced because of the inherent uncertainty in deploying new technologies, we can predict that the carbon intensity of the economy will decline dramatically once such technologies become readily available.

Does a technology-led policy imply particular R&D funding choices? No. With $100 billion a year to spend, there will be various initiatives, including carbon-capture and storage technologies, advanced nuclear and deep-geothermal energy, next-generation biofuels, ocean-wave energy and storage for solar and wind energy. There should be no need to pick 'winners' or to get locked into inferior technologies.

What about the choice between R&D in 'enabling' and 'breakthrough' technologies? Both will be needed, and in many cases they overlap. Given the longer lead times needed for breakthrough technologies, a balance between investments in technologies with near- to medium-term pay-offs and those with long-term pay-offs seems sensible. No matter how one approaches climate change, stabilization is a long-term process with numerous paths[10].

There are other advantages to a technology-led climate policy. It is attractive to developing countries such as China, India, Brazil and Indonesia because, instead of asking for emissions reductions, a technology-led policy invites them to participate in an energy-technology race. Many of these countries are already pressing ahead with R&D of low-carbon technologies.

A global technology race could attract new generations to the challenge of stabilizing climate by asking them to share their scientific and creative capabilities rather than to sacrifice future development. It would yield benefits in terms of spillovers to non-energy technologies and uses, just as the world is still benefiting from innovations developed in the race to put a man on the Moon. Ultimately, a technology-led policy is optimistic because it rewards success rather than punishing failure. ■

Isabel Galiana and **Christopher Green** are in the Department of Economics at McGill University, Montreal, Quebec H3A 2T7, Canada.
e-mails: isabel.galiana@mail.mcgill.ca; chris.green@mcgill.ca

1. Hoffert. M. I. *et al. Nature* **395,** 881–884 (1998).
2. Caldeira, K., Jain, A. K. & Hoffert, M. I. *Science* **299,** 2052–2054 (2003).
3. Pielke, R. Jr, Wigley, T. & Green, C. *Nature* **452,** 531–532 (2008).
4. Hoffert, M. I. *et al. Science* **298,** 981–987 (2002).
5. Barrett, S. *J. Econ. Perspect.* **23,** 53–75 (2009).
6. Galiana, I. & Green, C. *An Analysis of a Technology-led Climate Policy as a Response to Climate Change* (Copenhagen Consensus Center, 2009).
7. Meinshausen, M. *et al. Nature* **458,** 1158–1162 (2009).
8. Montgomery, W. D. & Smith, A. E. in *Human-induced Climate Change* (Cambridge Univ. Press, 2007).
9. Nemet, G. F. *Res. Policy* **38,** 700–709 (2009).
10. Wigley, T. M. L. *et al. Nature* **379,** 240–243 (1996).

See Opinion, page 568. Discuss this article at go.nature.com/9eLudp and see online at www.nature.com/roadtocopenhagen.

BOOKS & ARTS

Freezes, floes and the future

The story of Earth's glaciers and ice caps is key to understanding climate science, but this kaleidoscopic account lacks a strong narrative, argues **Johannes Oerlemans**.

A World Without Ice
by Henry Pollack
Avery: 2009. 304 pp. $26

A World Without Ice opens with a strong foreword from Al Gore: the science has been done — now we must act. Eminent scholars such as geophysicist Henry Pollack, says Gore, are needed to communicate the severity of the climate-change problem to the public. But for several reasons Pollack's patchwork assessment of the science of ice and climate, doesn't quite meet that challenge.

His book gets off to a bad start by adding drama. Writing in his preface that "Throughout most of Earth's history, ice has been an indomitable force of nature", Pollack sidesteps the consensus view that for the majority of Earth's past there was little or no ice, even though signatures of glacial action are visible in many places today. He also overstates the contribution of the polar ice caps to the planet's albedo — the amount of sunlight reflected from Earth's surface — saying that the ice caps account for much of it. Climate models show that their contribution is less than 1% now, and was only 3% during the Last Glacial Maximum 18,000 years ago, when the land-ice area was three times that of today. Similarly, he cites mountain glaciers as the direct source of water for almost a quarter of the world's population, when in reality the bulk comes from rain and seasonal snow.

The book's mix of history and scientific fact makes for an informative but unfocused account. Pollack gives an entertaining overview of polar exploration, recalling famous expeditions and how polar science became increasingly organized following the stimulus of the 1957–58 International Geophysical Year. He describes the ratification of the Antarctic Treaty by 12 nations in 1961, which declared Antarctica to be a peaceful, non-militarized continent open to scientific investigation and cooperation, and discusses the problems of polar tourism. Pollack explains how ice sheets work, how past climate can be deduced from deep-sea sediments and how changes in Earth's orbit have affected climate by redistributing incoming solar radiation. After describing the melting of the big ice sheets after the Last Glacial Maximum and the evolution of climate during the Holocene epoch, which started around 12,000 years ago, he jumps to the workings of the Intergovernmental Panel on Climate Change. This short but balanced section deserved a more distinct place in the book.


Melting Antarctic ice may be a good barometer of climate change but it is no crystal ball.

J. BRACK/NATL SCIENCE FOUNDATION


An ice core reveals historical and ongoing processes.

C. NAUGHTON/NATL SCIENCE FOUNDATION

Pollack goes on to compare natural versus anthropogenic mechanisms of climate change, looking at solar activity, volcanism and greenhouse gases such as carbon dioxide. He describes how the human imprints of deforestation, industrial revolution, mining, erosion, irrigation and energy consumption have accreted at an increasing pace, bringing us into the 'Anthropocene' era of man-made environmental change.

In his investigation of the regional effects of global warming on ice, snow and permafrost, Pollack adopts a fearful tone, suggesting that any change in the environment should be interpreted as a local disaster. He lists the many locations where glaciers are retreating, sea-ice coverage is shrinking, permafrost thawing and ski areas declining. And he cautions that "in only a few decades the Arctic Ocean may be ice-free in the summer, for the first time in 55 million years". Yet he forgets that, during the Holocene climatic optimum about 9,000 to 6,000 years ago when summer temperatures in the subarctic regions were 2–5 °C higher than today, the Arctic Ocean in summer was probably ice-free on a regular basis.

Pollack's main message lies in the chapter 'Choices Amid Change'. He places global warming in the context of other environmental problems related to the luxurious lifestyles sought by more and more people. He makes a strong case for relatively simple measures that can reduce energy consumption: more efficient cars, smaller distances between home and work, investing in public transport and insulating buildings. I agree. Big solutions can take forever to become reality, whereas many small steps can make a difference.

It is not until the final chapter that he begins a welcome discussion on the uncertainties, merits and problems of large computer models of the climate system, stating that we must live with change and surprises. Again, his discussion of the ice sheets and sea-level rise is too dramatic: for example, it has not been established that the Greenland ice sheet will melt away in a few centuries once we pass the 'tipping point'. I also missed a more thorough discussion of what would happen to sea levels and all of Earth's ice, known collectively as the cryosphere, if there were no anthropogenic influences on climate.

The slow, ongoing processes in ice sheets may be poorly understood, but they are likely to have implications for the future. For example, it is clear that the West Antarctic Ice Sheet is currently losing mass, but there is abundant evidence that the shrinkage has been happening for the past 15,000 years, mainly in response to rising sea levels initiated by deglaciation in the Northern Hemisphere. Thus, limiting greenhouse warming may not lead to the desired stabilization of the ice sheet. We must then ask whether we want the cryosphere to remain as it is today, or for it to follow its own course with a minimum of anthropogenic influence. To answer this question we need to know the consequences. Fortunately, large amounts of ice-sheet data are now becoming available, particularly from the many different sensors on various space satellites. With clever modelling, in ten years' time we should be able to quantify the effect of slow ongoing processes on the future of the ice sheets.

A World Without Ice is kaleidoscopic: a collection of anecdotes, scientific lectures, travel reports, political statements and repeated arguments. In lacking a strong narrative it doesn't match my personal favourite in climate science, *Ice Ages: Solving the Mystery* (Enslow, 1979) by John Imbrie and Katherine Palmer Imbrie. But Pollack's entertaining book is worth reading, even if it does not offer the full story. ■

Johannes Oerlemans is professor of meteorology at the Institute for Marine and Atmospheric Research, Utrecht University, Princetonplein 5, 3584 CC Utrecht, the Netherlands.
e-mail: j.oerlemans@uu.nl

Gail Wight, artist of science

Restless Dust
by Gail Wight
Imprint of the San Francisco Center for the Book: 2009. 36 pp. $280

The artist Gail Wight has examined X-rays with neuroscientists, pored over skeletal remains alongside archaeologists and thawed insects with animal behaviourists.

Wight, an associate professor in the Department of Art and Art History at Stanford University in California, likens her role in these short-term apprenticeships to that of a "lurker". The term is characteristic of her approach, in that it simultaneously diminishes her stature and emphasizes the tension inherent in the relationship. Some of these professional associations are internship-like shadowing sessions; others develop into something akin to collaboration. Each has led Wight to produce unique works of art that, over 20 years or so, have spanned media from photography and sculpture to film and performance art.

Among Wight's videos are colourful grids of the slime mould *Physarum polycephalum*, its growth visible using accelerated footage. Her *Ground Plane* large-format prints, depicting fossilized materials collected by Elizabeth Hadly's lab at Stanford University, include what might be termed 'mandible mandalas'. The massive, snowflake-like patterns are hand-constructed from photos of tiny fragments of animal bones, some of which are more than 10,000 years old.

Wight's sculptures include large-scale Plexiglas musical instruments that are 'played' by the mice captured inside them. If the captive mice bring to mind those that are probed and dissected in laboratories, the parallel isn't misguided. "I'm really interested in a critique of science, and one that scientists would appreciate," says Wight. "Scientists are the best [placed] to understand what's wrong with their field. I've been pretty delightfully surprised to find that they really love the critique aspect."

Another piece in which she aligns herself with the specimens of science, rather than with scientists, is one of her early performance works from 1992, *School of Evolution*. Its form took that of a day-long seminar, during which her audience was a fountain full of fish; she encouraged them to get on with evolution and thus escape their aquatic confines.

Humour is never far from Wight's artistic toolbox; nor, of late, has been the topic of evolution, through another collaboration — not with a living scientist but the ghost of one. For *Restless Dust*, an artist's book she completed this autumn during a residency in California at the San Francisco Center for the Book, Wight imagined taking Charles Darwin on a tour of the San Francisco Bay area to record his impressions. A visit to a genetics lab makes him think about inbreeding — something that Darwin, who married his cousin, feared may have precipitated the premature deaths of three of their children.

Wight's book marks the occasion of the 150th anniversary of the publication of Darwin's *On the Origin of Species*, and the result is a leather-bound, letterpress-printed journal not unlike the one in which Darwin made the notes that eventually became *The Voyage of the Beagle*. These journals, which Wight produced in an edition of 50, are housed in small pine boxes — "little coffins", she jokes — that also contain life-sized, rice-paper models of a pair of mockingbirds, modelled on Darwin's own research material. The body cavity of each bird pulses with the dim light of a small, battery-powered bulb. The ghost may be Darwin's, but his specimens are memorialized too.

One of Wight's many scientist-collaborators is the animal behaviourist Lucia Jacobs at the University of California, Berkeley, who praises


Gail Wight explores the habits and history of scientific practice in this installation of pinned butterflies.

WWW.STANFORD.EDU/~GAILW

the depth of Wight's engagement: "There's a lot of 'pretty' art about science, and there's a lot of art that involves technology," says Jacobs, "but that's not the same thing. Science is a way of pursuing truth, and Gail really loves that about science — but she's also kind of calling us on it, and in spectacular ways."

Jacobs likens Wight's research to playwright Tom Stoppard's use of maths in *Arcadia*, and mentions several artists-turned-scientists, such as Santiago Ramón y Cajal, whose efforts in founding the field of neuroscience succeeded his early career as a painter. "You could say she is a scientist about science," remarks Jacobs. "Her subject is science, and she's experimenting with it, using art, to understand the truth of it." Wight, however, would never apply the term scientist to herself. "Science is not my forte, it's not my world," she says. "But I live in the world largely constructed by science, so I want to understand it better."

She is an 'artist of science' in the same way that there are historians and sociologists of science. Wight explores the habits and history of science by using its materials as her materials, and producing art that serves both as work open to interpretation and as a pointed appraisal of scientific analysis. Consider her piece *J'ai des papillons noirs tous les jours* (2006), a series of colourful, handmade butterflies, each affixed to the back of its own narrow box by 100 pins (pictured) — the overkill begs the question: how many pins is too many? As with the mockingbirds in *Restless Dust*, the butterflies' abdomens glow.

For all Wight's concerns about scientific practice, she has a firm sense of where the art ends and the science begins. Ultimately, she reserves her strongest critique not for scientists, but for her fellow artists, saying: "I have a lot of problems with the contemporary idea that an artist can enter into the world of science and practise science on living creatures without any of the background that should come with that." ■

Marc Weidenbaum is a writer based in San Francisco, California.
e-mail: marc@disquiet.com

See go.nature.com/aedPm2 for more on Gail Wight's *Restless Dust*.

The Internet of the ancient world

Traveling the Silk Road: Ancient Pathway to the Modern World
American Museum of Natural History, New York
Until 15 August 2010

The time is 8.00 p.m. in Baghdad on 14 November — 1000 AD. I discovered this with the aid of a brass astrolabe, a map of the sky with moveable parts (pictured). I peered through sight-holes to measure the altitude of the red supergiant star Betelgeuse (a twinkling bulb in a model night sky), spun the astrolabe around, rotated a dial, aligned the rule to a mark on the inner rim and read the time on the instrument's outer rim. A simple task for a Muslim astronomer at the turn of the first millennium — or if you're reading the instrument's instructions while visiting the American Museum of Natural History's enchanting exhibition, *Traveling the Silk Road*.

Baghdad was an important stop on a 7,400-kilometre journey along the Silk Road, a network of trade routes that, at its peak 1,000 years ago, stretched from China to the Mediterranean. The road guided merchants on foot or on camelback through scorching deserts and snow-capped mountains, where they confronted bandits, sandstorms, frostbite and hunger, among other dangers. The trek through the exhibition is much safer, but there is still plenty to do. Visitors can watch live silkworms chomping on mulberry leaves, play with interactive maps and displays, listen to an orchestra of ancient Chinese instruments and listen to familiar folk-tales carried to a global audience along the Silk Road — which was "the Internet of the ancient world", says the museum's president Ellen Futter.

AMNH/D. FINNIN

To guide their journey, visitors are equipped with a passport that they can stamp at four cities en route through the exhibition: Xi'an, the ancient capital of China, from which camel caravans set out loaded with silk for trade with the West; Turpan, an oasis city on the edge of the Taklimakan Desert in northwest China; Samarkand, a merchant city now in present-day Uzbekistan; and Baghdad, once a meeting place for scholars and known as the City of Peace. On the way, visitors grasp the great cultural and technological developments that flowed along the route: music, religion, language, numerals, medicines and innovations such as paper-making.

A two-metre-long, table-top interactive map charts these advances. Paper was invented by Chinese craftsmen in 50 BC, who mashed plant fibres such as hemp and flax with water, shaping the pulp into paper using a wooden mould. One such mould, with a screen made of woven reeds, is on display. The Chinese kept their paper-making technique a secret for centuries, but it slowly wound its way westwards. The Diamond Sutra, an ancient Buddhist text printed in 868 AD, is the oldest known dated example of a paper book and was found in a cave in Dunhuang, an oasis town that formed an interchange between China and western Asia. The first paper Koran was copied in Persia in 971 AD, and in 1150 Arabs set up the first paper mill in Al-Andalus, the Islamic territories in southern Spain.

Innovations travelled in the other direction, too. Glass-blowing was developed in the Middle East in around 100 BC, and then progressed eastwards: one beautiful, long-necked cloudy-white bottle from western Asia dating to 800–1000 AD is imprinted with pairs of graceful, curved-horned ibex. Irrigation techniques also spread along the Silk Road: artificial underground rivers called karezes, which rely on gravity to carry mountain water to distant desert fields, were first used in Persia more than 2,500 years ago and still move nearly 300 million cubic metres of water in the Turpan Basin each year.

A mock market displays luxury goods offered by traders in Turpan: items such as leopard, tiger and ermine fur; red coral and green jade; a profusion of fruit including figs, peaches, pistachios and melons — the latter transported in lead containers packed with snow — and medicines such as rhino horn, rhubarb, ginseng and bezoars, the brown pellets of undigested food extracted from the stomachs of goats and cows, purported to cure a range of ills and rid you of "perverse goblins", according to one Chinese doctor.

A final video quiz reminds us that the globalization of trade and culture continues unabated. China, which grew wealthy on silk sales, is now the world's biggest exporter of manufactured goods. To get the best price for their catch, rural fishermen in India call ahead on mobile phones — more than four billion of which are now used worldwide. And in a 2001 speech to "celebrate Britishness", then UK foreign secretary Robin Cook called a sauce-doused ensemble "a true British national dish". It was chicken tikka masala. ■

Josie Glausiusz is a journalist based in New York.
e-mail: jg@planetjosie.net

PLANT BIOLOGY

Signal advance for abscisic acid

Laura B. Sheard and Ning Zheng

The hunt for the receptor for abscisic acid, initially marked by false starts and lingering doubts, has met with success. Converging studies now reveal the details of how this plant hormone transmits its message.

To survive and flourish in rapidly changing environmental conditions, plants use a complex family of hormones to regulate growth and reproduction. One plant hormone, abscisic acid (ABA), coordinates responses to stressors such as drought, extreme temperature and high salinity, as well as regulating non-stress responses including seed maturation and bud dormancy. Because of its essential function in plant physiology, targeting the ABA signalling pathway holds enormous promise for future application in agriculture. A burst of papers, four in this issue[1–4] and two published elsewhere[5,6], provide structural and mechanistic insights into ABA signalling that take us closer towards realizing this promise.

Plant hormone research has been the subject of recent attention owing to the identification of receptors for two other hormones, gibberellin and auxin, that like the current model for ABA have soluble (non-membrane-bound) receptors. But identification of the ABA receptor has been unusually challenging. Since 2006, several proteins have been suggested as possible ABA receptors, but their exact roles in ABA signalling remained controversial[7]. In May 2009, a new family of proteins was reported as candidate ABA sensors[8,9]. Members of this family, known as PYR/PYL/RCAR proteins, were found to bind ABA and inhibit the activity of specific protein phosphatase enzymes, the type 2C plant PP2Cs, which were previously implicated in the ABA response. Six independent groups[1–6] now simultaneously define the structural and functional mechanisms by which ABA is sensed by this newly identified protein receptor.

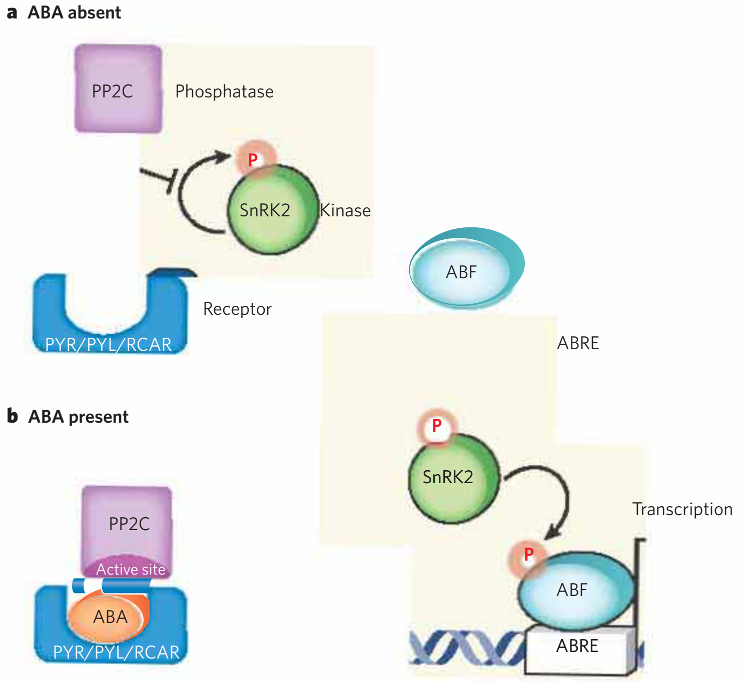


Figure 1 | Minimal abscisic acid (ABA) signalling pathway. a, In the absence of the plant hormone ABA, the phosphatase PP2C is free to inhibit autophosphorylation of a family of SnRK kinases. **b,** ABA enables the PYR/PYL/RCAR family of proteins to bind to and sequester PP2C (see Figure 2, overleaf, for mechanistic details). This relieves inhibition on the kinase, which becomes auto-activated and can subsequently phosphorylate and activate downstream transcription factors (ABF) to initiate transcription at ABA-responsive promoter elements (ABREs).

Fujii *et al.*[3] (page 660 of this issue) report elegant reconstitution assays that pinpoint the minimal pathway sufficient to recapitulate ABA signalling in plant protoplasts (cells with their cell wall removed) and *in vitro*, tying together the receptor, phosphatase and downstream kinase signalling components. Phosphatases and kinases exert opposite regulatory effects by respectively removing and adding phosphate groups to substrate proteins. As described in Figure 1, in the absence of ABA, the phosphatase PP2C acts as a constitutive negative regulator of a family of kinases (SnRK2) whose autophosphorylation is required for kinase activity towards downstream targets. When ABA binds, it enables the PYR/PYL/RCAR receptor to subsequently bind to and repress PP2C. Sequestration of PP2C permits auto-activation of the kinase, which phosphorylates downstream transcription factors and facilitates transcription of ABA-responsive genes. This pathway is attractive in its simplicity and offers a seamless complement to the known body of ABA literature.

The crucial step in this pathway is perception of ABA by the PYR/PYL/RCAR proteins and the inhibition of PP2C by the ligand (ABA)-bound receptor. Five crystallographic studies[1,2,4–6], including those on pages 602, 609 and 665, have converged to paint a complete picture of these events. Together, they reveal the atomic structures of several PYR/PYL/RCAR proteins in different functional states. Studies by Melcher *et al.*[1], in particular, have captured the structures of PYL2 in all critically relevant forms (ligand-free, ligand-bound and ligand/phosphatase-bound), and allow detailed analyses of the conformational changes that PYL2 undergoes on binding first to the hormone, and subsequently to the phosphatase.

The first highlight of these studies is the ligand-binding mechanism, which is regulated by the opening and closing of a gating loop on the ABA-binding pocket. In the absence of the hormone, the PYR/PYL/RCAR proteins present an open and accessible cavity. Two flexible surface loops, together with several

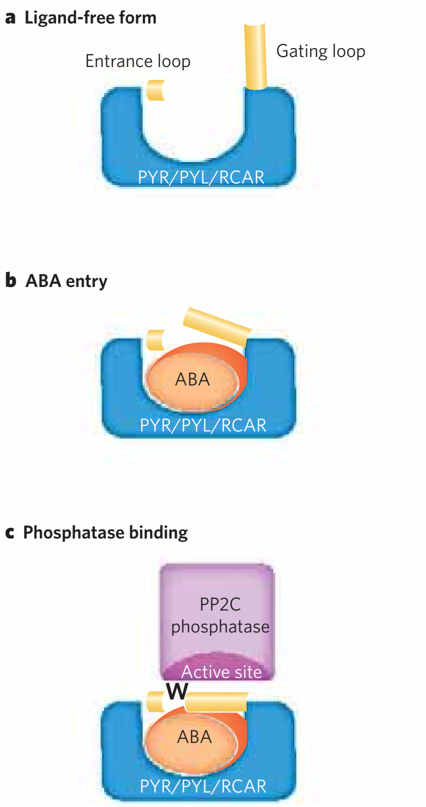


Figure 2 | Structural mechanism of ABA action. **a,** In the ligand-free form, the ABA receptor PYR/PYL/RCAR presents an open and accessible cavity with two flexible surface loops that guard the cavity's entrance. **b,** ABA initiates an allosteric open-to-close transition of the gating loop, allowing it to approach the second entrance loop and sequester ABA in the pocket. This exposes a hydrophobic binding site on the gating loop. **c,** The phosphatase PP2C binds to the hydrophobic site on the gating loop, inserting a conserved tryptophan (W) next to the gating loop and locking it closed. In turn, the gating loop interacts closely with the active site of the phosphatase, blocking its ability to bind to its substrate.

nearby structural elements, guard the entrance of the cavity. When this water-filled pocket is occupied by ABA, one of the loops closes like a gate, approaching the other loop and sequestering ABA within the pocket (Fig. 2). Because most of the amino-acid residues in contact with ABA, as well as the sequences of the two entrance loops, are evolutionarily conserved among all PYR/PYL/RCAR proteins, the ABA-binding and open-to-close gating mechanisms are likely to be common for all members of the receptor family.

Three of the five groups[1,2,6] extend their crystallographic studies to describe the architecture of the complex formed between an ABA-bound receptor and PP2C. These structures reveal that the ABA-bound receptors dock onto PP2Cs through a large complementary interface that involves the active site of the phosphatase and the two entrance loops of the receptor. A conserved tryptophan residue of the phosphatase inserts its side chain next to the gating loop, locking it closed. In turn, the gating loop interacts closely with the substrate-binding and active site of the phosphatase, blocking its ability to bind and dephosphorylate its substrate. Together, these structural features comprehensively explain how the PYR/PYL/RCAR proteins inhibit PP2C activity in an ABA-dependent manner, and how PP2Cs act as a potent co-receptor to enhance the affinity of the hormone for its receptor.

These reports present us with a consistent view of ABA signalling that also raises questions for future work. In three of the structural studies[4–6], receptor dimerization is observed in the absence of PP2C; and although it is not discussed in the text, dimerization is also present in structural models of the remaining two studies[1,2]. But the exact purpose of receptor dimerization remains unclear. In the structure of homodimeric PYR1, only one molecule of ABA can bind per dimer, whereas in the related structure of PYL homodimers, both subunits are occupied. In all of the structural models, the dimer interface involves the flexible gating loop of the receptor, suggesting that dimerization may be functionally relevant. However, receptor dimerization is clearly not required for the final action of the hormone, because only monomeric ABA-bound receptor is found in complex with PP2C.

The movement of the gating loop in PYR/PYL/RCARs to create a hormone-dependent PP2C-binding site is reminiscent of the 'closing lid' mechanism used by the receptor GID1 to sense gibberellin. In this case, hormone binding induces a movement of parts of GID1 that cover the hormone-binding pocket, creating a site for GID1 to bind to substrate proteins and initiate another form of chemical modification, ubiquitylation[10,11]. Both ABA and gibberellin allosterically remodel their respective receptors, in contrast to the 'molecular glue' mechanism used by auxin[12]. Although the precise mechanistic details may differ, there is a common feature in plant hormone action at soluble receptors: the hormone signal enhances protein–protein interactions to modulate critical modifications — either phosphorylation or ubiquitylation — that will alter the activity of the target protein. In addition, in all cases, the hormone binds to a site that is directly at or near the protein–protein interface, engaging the associating protein as a co-receptor. ■

Laura B. Sheard and Ning Zheng are at the Howard Hughes Medical Institute and Department of Pharmacology, University of Washington, Seattle, Washington 98195, USA.
e-mail: lsheard@uw.edu

1. Melcher, K. *et al. Nature* **462,** 602–608 (2009).
2. Miyazono, K. *et al. Nature* **462,** 609–614 (2009).
3. Fujii, H. *et al. Nature* **462,** 660–664 (2009).
4. Santiago, J. *et al. Nature* **462,** 665–668 (2009).
5. Nishimura, N. *et al. Science* doi:10.1126/science.1181829 (2009).
6. Yin, P. *et al. Nature Struct. Mol. Biol.* doi:10.1038/nsmb.1730 (2009).
7. McCourt, P. & Creelman, R. *Curr. Opin. Plant Biol.* **11,** 474–478 (2008).
8. Ma, Y. *et al. Science* **324,** 1064–1068 (2009).
9. Park, S. Y. *et al. Science* **324,** 1068–1071 (2009).
10. Murase, K. *et al. Nature* **456,** 459–463 (2008).
11. Shimada, A. *et al. Nature* **456,** 520–523 (2008).
12. Tan, X. *et al. Nature* **446,** 640–645 (2007).

BIOLOGICAL CHEMISTRY

Dehydrated but unharmed

Justin L. P. Benesch and Carol V. Robinson

The weakest interactions of protein complexes are thought to be lost when such assemblies are removed from their natural, watery environments. Not so, reveals a study in the vacuum chamber of a mass spectrometer.

A long-accepted tenet of biochemistry is that water is integral to protein folding and function. Nevertheless, proteins can remain active in the absence of bulk water — for example, enzymes have been shown to carry out their functions in organic solvents[1], implying that the proteins' structures remain at least partially folded in these conditions. However, it is also known that, in a vacuum, proteins can populate many more conformations than are accessible to them in solution[2]. Bearing in mind this apparent conflict of experimental data, it is interesting to read Liu and colleagues' report[3] in the *Journal of the American Chemical Society*, in which the authors reveal that even water-mediated interactions can be preserved in a vacuum. By introducing a protein and its associated ligands into the 'dry' vacuum inside a mass spectrometer, the researchers demonstrate not only the survival of hydrophobic protein–ligand interactions, but also that the protein retains the structure of its binding cavity without assistance from water. This suggests that many analytical techniques that operate under a vacuum could have applications in structural biology.

After the introduction of electrospray ionization — a means of transferring proteins from an aqueous solution to the gas phase for mass spectrometry — researchers were quick to capitalize on the opportunities presented to study proteins and the non-covalent complexes into which they assemble[4]. Since those early days, technological and methodological advances have continued apace, so that fragile assemblies of proteins, nucleic acids, ligands and co-factors can now be studied intact using mass spectrometry[5]. As progressively more complex systems are studied, however, a crucial question remains: to what extent do these macromolecular assemblies maintain their solution-phase characteristics in the gas phase?

One approach that can address this question — and indeed one of the approaches used by Liu and colleagues[3] — is ion-mobility spectrometry. This technique monitors the progression of ions through an inert buffer gas, and provides information about the size of molecules that is complementary to that obtained using traditional mass spectrometry. Such experiments have shown that the global structure of macromolecules, including the topology of fragile multi-protein assemblies, can be preserved in a vacuum[6]. But the extent and, importantly, the timescale over which local elements of structure are maintained remain less well understood[2], particularly for hydrophobic interactions. It is, therefore, vital that less complex systems are examined in detail, such that individual interactions can be addressed at almost the atomic level.

Liu *et al.*[3] have done just this, by examining the binary complexes formed between the protein β-lactoglobulin and different fatty acids in the gas phase. These small molecules, composed of a water-repellent hydrocarbon chain and an acidic head-group, bind in a hydrophobic pocket within the protein, and the resultant complexes can be maintained intact in a vacuum. The researchers slowly heated these complexes in a mass spectrometer, causing the fatty acids to dissociate from the protein, and monitored the process with mass spectrometry. The great advantage of the authors' approach is that it allows temperature-dependent kinetic parameters for the dissociations to be obtained.

Backed up by ion-mobility measurements and molecular-dynamics simulations, Liu *et al.* found evidence for two distinct types of structure. In the first, the protein adopts a 'closed' conformation, wherein the head-group of the fatty acid forms hydrogen bonds with residues at the entrance to the pocket. It is therefore not surprising that this protein–ligand complex survives in the gas phase.

In the second, more 'open' structure, however, such stabilizing hydrogen-bonding interactions are absent, yet the complex nevertheless remains intact within the mass spectrometer. The authors observed that the energy required to dissociate fatty acids from this structure (the activation energy for dissociation) scales linearly with the length of the acid's hydrocarbon chain, and hence with the amount of nonpolar interaction between the acid and the protein. Moreover, the authors found the activation energies to be similar to the energies expected to be required to dissolve the fatty-acid chains in organic, nonpolar solvents. So, although in both kinds of structure the fatty acid is bound within the hydrophobic cavity, in the open case the lack of hydrogen bonding means that the interactions are solely nonpolar, occurring between the fatty acid's hydrocarbon chain and the hydrophobic amino-acid residues lining the protein's binding pocket.

The fact that these hydrophobic interactions can be maintained in the gas phase (albeit in a simple bimolecular interaction) on the timescale of mass-spectrometry investigations is a crucial observation. The extent to which these findings apply to other systems remains to be seen. Nevertheless, Liu and colleagues' results[3] provide a clear rationale for how the structural integrity of various noncovalently bound complexes is maintained in the gas phase — particularly for those that rely heavily on hydrophobic interactions, such as membrane–protein complexes[7], or assemblies of molecular 'chaperones' with unfolded client proteins[8]. Moreover, the authors' observations demonstrate that even those interactions that seem, intuitively, to be the most susceptible to deformation upon dehydration can be maintained in the gas phase. This provides strong validation of the idea that mass spectrometry can be used not only as a tool for identifying and quantifying the proteins in cells, but also for characterizing the complexes they form.

The emergence of mass spectrometry as a tool for structural biology comes as the field itself is undergoing dramatic changes. Although initiatives in structural genomics have been determining protein structures at a remarkable rate, the structures of many vital targets remain frustratingly elusive. Furthermore, the essential role of conformational fluctuations in protein function has become increasingly apparent, although the characterization of these protein dynamics is difficult. As such, there is a realization that new tools need to be developed to complement traditional methods in structural biology. Hybrid techniques are emerging that combine information collated using several different methods[9]; mass spectrometry, with its ubiquity, speed and ability to work with small samples, is well placed to contribute. Indeed, with Liu and colleagues' report[3], the advent of reliable structural information — everything from simple binding stoichiometries to diffraction patterns of individual macromolecules[10] — obtained from gas-phase measurements alone moves one step closer. ■

Justin L. P. Benesch and Carol V. Robinson are in the Department of Chemistry, University of Oxford, Oxford OX1 3QY, UK.
e-mails: carol.robinson@chem.ox.ac.uk; justin.benesch@chem.ox.ac.uk

1. Klibanov, A. M. *Nature* **409,** 241–246 (2001).
2. Breuker, K. & McLafferty, F. W. *Proc. Natl Acad. Sci. USA* **105,** 18145–18152 (2008).
3. Liu, L., Bagal, D., Kitova, E. N., Schnier, P. D. & Klassen, J. S. *J. Am. Chem. Soc.* **131,** 15980–15981 (2009).
4. Loo, J. A. *Mass Spectrom. Rev.* **16,** 1–23 (1997).
5. Benesch, J. L. P. *et al. Chem. Rev.* **107,** 3544–3567 (2007).
6. Wyttenbach, T. & Bowers, M. T. *Annu. Rev. Phys. Chem.* **58,** 511–533 (2007).
7. Barrera, N. P. *et al. Science* **321,** 243–246 (2008).
8. van Duijn, E. *et al. J. Am. Chem. Soc.* **131,** 1452–1459 (2009).
9. Robinson, C. V., Sali, A. & Baumeister, W. *Nature* **450,** 973–982 (2007).
10. Chapman, H. N. *et al. Nature Phys.* **2,** 839–843 (2006).

50 & 100 YEARS AGO

50 YEARS AGO

Composition with an Electronic Computer. By Prof. L. A. Hiller, Jr., and Leonard M. Isaacson — So much is heard nowadays of the capabilities of the computer robot in imitating every human activity that it is not surprising to read of experiments in which the computer is made to compose music. Perhaps to say that the machine 'composes' is putting the experiments described in this book at too high a level … Nevertheless, these experiments are amusing and the account of them is well written, so that one does not need to be a *habitué* of the computer room to understand what the authors were planning.

From *Nature* 5 December 1959.

100 YEARS AGO

My attention has just been directed to a letter which appeared in NATURE of March 11 … It was signed by Prof. McKendrick, and dealt with the vexed question of the blind and their faculties. I am a blind man, and have mixed with blind people of all ages for the past thirty years … Permit me to thank you for what you say about the popular notion that when a person loses his sight he is compensated by a gift of ability in one, if not all, his other faculties. The intelligent blind know how foolish this idea is, and constantly protest against it … We are credited with marvellous powers in music, basket-making, &c., and yet when we assert our claim to live the ordinary life of the citizen these people are shocked at our audacity … My own experience has compelled me to take heed of the varying degrees of what I shall call, for want of a better name, ear-power … When people are speaking to me, they are never on guard to control their countenance as they would be if conversing with a sighted person … I know when a person smiles, frowns, when the face lights up with an intelligence or when apathy and want of perception cloud the countenance.

From *Nature* 2 December 1909.

ASTROPHYSICS

Different stellar demise

Norbert Langer

A decades-old theory of stellar evolution — that the most massive stars end their life in a peculiar type of explosion termed a pair-instability supernova — finally seems to have been confirmed by observations.

Whereas the final evolutionary stage of a low-mass star such as the Sun is that of a simple white dwarf, the life of a massive star ends in a spectacular explosion called a supernova (SN). This theoretical view of stellar demise has been verified many times, most prominently through observations of SN 1987A, a supernova that occurred in a satellite galaxy of the Milky Way, the Large Magellanic Cloud. Neutrinos detected from SN 1987A confirmed the idea that such supernovae are triggered by the gravitational collapse of the iron core of their progenitor star[1]. But theory predicts that stars more massive than 140 solar masses could find another way to blow up — through the thermonuclear explosion of oxygen. On page 624 of this issue, Gal-Yam *et al.*[2] report that SN 2007bi provides the first evidence of such an explosion. Their study opens up the way to understanding the fate — and the mere existence — of the most massive stars, which may have dominated the early evolution of the Universe.

Stars born with a mass in the range 10–140 solar masses form a massive iron core in the final stage of their life[3]. At this point, nuclear fusion or fission ceases in the core, which collapses, owing to its own gravity, into a neutron star. Copious production of neutrinos is then thought to push the star's outer envelope away and produce a supernova. Besides uncertain neutrino interaction and transport processes, the physics of these iron-core-collapse supernovae involves poorly constrained physical processes such as turbulence, pulsations, general-relativity mechanisms, and perhaps rotation and magnetic fields[4]. It remains a challenge for supernova modellers to successfully reproduce iron-core-collapse supernovae on the computer. By contrast, models for the explosions of the most massive stars — with masses larger than about 140 solar masses — are robust.

For such very massive stars, when the core temperature at advanced evolutionary stages — namely before the ignition of oxygen — exceeds about 10^9 kelvin, photons produced in the core are sufficiently energetic to create electron–positron pairs[5]. This pair production reduces the star's radiation pressure, which would otherwise keep it in hydrostatic equilibrium, and the star collapses, igniting oxygen explosively in an event termed a pair-instability supernova (PISN; Fig. 1). Unless the mass of the core is overwhelmingly large (above 130 solar masses[6,7]), the energy released in the explosion disrupts the whole star without leaving a stellar remnant[5,6,8–10]. Although, as in the case of iron-core collapse, the core of the star has to turn a collapse into an explosion, its density and binding energy is much smaller in the PISN case, and so an explosion is much easier to obtain.

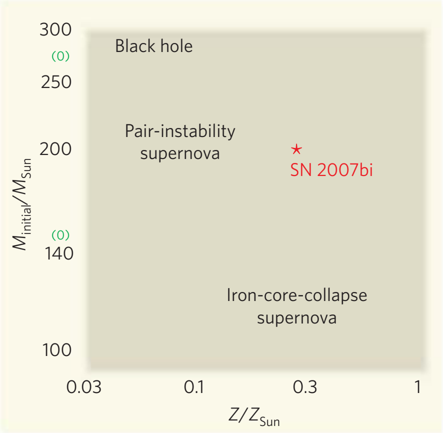


Figure 1 | Fate of the most massive stars. The ultimate fate of a massive star depends on the metallicity (the abundance of elements other than hydrogen and helium; Z) and mass with which it is born (M_{initial}); plotted values are relative to those of the Sun. The solid line represents the boundary between an iron-core-collapse supernova (triggered by the collapse of a star's iron core) and a pair-instability supernova (driven by instability caused by the production of electron–positron pairs in a star's core before oxygen fusion takes place). The blue shaded area denotes a transition region in which stars first become 'pair unstable', but eventually undergo iron-core collapse (pulsational pair-instability supernovae)[5,6]. The dotted line marks the point above which pair-unstable stars are thought to form black holes instead of exploding[5]. Lines are schematic and their exact location is uncertain. Triangles and squares denote values obtained by different theoretical studies: blue triangles and squares for finite metallicity[16]; green triangles for zero metallicity[5]. Gal-Yam and colleagues[2] estimate that the progenitor mass of supernova SN 2007bi (red asterisk) is about 200 solar masses or more.

The theoretical prediction of PISNs was never in doubt. In fact, the most recent model predictions for PISNs seem to be beautifully confirmed by Gal-Yam and colleagues' analysis[2] of SN 2007bi. In particular, the large derived amount of the radioactive isotope ^{56}Ni, the very high total mass and kinetic energy, and the slow expansion velocity and brightness evolution of SN 2007bi fit well the models of PISNs, and are incompatible with the classic iron-core-collapse supernovae.

So is SN 2007bi a textbook example of a PISN? Maybe, but it leaves us with one puzzle. First, one might wonder why it took so long (decades) to discover a PISN, which can obviously be very bright. The fact that none has been detected before did not discourage theorists because it was known that stars in our Galaxy, however massive they are at birth, would lose so much mass during their evolution, owing to radiation-driven stellar winds, that they would end up undergoing iron-core collapse[11,12]. To observe PISNs, one would have to withdraw to the early Universe, in which the abundance of metals (in astronomy, elements other than hydrogen and helium) was low — preferably even to the cosmic epoch that witnessed the formation of the very first stars, which lacked any metals. Indeed, most PISN models, including those used by Gal-Yam and colleagues[2], were constructed for metal-free stars. It was only recently that PISNs were predicted — on the basis of the most recent stellar-wind theories and provided that the metal content of the host galaxy is sufficiently low — to occur also in the local, present-day Universe (Fig. 1). The criterion of low-metal content is met by local dwarf galaxies such as the Small Magellanic Cloud. In view of this, it is reassuring that SN 2007bi was found in a metal-poor dwarf galaxy[13].

The puzzle about SN 2007bi is that no hydrogen was detected in its spectrum[2]. This means, on the one hand, that the progenitor star must have undergone substantial mass loss, but, on the other hand, that it must have avoided the drastic mass loss that massive hydrogen-free stars (called Wolf–Rayet stars) are known to suffer[14,15]. Such drastic loss would have swiftly reduced the mass of the progenitor star to below that required for a PISN. So we are left with an unlikely, finely tuned timing that led the star to explode shortly after the last of its hydrogen stores were lost to the star's wind, or — perhaps more likely — a situation in which some amount of hydrogen was still present in the supernova progenitor but was insufficient to produce a clear signal in the supernova spectrum.

Whatever the case, SN 2007bi pushes the door wide open for studies of the early, metal-poor Universe. The consequences of PISNs for the chemical evolution of metal-poor galaxies can be enormous, because a single PISN can release more metals than a whole generation of iron-core-collapse supernovae[16]. Gal-Yam and colleagues' study[2] indicates that SN 2007bi produced about 22 solar masses of silicon and more than three solar masses of radioactive nickel. This and future PISNs will also remain the only empirical way to quantify the stellar-wind mass loss in the most massive stars. Knowledge of the mere existence of stars sufficiently massive to undergo a PISN may have a considerable impact on our understanding of star formation. And, perhaps most importantly and exciting of all, finding PISNs in the local

Universe argues most convincingly that they existed — or even dominated — in the early Universe, in which stars are thought to have been more massive than they are today, and where stellar winds are predicted to have been weak. ■

Norbert Langer is at the Argelander-Institut für Astronomie, University of Bonn, Auf dem Hügel 71, 53121 Bonn, Germany.
e-mail: nlanger@astro.uni-bonn.de

1. Krauss, L. M. *Nature* **329,** 689–694 (1987).
2. Gal-Yam, A. *et al. Nature* **462,** 624–627 (2009).
3. Heger, A., Fryer, C. L., Woosley, S. E., Langer, N. & Hartmann, D. H. *Astrophys. J.* **591,** 288–300 (2003).
4. Woosley, S. & Janka, T. *Nature Phys.* **1,** 147–154 (2005).
5. Fowler, W. A. & Hoyle, F. *Astrophys. J.* (Suppl.) **9,** 201–319 (1964).
6. Heger, A. & Woosley, S. E. *Astrophys. J.* **567,** 532–543 (2002).
7. Woosley, S. E., Blinnikov, S. & Heger, A. *Nature* **450,** 390–392 (2007).
8. Bond, J. R., Arnett, W. D. & Carr, B. J. *Astrophys. J.* **280,** 825–847 (1984).
9. Glatzel, W., El Eid, M. F. & Fricke, K. J. *Astron. Astrophys.* **149,** 413–422 (1985).
10. El Eid, M. F. & Langer, N. *Astron. Astrophys.* **167,** 274–281 (1986).
11. Meynet, G., Maeder, A., Schaller, G., Schaerer, D. & Charbonnel, C. *Astron. Astrophys.* (Suppl.) **103,** 97–105 (1994).
12. Langer, N. *et al. Astron. Astrophys.* **290,** 819–833 (1994).
13. Young, D. R. *et al. Astron. Astrophys.* (in the press); preprint at http://arxiv.org/abs/0910.2248.
14. Vink, J. S. & de Koter, A. *Astron. Astrophys.* **442,** 587–596 (2005).
15. Gräfener, G. & Hamann, W.-R. *Astron. Astrophys.* **482,** 945–960 (2008).
16. Langer, N. *et al. Astron. Astrophys.* **475,** L19–L23 (2007).

NEUROSCIENCE

Unbearable lightness of touch

Liam J. Drew and Amy B. MacDermott

Following inflammation or nerve injury, stimuli that are normally perceived as innocuous can evoke persistent pain. A population of neurons that contributes to this syndrome has now been identified.

For many people with persistent pain syndromes, even the touch of something as innocuous as a shirt on their skin can be agonizing. Determining the underlying cause of this debilitating pain is essential for developing effective therapies. However, the precise identity of the neural circuits involved has so far remained elusive. In this issue (page 651), Seal *et al.*[1] pinpoint a previously enigmatic population of small sensory neurons that respond to light touch, and show that they have a crucial role in the painful sensitivity to touch or pressure that follows injury or inflammation.

The sensations of pain and touch, like other bodily sensations such as temperature and itch, begin with the activation of distinct subsets of primary sensory neurons by physical stimuli affecting the body. Once activated, these neurons excite circuits in the dorsal horn of the spinal cord. It is there that the sensory input is processed and then relayed to centres in the brain, where it is perceived and interpreted.

The dorsal horn has a laminated, or layered, structure, with each lamina receiving input from multiple but distinct classes of sensory neurons that extend from the body's periphery and convey information about various sensations (Fig. 1). These neurons release the excitatory neurotransmitter glutamate to activate neurons in the dorsal horn. Glutamate is packaged into synaptic vesicles in neurons by proteins called vesicular glutamate transporters (VGLUTs), of which there are three types. Previously, only the two most common types, VGLUT1 and VGLUT2, had been characterized in sensory neurons. VGLUT1 is expressed in touch-sensing fibres (heavily myelinated Aβ fibres with low activation thresholds) that terminate in lamina III/IV and the inner part of lamina II (IIi) in the dorsal horn[2,3]. By contrast, sensory neurons that innervate the more superficial laminae, I and II, are generally smaller and mainly associated with sensing painful (including strong mechanical) stimuli, temperature and itch. These neurons have thin, either lightly myelinated Aδ fibres or unmyelinated C fibres, most of which express VGLUT2 (ref. 3). (Myelination increases the speed at which electrical impulses pass along a neuron.)

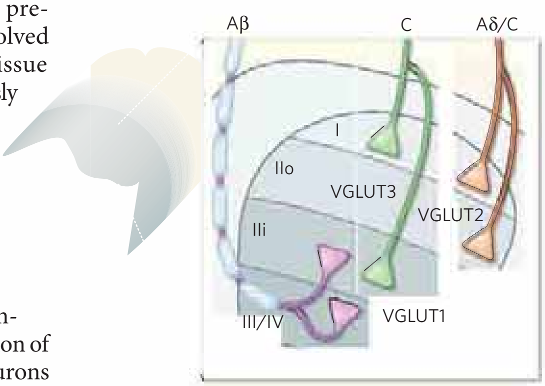


Figure 1 | Schematic view of peripheral sensory neurons terminating in the spinal-cord dorsal horn. Sensory neurons activate dorsal-horn neurons by releasing the neurotransmitter glutamate, which is packaged into synaptic vesicles by vesicular glutamate transporters (VGLUTs). Low-threshold Aβ fibres (pink) enter the dorsal horn medially, terminating in lamina III/IV and in lamina IIi. These heavily myelinated touch-sensitive fibres express VGLUT1. High-threshold, lightly myelinated Aδ fibres and unmyelinated C fibres (orange) that sense pain, temperature and itch enter the dorsal horn more laterally and terminate in laminae I and II. These fibres may express VGLUT2. The novel mechanosensors with C fibres described by Seal *et al.*[1] (green) terminate at the lamina II/III border and in lamina I, and express VGLUT3.

Seal *et al.*[1] show that the little-studied VGLUT3 is expressed in a specific subpopulation of small sensory neurons that have unmyelinated C fibres. These fibres terminate in lamina I and also in lamina IIi, where they overlap with the most dorsal innervation of conventional, touch-sensing Aβ fibres (Fig. 1). It has been known for some time that there is a small subset of sensory neurons in the C-fibre population that terminates primarily in lamina II and that responds to innocuous mechanical stimuli, such as light touch, rather than to noxious stimuli[4]. By performing electrical recordings on a preparation of sensory neurons with their fibres still attached to skin, Seal *et al.* demonstrate that the VGLUT3-expressing sensory neurons are activated by brush and light touch, and so represent these previously ill-defined, low-threshold C-fibre mechanoreceptors.

The authors then assessed the function of low-threshold C-fibre mechanoreceptors by studying mice in which the *vGlut3* gene had been knocked out[5]. The absence of VGLUT3 functionally disconnects this subset of peripheral sensory neurons from their dorsal-horn targets by preventing the normal release of glutamate. Surprisingly, in the absence of injury or inflammation, these knockout mice had small defects in their responses to noxious mechanical stimuli, but responded normally to low-intensity mechanical stimuli and also to hot and cold stimuli. This somewhat paradoxical result suggests that, despite the low-threshold mechanical sensitivity of the VGLUT3-expressing neurons, they can contribute to the detection of noxious mechanical pain.

Seal and colleagues' most striking findings[1], however, came in the next round of tests, in which the authors used three manipulations in mice as models of persistent pain states. These comprised: injection of an irritant substance into the footpad, causing a massive inflammatory response; an incision in the paw that is considered a model of post-surgical pain; and a lesion of the peripheral nerve to model neuropathic pain. In the last case, injury to peripheral nervous tissue causes a persistent pain state, best defined[6] as a pathological response of the pain system to nerve damage. In all of these conditions, animals whose paws are exposed to mechanical and thermal stimulation will usually tolerate a far lower stimulation intensity than normal mice before withdrawing their paws. In all three models, the VGLUT3-knockout mice and normal control mice showed similar hypersensitive responses to heat. However, knockout animals showed much less hypersensitivity to mechanical stimulation than did control animals. In fact, in the model of inflammation, the VGLUT3-knockout mice did not develop mechanical hypersensitivity. Thus, preventing synaptic input from these low-threshold C-fibre mechanoreceptors seems to remove a crucial mechanical input that is either partially or completely required for responses to mechanical touch in these hypersensitive states.

One of many questions highlighted by this work is why two very different types of low-threshold mechanoreceptor — a subset of Aβ fibres and a subset of VGLUT3-expressing C fibres — both terminate in the same lamina IIi region of the dorsal horn. This region has been implicated[7,8] as a key processor of low-threshold inputs that can become painful when there is inflammation or nerve injury. Understanding the physiological changes occurring in this region in such conditions may be crucial for understanding how normally innocuous inputs can activate pain pathways.

Why are there two distinct low-threshold inputs? If we consider that the Aβ-fibre input is important for identifying the location of touch, perhaps the C-fibre input instead works to amplify the Aβ-mediated responses under injury conditions and so to modify the accompanying protective responses and perception of pain. The authors note that, under normal conditions in humans, activation of low-threshold mechanoreceptor C fibres by brushing correlates with the sensation of pleasurable touch[9]. Although it remains to be confirmed that the VGLUT3 neurons are responsible for sensing pleasurable touch in humans, a sensory subsystem that can change from eliciting feelings of pleasure to evoking pain following inflammation or injury would indicate a hitherto unexpected level of plasticity in an already remarkably flexible system. ■

Liam J. Drew is in the Department of Physiology and Cellular Biophysics, and Amy B. MacDermott is in the Department of Physiology and Cellular Biophysics, and the Department of Neuroscience, Columbia University, New York, New York 10032, USA.
e-mails: ljd2109@columbia.edu; abm1@columbia.edu

1. Seal, R. P. *et al. Nature* **462,** 651–655 (2009).
2. Neumann, S., Braz, J. M., Skinner, K., Llewellyn-Smith, I. J. & Basbaum, A. I. *J. Neurosci.* **28,** 7936–7944 (2008).
3. Todd, A. J. *et al. Eur. J. Neurosci.* **17,** 13–27 (2003).
4. Sugiura, Y., Lee, C. L. & Perl, E. R. *Science* **234,** 358–361 (1986).
5. Seal, R. P. *et al. Neuron* **57,** 263–275 (2008).
6. Costigan, M., Scholz, J. & Woolf, C. J. *Annu. Rev. Neurosci.* **32,** 1–32 (2009).
7. Malmberg, A. B., Chen, C., Tonegawa, S. & Basbaum, A. I. *Science* **278,** 279–283 (1997).
8. Miraucourt, L. S., Dallel, R. & Voisin, D. L. *PLoS ONE* **2,** e1116 (2007).
9. Löken, L. S., Wessberg, J., Morrison, I., McGlone, F. & Olausson, H. *Nature Neurosci.* **12,** 547–548 (2009).

STRUCTURAL BIOLOGY

Steps in the right direction

Smita S. Patel

The ring-shaped helicase enzyme Rho moves along RNA using ATP as an energy source. Coordinating ATP hydrolysis with nucleic-acid binding seems to determine the direction and mechanism of helicase movement.

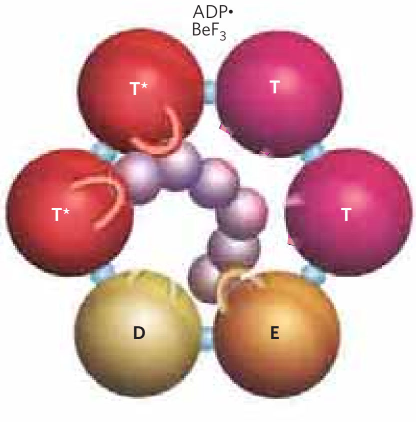


Figure 1 | Structure of the Rho hexameric helicase. Thomsen and Berger's structure[2] of Rho bound to RNA and ATP mimics (ADP•BeF$_3$) shows that Rho's six subunits interact with the RNA (purple) in the central channel through the Q and R loops (shown here as one loop per subunit). The ATP mimics are bound at the subunit interfaces. The different colours of the spheres depict the various conformational states of their ATP-binding site. Hydrolysis-competent T* subunits bind ATP and RNA tightly. The T subunits (before ATP hydrolysis), E subunit (involved in nucleotide exchange) and D subunit (bound to the product of ATP hydrolysis) all bind ATP and the central RNA weakly. The T and E subunits each interact with one nucleotide of RNA at a time, whereas the D subunit has limited interactions with RNA.

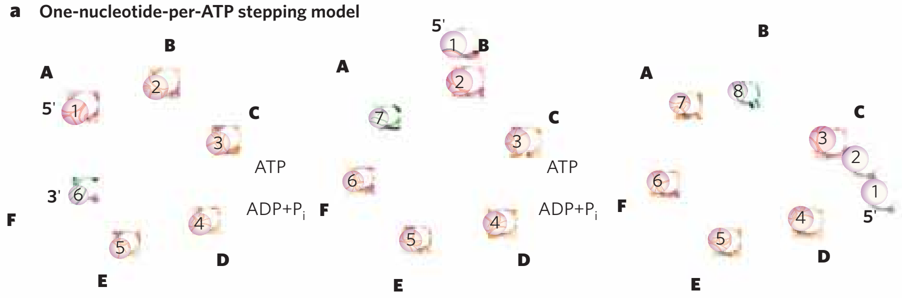
Helicases are molecular motors, fuelled by energy from ATP hydrolysis, that move along nucleic-acid molecules opening up double strands of DNA, RNA or DNA–RNA hybrids. These enzymes function in fundamental cellular processes such as genome replication, recombination and transcription[1]. The six subunits of the hexameric helicases are arranged in a ring that encircles a nucleic-acid strand. The ATP-binding sites, which catalyse ATP hydrolysis, are located at the interfaces between each subunit.

Deciphering how the six subunits coordinate their ATPase cycles to move the ring in a specific direction has been a major research topic, one that is advanced by two studies investigating the hexameric helicase Rho. In *Cell*, Thomsen and Berger[2] report a high-resolution structure of Rho bound to RNA and ATP mimics. The structure suggests that ATP is hydrolysed in an ordered, sequential manner, and that the helicase tracks along RNA taking steps of one nucleotide per molecule of ATP hydrolysed. Meanwhile, Schwartz *et al.*[3] investigate Rho movement using a chemical method, and report in *Nature Structural and Molecular Biology* that Rho makes essential interactions with the RNA every seven nucleotides, which suggests a much larger step size.

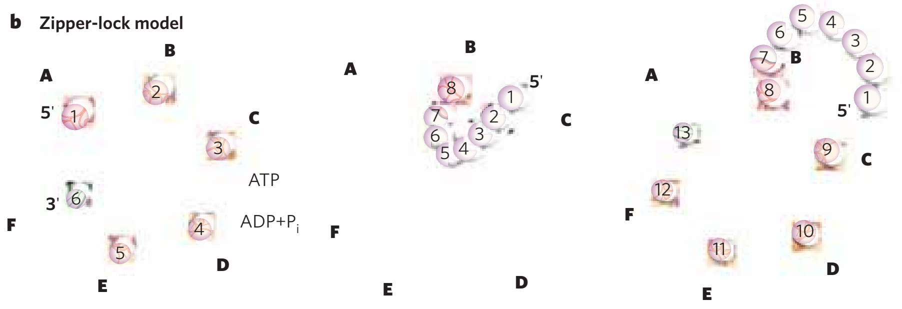
The Rho helicase is involved in terminating transcription in bacteria. When a copy of RNA is made from DNA, a short segment of the RNA at the leading edge of synthesis is paired with the DNA template to form a stable RNA–DNA hybrid within the active site of the transcribing RNA polymerase. To release the newly formed RNA, the Rho helicase binds to the RNA, moves along it in the 5′-to-3′ direction, and dissociates the RNA transcript from the RNA polymerase by disrupting the RNA–DNA complex.

A previous structure[4] captured Rho as an open ring, with RNA wrapped around the primary RNA-binding sites involved in initiation of transcription termination. In this structure, the RNA was absent from the central channel that couples ATP hydrolysis[5] to the conformational changes required for movement along nucleic acids. Thomsen and Berger's structure[2] captures Rho in a closed-ring conformation with six nucleotides of RNA bound in the central channel of the ring, and with six ATP mimics (ADP•BeF$_3$) bound to the ATPase sites (Fig. 1).

The present structure[2] provides a snapshot of the ring moving along RNA, and shows that each of the six helicase subunits interacts differently with the RNA, which makes a complete turn in the central channel. Two RNA-binding elements, the Q and R loops, extend from each subunit of the Rho hexamer into the central channel, with several amino-acid residues from each of the loops associating with the backbone ribose and phosphates of the RNA. Interaction of the Q loop with the 2′-hydroxyl of the ribose of RNA explains the strict specificity of Rho for RNA. These RNA-binding elements organize into a spiral staircase to track the RNA backbone in a manner that is remarkably similar to that observed in the viral DNA-processing hexameric helicase E1 (ref. 6). As E1 and Rho



a One-nucleotide-per-ATP stepping model



b Zipper-lock model

Figure 2 | Two models of hexameric helicase nucleic-acid translocation. As indicated by the crystal structure[2], five nucleotides of RNA (spheres) interact with five subunits of the helicase (hands A–E). The red closed hand is the subunit that has the tightest grip on the RNA, and is activated for ATP hydrolysis. The subunits represented by orange hands are in various tight and weak binding states. The green, open hand is the D subunit, which has limited interactions with RNA. **a**, The one-nucleotide-per-ATP stepping model proposes that, with every ATP-hydrolysis cycle, the subunit with the tightest grip on RNA (red hand) releases one nucleotide of RNA from the 5′ end, and a subunit grabs one nucleotide of RNA from the 3′ end (green hand). Thus, Rho moves along RNA in the 5′-to-3′ direction with steps of one nucleotide per ATP hydrolysed. **b**, In the zipper-lock model, which involves movement along more than one nucleotide per ATP hydrolysed, the tightly bound subunit functions as the zipper-lock. With every ATP-hydrolysis cycle, the zipper is unlocked, which causes a global change in the RNA-binding staircase (orange hands become blue), releasing a large segment of RNA from the 5′ end. A new zipper-lock is formed at an adjacent subunit, which restores the RNA-binding staircase, allowing it to spontaneously bind a new segment of RNA from the 3′ end. Thus, by fluctuating between tight and weak states, the hexamer ring translocates more than one nucleotide per ATP hydrolysed. P$_i$, inorganic phosphate.

share little sequence similarity, these findings indicate that hexameric helicases have converged on a common mode of nucleic-acid binding.

The fully liganded Rho structure[2] has six ATP mimics bound at the subunit interfaces (Fig. 1). Each of the sites binds ATP either tightly (T* state in Fig. 1) or weakly (T, E or D states). The subunits that bind ATP tightly also bind RNA tightly, whereas the subunits that bind ATP weakly bind RNA weakly. Rho helicase moves in the 5′-to-3′ direction along RNA, whereas E1 moves in the 3′-to-5′ direction along its DNA substrate. Knowledge of the structures of two hexameric helicases that move in opposite directions on nucleic acid represents a unique opportunity to examine the structural basis of directionality.

One might assume that helicases moving on nucleic acid in opposite directions would bind nucleic acid in opposite orientations. Surprisingly, this is not the case. Rho and E1 bind nucleic acid with the same relative polarity, and the chirality of their amino-acid spiral staircase is also the same. But the order of ATPase sites with weakly and tightly bound ATPs around the ring in E1 is the opposite in Rho. This suggests that the cycles of ATP hydrolysis that drive movement of the helicase along the nucleic acid might occur in opposite directions around the ring in the two enzymes. Moreover, in contrast to Rho, the subunits of E1 that make tight interactions with DNA are those that have hydrolysed ATP (rather than those that bind it tightly). This raises an intriguing alternative possibility to explain the opposing directions of translocation — the on and off switches for nucleic-acid binding and release might be reversed in the two helicases.

Although the sequence of ATP binding and hydrolysis cannot be established from a single structural snapshot, the arrangement of Thomsen and Berger's Rho structure[2] supports the ordered sequential model of ATP hydrolysis and translocation (the one-nucleotide-per-ATP model)[6]. Conceptually, this model of helicase function is similar to the binding change or rotational catalysis mechanism of the ATP synthase[7], which has sequence and structural similarity to the Rho protein. The one-nucleotide-per-ATP model[2,6] (Fig. 2a) proposes that each Rho subunit hydrolyses ATP sequentially around the ring. After ATP hydrolysis at the subunit that is tightly bound to RNA, one nucleotide of RNA is released at the 5′ end from the subunit at the top of the amino-acid staircase. ATP binding and product release in subunits at the bottom of the staircase trigger binding, in an orchestrated manner, of one nucleotide from the 3′ end of RNA. In this way, a single cycle of ATP hydrolysis leads to movement with a step of one nucleotide in the 5′-to-3′ direction.

The one-nucleotide stepping model predicts that each nucleotide of RNA goes through uniform interactions with the helicase subunits during translocation. The biochemical studies described by Schwartz *et al.*[3] challenge the predictions of this model. The authors used a chemical-interference method to probe functionally essential RNA interactions that occur as Rho tracks along and unwinds the RNA–DNA duplex. Surprisingly, they found that Rho interacts with the 2′-hydroxyl group of the RNA only about every seven nucleotides. This seems inconsistent with the one-nucleotide-per-ATP stepping mechanism[2,6]. To explain the large step size, Schwartz and colleagues[3] instead propose a mechanism similar to that used by the ring-shaped, ATP-powered motor of the bacterial virus phage Φ29, which packages DNA into its empty phage capsid[8]. The authors[3] suggest that Rho fluctuates between open and closed ring structures and that the two subunits at the gap in the open structure act as a 'latch' and 'lever' to bind and release RNA periodically with large ~7-nucleotide movements.

To reconcile these structural[2] and biochemical[6] findings, a modified version of the latch–lever model, dubbed the zipper-lock model (Fig. 2b), is proposed here. As with the one-nucleotide stepping model, the zipper-lock model assumes that each subunit transitions through a round of ATP binding, hydrolysis and product release and that this occurs in a sequential manner around the ring.

In contrast to the one-nucleotide stepping model, which assumes that ATP hydrolysis causes a local conformational change that releases one nucleotide of RNA, the zipper-lock model posits that ATP hydrolysis causes a global conformational change in the ring that releases more than one nucleotide of RNA. The tightly binding subunit acts like a zipper-lock (latch) to hold the RNA in place. The zipper is unlocked after ATP hydrolysis at the locking subunit, which disrupts the RNA-binding staircase, unzipping the remaining protein–RNA interactions, while an adjacent subunit forms a new zipper-lock with a distant RNA nucleotide (Fig. 2b). The new zipper-lock reorganizes the amino-acid staircase to spontaneously bind a new, equally large segment of RNA from the 3′ end. The ~7-nucleotide periodicity in the chemical-interference pattern[3] can be explained by clockwise ATP hydrolysis and periodic zipper-lock formation with roughly every seventh nucleotide of RNA. Variations of this model can explain helicase steps ranging from one nucleotide to seven nucleotides. Greater than one-nucleotide movement per ATP molecule hydrolysed is also consistent with biochemical studies[9] of the hexameric T7 gp4 helicase.

The structures of the fully liganded hexameric helicases[2,6] represent a turning point for understanding the workings of ring-shaped helicases. And the two studies of Rho[2,3] offer challenges for mechanistic deliberations and investigations to fully understand how ring-shaped helicases move on nucleic acids. Future work will involve obtaining additional snapshots of the hexameric helicase in various ATP-liganded states to sort out the directionality of ATP hydrolysis around the ring; applying new methods to elucidate the dynamics of the nucleic-acid translocation process; and measuring the chemical and physical step sizes of the hexameric helicases by biochemical and single-molecule kinetics. ■

Smita S. Patel is in the Department of Biochemistry, UMDNJ-Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, USA.
e-mail: patelss@umdnj.edu

1. Patel, S. S. & Picha, K. M. *Annu. Rev. Biochem.* **69,** 651–697 (2000).
2. Thomsen, N. D. & Berger, J. M. *Cell* **139,** 523–534 (2009).
3. Schwartz, A. *et al. Nature Struct. Mol. Biol.* doi:10.1038/nsmb.1711 (2009).
4. Skordalakes, E. & Berger, J. M. *Cell* **114,** 135–146 (2003).
5. Wei, R. R. & Richardson, J. P. *J. Biol. Chem.* **276,** 28380–28387 (2001).
6. Enemark, E. J. & Joshua-Tor, L. *Nature* **442,** 270–275 (2006).
7. Boyer, P. D. *Annu. Rev. Biochem.* **66,** 717–749 (1997).
8. Moffitt, J. R. *et al. Nature* **457,** 446–450 (2009).
9. Donmez, I. & Patel, S. S. *EMBO J.* **27,** 1718–1726 (2008).

ATOMIC PHYSICS

Neutral atoms put in charge

Martin Zwierlein

An elegant experiment shows that atoms subjected to a pair of laser beams can behave like electrons in a magnetic field, as demonstrated by the appearance of quantized vortices in a neutral superfluid.

Ultracold gases of atoms — a million times thinner than air and a million times colder than interstellar space — allow the observation and control of many-body quantum phenomena at macroscopic scales. They can thus serve as model materials[1] for condensed-matter systems in which such phenomena arise. From superfluids (fluids that flow without friction) to insulators, and from weakly to strongly interacting systems, a great variety of fundamental states of matter can be realized, observed in real time and probed with the precision of atomic physics. But there seems to be one obvious limitation: atoms are neutral, a fact that in principle precludes the observation of a wealth of phenomena tied to charged particles, for example their behaviour in a magnetic field. On page 628 of this issue, Lin *et al.*[2] get round this problem and in striking fashion demonstrate their ability to create synthetic magnetic fields for a neutral ultracold atomic system. In their study, a collective state of matter termed a Bose–Einstein condensate develops quantized vortices of magnetic flux similar to mini-tornadoes.

Magnetic fields are directly tied to the rotational motion of charged particles: an electron placed in a uniform magnetic field rotates about the field axis. One way to mimic the effect of a magnetic field on a cloud of neutral atoms is thus to make the cloud rotate. But this approach comes with limitations. Achieving the equivalent of a large magnetic field requires fast rotation, which causes atoms to fly apart. Also, the method requires rotationally symmetrical confinement of the gas and cannot be applied to atoms trapped in a static optical lattice — an artificial 'crystal' of light used to model solid-state systems[1]. A more direct way of mimicking effective magnetic fields should thus not be restricted to certain system geometries and configurations.

To get a picture of what is needed, consider what happens to a paddle wheel placed in a river (Fig. 1). If the river flows with uniform velocity across the paddle wheel, it will not rotate. But if the water on one side of the paddle wheel flows faster than that on the other side or in a different direction, or if there is a swirl (vortex) in the flow, the paddle wheel will rotate. In this analogy, the axis and speed of rotation represent the direction and strength of the magnetic field, and the flow pattern is called a vector potential. Just as the paddle wheel rotates in the presence of vortices in the flow, so a cloud of atoms will experience a magnetic field if a vector potential that carries vorticity is imprinted on the system.

This is precisely what Lin and colleagues[2] achieved in their experimental set-up, which is an inspired realization of earlier proposals[3–6]. Using a pair of laser beams, they first imprinted the equivalent of a river's uniform flow — a uniform vector potential — on their ultracold cloud of atoms[7,8]. Next, by judiciously tuning the lasers in a spatially dependent way, they endowed the vector potential with a swirl: atoms on different sides of the cloud experienced a different vector potential. This non-uniform flow created a synthetic magnetic field.

Quantum mechanics predicts a wide range of peculiar effects for systems subjected to magnetic fields. One dramatic example, the Meissner effect, is found in superconductors, in which currents of electron pairs flow without resistance. In the Meissner effect, a superconductor placed in a magnetic field expels the field from its interior. But beyond a certain critical field strength, many superconductors, termed type II superconductors, allow magnetic flux to pierce their interior in the form of tiny quantized vortices or flux tubes. This is a direct consequence of the fact that

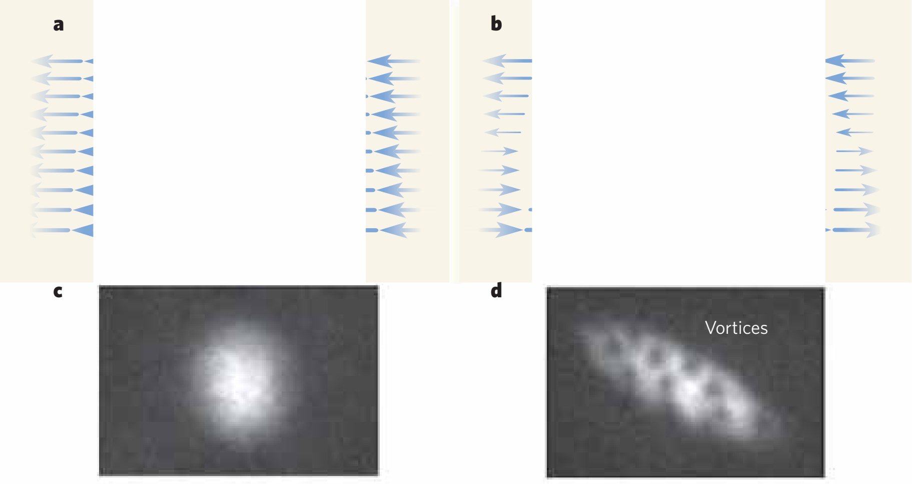


Figure 1 | Paddle-wheel analogy for magnetic fields. a, A paddle wheel placed in water flowing with uniform velocity (blue arrows) does not rotate. **b,** If the water flows non-uniformly and in a different direction on one side of the paddle wheel from that on the other, the wheel rotates (green arrow). The flow pattern is analogous to the 'vector potential' created in Lin and colleagues' experiment[2] to generate a synthetic magnetic field for an ultracold cloud of atoms known as a Bose–Einstein condensate. The rotation axis (purple arrow) and the speed with which the paddle wheel rotates correspond respectively to the direction and strength of the magnetic field. **c, d,** Images of the Bose–Einstein condensate before (**c**) and after (**d**) application of the synthetic magnetic field. The appearance of quantized vortices (**d**) is a direct demonstration of a synthetic magnetic field.

superconductors can be described by a single collective macroscopic wavefunction that is shared among all of their constituent electron pairs. Such 'flux quantization' is the principle behind the working of SQUIDs (superconducting quantum interference devices), the extremely accurate magnetometers that are used to measure small magnetic fields.

Superconductors are nothing other than charged superfluids. The behaviour of superconductors in a magnetic field is thus directly analogous to that of neutral superfluids under rotation. Observations of a rotating lattice of vortices in neutral superfluids such as Bose–Einstein condensates[9] and gases of fermionic atoms[10] (particles with half-integer spin, such as electrons) have in fact served as direct proof of these systems' superfluidity. Lin *et al.*[2] applied their novel technique of creating synthetic magnetic fields, which do not require the system to rotate, to a Bose–Einstein condensate, and found that the system develops a striking array of vortices — the smoking gun for magnetic-flux quantization. This is the first time that a stable, long-lived, non-rotating array of vortices has been formed in a neutral superfluid.

In charged systems, many intriguing magnetic phenomena arise at large magnetic fields. One example is given by fractional quantum Hall states — collective states of matter in which electrons behave as if their elementary charge is only a fraction of their actual charge. Reaching the required field strengths in neutral systems is, however, probably beyond the reach of experiments that rely on rotation to mimic the effect of such fields. Lin and colleagues' experiment[2] offers a new approach to generate synthetic fields and might point the way towards observing fractional quantum Hall and other exotic states of matter in neutral systems. Of course, the task of extending the authors' technique to strong magnetic fields won't be easy. Their technique also comes with shortcomings: for example, the finite lifetime of the atomic cloud in the presence of the laser beams; a limit on the highest possible vector potential; and the difficulty of extending the approach to fermionic systems. That said, the challenges that accompany the authors' technique might be easier to tackle than those that plague the rotational approach.

Lin and colleagues' spectacular demonstration[2] of a synthetic magnetic field for neutral atoms signals the advent of synthetic electrodynamics in the field of ultracold atomic gases. Future applications of their method might include a measurement of the superfluid fraction in ultracold atomic gases[11] and the creation of unusual quantum states in two-dimensional optical lattices at high effective fields[5]. The demonstration of quantum Hall physics in such lattices[6] using the authors' approach might also be within reach. Their work opens up exciting avenues for producing novel many-body quantum systems. ■

Martin Zwierlein is at the MIT-Harvard Center for Ultracold Atoms, Research Laboratory of Electronics, and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mail: zwierlein@mit.edu

1. Bloch, I., Dalibard, J. & Zwerger, W. *Rev. Mod. Phys.* **80,** 885–962 (2008).
2. Lin, Y.-J., Compton, R. L., Jiménez-García, K., Porto, J. V. & Spielman, I. B. *Nature* **462,** 628–632 (2009).
3. Berry, M. V. *Proc. R. Soc. A* **392,** 45–57 (1984).
4. Higbie, J. & Stamper-Kurn, D. M. *Phys. Rev. Lett.* **88,** 090401 (2002).
5. Jaksch, D. & Zoller, P. *New J. Phys.* **5,** 56 (2003).
6. Sørensen, A. S., Demler, E. & Lukin, M. D. *Phys. Rev. Lett.* **94,** 086803 (2005).
7. Lin, Y.-J. *et al. Phys. Rev. Lett.* **102,** 130401 (2009).
8. Spielman, I. B. Preprint at http://arxiv.org/abs/0905.2436 (2009).
9. Fetter, A. L. *Rev. Mod. Phys.* **81,** 647–691 (2009).
10. Zwierlein, M. W. *et al. Nature* **435,** 1047–1051 (2005).
11. Cooper, N. R. & Hadzibabic, Z. Preprint at http://arxiv.org/abs/0910.4767 (2009).

CELL BIOLOGY

Stairway to the proteasome

Malavika Raman and J. Wade Harper

The study of fast and intricate enzyme reactions requires methods that have the speed and sophistication to match. Such an approach reveals the way in which proteins are tagged with ubiquitin for destruction.

The ubiquitin–proteasome system is the primary pathway through which the stability of cellular regulatory proteins is controlled. Proteins marked with a chain of ubiquitin molecules are recognized by a cellular machine called the proteasome, and are rapidly degraded. The ubiquitins are added by an E1–E2–E3 enzyme cascade. This simple model has dominated the field for several years, but the actual mechanisms that control the rates and types of ubiquitin-chain extensions have remained poorly understood[1]. In particular, a minimum of four ubiquitins must be added to the substrate destined for destruction before it can be delivered to the proteasome[2]. As such, four ubiquitins must be added during a single encounter with the E3, lest the substrate dissociate prematurely and its ubiquitins be rapidly removed by cellular deubiquitylating enzymes.

There are two competing models for this chain extension — stepwise addition of ubiquitin monomers, and *en-bloc* transfer, wherein the ubiquitin chain is first built on the enzymatic machinery before being transferred to the substrate in a single step[1,3,4]. Differentiating between these models has been difficult owing to the speed of ubiquitin-transfer reactions. On page 615 of this issue[5] and in a companion paper in *Cell*[6], Raymond Deshaies and colleagues describe how they have dissected the ubiquitin-transfer reaction on a millisecond timescale, and provide compelling evidence in favour of the stepwise model.

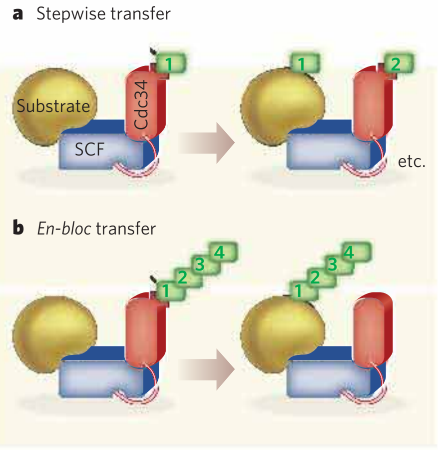


Figure 1 | Models of ubiquitin transfer to a substrate protein that requires degradation. a, Stepwise transfer. The E1 enzyme (not shown) charges E2 (Cdc34, red) with one ubiquitin (1, 2, etc.) at a time. In a reaction mediated by E3 (SCF, blue), these are then singly transferred to the substrate to build the minimal chain of four ubiquitins required for recognition by the proteasome. The full sequence of events is shown in Figure 2 (overleaf). **b,** *En-bloc* transfer. Cdc34 is charged with multiple ubiquitins, with SCF mediating the *en-bloc* addition of the entire ubiquitin chain to the substrate.

Ubiquitylation of proteins occurs via transfer of ubiquitin from a ubiquitin-activating enzyme (E1) to a ubiquitin-conjugating enzyme (E2), which, in collaboration with a substrate-selective ubiquitin ligase (E3), covalently attaches ubiquitin to a lysine amino-acid residue on the substrate. The primary E2–E3 system employed by Deshaies and colleagues is the Cdc34–$\mathrm{SCF}^{\mathrm{Cdc4}}$ complex, whose substrates include cyclin E (CycE). $\mathrm{SCF}^{\mathrm{Cdc4}}$ is the prototypical form of a family of enzymes known as cullin-RING ubiquitin ligases (CRLs)[7,8], and has been used as a model for some 300 CRLs in mammals[9].

Using an assay that allowed visualization of the number of ubiquitin molecules attached to a model CycE peptide substrate, Pierce *et al.*[5] observed that three or more ubiquitins are linked within 30 seconds of mixing the CycE–$\mathrm{SCF}^{\mathrm{Cdc4}}$ complex with ubiquitin-charged Cdc34 (Cdc34~Ub, where ~ represents a

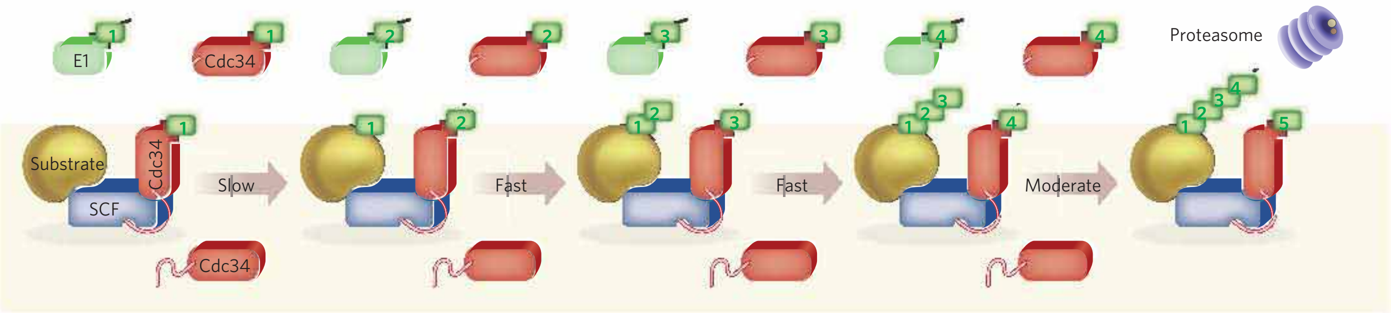


Figure 2 | The stepwise model of ubiquitin (Ub) transfer, validated by Deshaies and colleagues[5,6]. The E1 ubiquitin-activating enzyme charges E2 (Cdc34, red) with ubiquitin (1, 2, etc.), which in turn transfers the ubiquitin sequentially to substrate bound to E3 (SCF, blue), a cullin-RING ligase. Rapid binding cycles between the cellular pool of Cdc34~Ub and SCF is achieved via interaction between the acidic tail of Cdc34 and a basic 'canyon' on the cullin component of the SCF (shown as a notch). This electrostatic interaction enables Cdc34~Ub to rapidly bind SCF and processively ubiquitylate the substrate in a stepwise manner, before the substrate dissociates and is taken to the proteasome for degradation. The relative rates of ubiquitin transfer (k_{Ub}) are shown for each step in the process. The light blue circle depicts the Rbx1 site of SCF association with Cdc34~Ub.

thioester bond between Cdc34 and ubiquitin). However, the precise manner in which Cdc34 transfers ubiquitin to CycE — stepwise, *en bloc* or combinations thereof — was not discernible. So the authors developed a theoretical model to test potential pathways of chain elongation, on the basis of the observed number of ubiquitins conjugated to CycE during a single encounter with the SCF. From a consideration of the distribution of ubiquitin molecules on a pre-assembled chain, and the number of transfer events from Cdc34~Ub to substrate, it was apparent that only stepwise or *en-bloc* transfer was occurring.

With stepwise transfer, one would expect to observe Cdc34 molecules charged with a single ubiquitin, whereas the accumulation of charged Cdc34 molecules containing multiple ubiquitins would favour the *en-bloc* model (Fig. 1). Using mass spectrometry, Pierce *et al.*[5] found that Cdc34 is linked with a single ubiquitin *in vitro*, consistent with stepwise transfer. These results are in agreement with studies[10] indicating that human Cdc34 is charged with a single ubiquitin *in vivo*.

To distinguish further between stepwise and *en-bloc* transfer, Pierce *et al.*[5] performed mixing reactions on a millisecond timescale in a quench-flow device, and subsequently visualized the reaction products. In this setting, the CycE substrate is modified with a single ubiquitin within 10 milliseconds, and subsequent conjugates appear sequentially. These data, along with the fact that Cdc34 is charged with a single ubiquitin, argue that Cdc34 transfers ubiquitin to the substrate in a stepwise manner (Fig. 2). In extending the approach to SCF$^{\beta\text{-TrCP}}$ and its substrate β-catenin, Pierce *et al.* show that the stepwise mechanism is generalizable to other CRLs.

Several observations emerge from this study that potentially explain how CRLs modulate substrate degradation half-lives *in vivo*. First, most SCF–substrate encounters may be unproductive because the substrate dissociation rate (k_{off}) is faster than the rate of transfer of the first ubiquitin (k_{Ub1}). Second, the rate of ubiquitin transfer is not constant. Once a substrate is monoubiquitylated, the rate of chain elongation increases dramatically before eventually declining (Fig. 2), thereby allowing multiple ubiquitin transfers before substrate dissociation[5,11]. This decline in rate probably reflects topological constraints conferred by the growing ubiquitin chain. Third, individual k_{Ub1} values differ for SCF$^{\mathrm{Cdc4}}$ and SCF$^{\beta\text{-TrCP}}$ by an order of magnitude, whereas k_{off} values for their individual peptide substrates are similar regardless of ubiquitin-chain length[5]. A substrate with a larger k_{Ub1} and a smaller k_{off} is more likely to productively encounter the SCF. Given the diversity of substrates and mechanisms of recognition by CRLs, it seems likely that the k_{Ub1} and k_{off} values, and hence rates of polyubiquitylation and turnover, will vary significantly with full-length substrates *in vivo*.

The term 'processivity' describes the number of ubiquitins transferred before the substrate dissociates, and the process is key to the production of chains that are long enough for proteasomal targeting. Because there is a single RING domain in the subunit (Rbx1) of the CRL used for Cdc34~Ub recruitment, discharged Cdc34 must dissociate from the CRL after ubiquitin transfer to allow for recruitment of the next Cdc34~Ub. Processivity does not depend on the rate of Cdc34 recharging as such, because most Cdc34 in the cell is pre-charged with ubiquitin under steady-state conditions[10]. However, Cdc34~Ub binds tightly to the SCF, raising the question of how Cdc34 balances affinity with processivity. Insight into this conundrum comes from an analysis of Cdc34–SCF interaction kinetics by Kleiger and colleagues[6].

In addition to the well-known interaction between Cdc34 and Rbx1, Kleiger *et al.*[6] identify a new interaction between an acidic carboxy-terminal tail on Cdc34 and a basic 'canyon' in the cullin subunit (Fig. 2). Remarkably, association of Cdc34 with SCF is about two orders of magnitude faster than the diffusion-controlled limit. Coupled with a fast dissociation rate and the abundance of Cdc34~Ub, this allows for many encounters with substrate-bound SCF before substrate dissociation. Mutations in either the basic canyon or the acidic tail substantially impede processivity[6], indicating that electrostatic complementarity has a key role in increasing the frequency of productive encounters. Thus, engagement of Cdc34~Ub by the SCF can be viewed as involving two major features: initial rapid recognition, perhaps depending mainly on the tail–canyon interaction, and a secondary positioning interaction via Rbx1 that orients the ubiquitin thioester for transfer to the substrate. The finding that the canyon exists across the cullin protein family makes it likely that this kinetic mechanism will be used by all CRLs.

CRL enzymes surmount seemingly impossible odds to polyubiquitylate substrates, thereby establishing the dynamics of much of the proteome. By exploiting small regions of charge conservation, CRLs have evolved to maintain high-affinity interactions as well as rapid on/off kinetics to sequentially ubiquitylate substrates. The acidic tail of Cdc34 is unique among the cellular E2 enzymes examined, so further studies will be required to determine the extent to which analogous mechanisms are at work for other E2–E3 pairs. The new papers[5,6] provide a framework for examining this question, while delivering an unprecedented view of how Cdc34–SCF generates a stairway to the proteasome. ■

Malavika Raman and J. Wade Harper are in the Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA.
e-mail: wade_harper@hms.harvard.edu

1. Ye, Y. & Rape, M. *Nature Rev. Mol. Cell Biol.* **10,** 755–764 (2009).
2. Thrower, J. S. *et al. EMBO J.* **19,** 94–102 (2000).
3. Li, W., Tu, D., Brunger, A. T. & Ye, Y. *Nature* **446,** 333–337 (2007).
4. Ravid, T. & Hochstrasser, M. *Nature Cell Biol.* **9,** 422–427 (2007).
5. Pierce, N. W., Kleiger, G., Shan, S. & Deshaies, R. J. *Nature* **462,** 615–619 (2009).
6. Kleiger, G., Saha, A., Lewis, S., Kuhlman, B. & Deshaies, R. J. *Cell* **139,** 957–968 (2009).
7. Skowyra, D. *et al. Cell* **91,** 209–219 (1997).
8. Feldman, R. M., Correll, C. C., Kaplan, K. B. & Deshaies, R. J. *Cell* **91,** 221–230 (1997).
9. Petroski, M. D. & Deshaies, R. J. *Nature Rev. Mol. Cell Biol.* **6,** 9–20 (2005).
10. Jin, J., Li, X., Gygi, S. P. & Harper, J. W. *Nature* **447,** 1135–1138 (2007).
11. Petroski, M. D. & Deshaies, R. J. *Cell* **123,** 1107–1120 (2005).

Vol 462|3 December 2009|**doi:10.1038/nature08533**

Forcing cells to change lineages

Thomas Graf[1] & Tariq Enver[2]

The ability to produce stem cells by induced pluripotency (iPS reprogramming) has rekindled an interest in earlier studies showing that transcription factors can directly convert specialized cells from one lineage to another. Lineage reprogramming has become a powerful tool to study cell fate choice during differentiation, akin to inducing mutations for the discovery of gene functions. The lessons learnt provide a rubric for how cells may be manipulated for therapeutic purposes.

Seemingly at odds with the stability of the differentiated state in metazoa are cell fusion and nuclear transfer experiments, which have shown that the epigenomes of differentiated cells can be remarkably plastic. Experiments performed several decades ago showed that dormant gene expression programs can be dominantly awakened in differentiated cells by the fusion of different pairs of cell types[1]. Subsequently, lineage conversions could be effected simply through the introduction of defined transcription factors[2,3] (Fig. 1a, b). Parallel experiments, conducted in a number of different species, showed that transfer of nuclei from both embryonic and adult somatic cell types can lead to the formation of all three germ layers and even to the generation of entire new animals[4–7], unequivocally demonstrating that the identity of differentiated cells can be fully reversed. The latest and most dramatic development is the demonstration that somatic cells can be reprogrammed to a pluripotent state by the expression of a transcription factor cocktail, generating induced pluripotent stem (iPS, for nomenclature see Box 1) cells[8] (Fig. 1c).

The facility with which cell fates can be altered experimentally raises the question as to whether such interconversions occur physiologically or in the context of disease. Arguably, gastrulation provides a first example of transdetermination, where an invagination of the ectoderm produces mesoderm (reviewed in refs 9, 10). Transdetermination and transdifferentiation may also have a role in regeneration, metaplasia and cancer. For example, removing the eye lens of a newt leads to depigmentation of dorsal iris cells and their redifferentiation into transparent lens cells consisting of specialized keratinocytes (reviewed in ref. 9). In another well-studied system of regeneration, limb regeneration in axolotls, it has long been assumed that the blastema that forms in response to injury contains de- and re-differentiating cells. However, recent work indicates that only dermis cells can 'transdifferentiate' into cartilage and tendons, whereas cartilage, muscle and neuronal precursors within the blastema do not change identity before generating a new limb[11]. Several types of metaplasia have been attributed to transdifferentiation[9], and epithelial mesenchymal transitions may be involved in the formation of metastatic breast cancers[10]. Here, as during normal epithelial mesenchymal transitions, activation of the transcription factors Snail, Slug and Twist are essential[9,10]. With the rapidly growing arsenal of lineage tracing tools it seems likely that many more physiological or pathogenic cell conversions will be discovered in the future.

Perhaps the best evidence that functionally differentiated cells can change fate during normal development comes from studies on the origin of blood. Fetal blood cells originate in the dorsal aorta after activation of Scl (also known as Tal1) and Runx1[12,13], two transcription factors essential for haematopoietic stem cell formation[14]. Most strikingly, time-lapse experiments recently showed that a small proportion of cells within cultured endothelial sheet colonies derived from embryonic stem (ES) cells undo their tight junctions, round up and begin to express erythroid and monocytic haematopoietic antigens[15] (Fig. 2). This process, which is unique to embryonic as

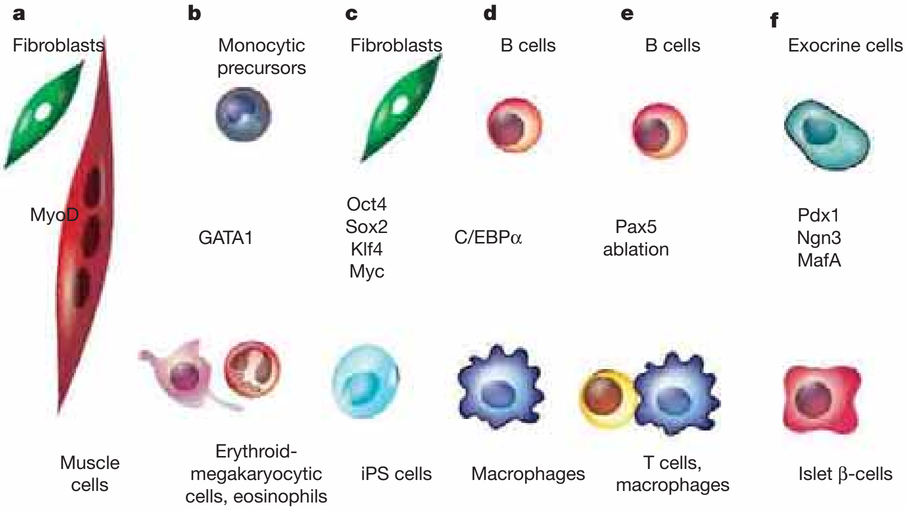


Figure 1 | Examples of transcription factor overexpression or ablation experiments that result in cell fate changes. For explanation of panels **a**–**f** see text.

[1]Center for Genomic Regulation and ICREA, 08003 Barcelona, Spain. [2]MRC Molecular Haematology Unit, Weatherall Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DS, UK.

Lineage Cells of the same developmental origin sharing a similar phenotype/function.
Cell differentiation Process by which cells become more specialized, acquiring new identities.
Cell determination Commitment to a lineage.
Commitment Stable activation of a gene expression program characteristic of a lineage.
Pluripotent Potential of a cell to generate all cell types except extra-embryonic tissue. Examples: embryonic stem cells and induced pluripotent stem (iPS) cells.
Multipotent Potential of a cell to form several lineages within a tissue. Example: haematopoietic stem cells.
Progenitor Cell with the capacity to differentiate and divide but with limited self-renewal potential.
iPS cell reprogramming Induced conversion of somatic cells into pluripotent stem cells.
Lineage reprogramming Conversion of cells from one lineage to another. Term covers both transdifferentiation and transdetermination.
Transdifferentiation Reprogramming of one specialized cell type into another, without reversion to pluripotent cells. Also called 'lineage switching' or 'lineage conversion'.
Transdetermination Reprogramming of a committed, but not yet fully differentiated, cell type into another.
Lineage priming Promiscuous expression in progenitors of transcriptional programs associated with different lineages.
Epigenome/epigenetics Changes in phenotype or gene expression caused by mechanisms other than changes in the underlying DNA.
Cell regeneration Replacement of cells lost by injury or attrition.
Plasticity Ability of a cell to convert into another cell type either spontaneously, by external cues or by gene perturbation experiments.
Metaplasia Replacement in a tissue of one differentiated cell type with another, generally caused by an abnormal chronic stimulus.

opposed to adult endothelium, is exacerbated by shear stress (mimicking blood flow) through production of nitric oxide, upregulation of Runx1, c-Myb and Klf2 (refs 16,17). Interestingly, c-Myb, like Runx1 and Scl, is a transcription factor that is also required for the formation of definitive blood cells[14]. The re-specification of endothelium into haematopoietic cells supports the notion that transdifferentiation may occur during normal development and alternates with classic 'forward' differentiation. This is where the fields of induced lineage conversions and developmental biology merge: we propose that the cell interconversions elicited experimentally by transcription factors may mimic specific physiological cell fate transitions and that the two processes are fundamentally similar.

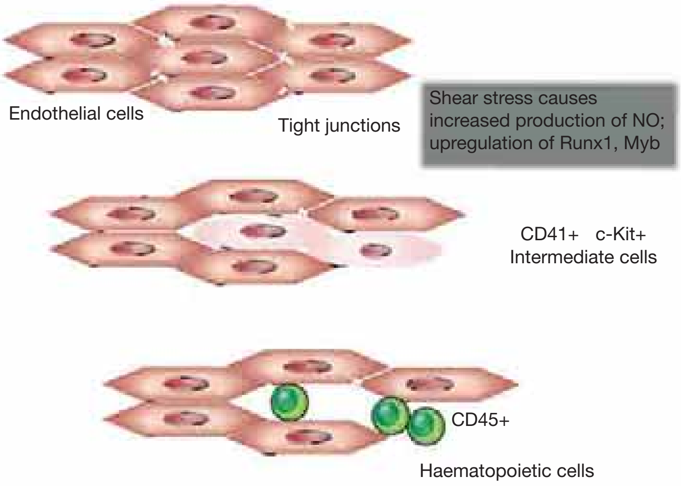


Figure 2 | Conversion of endothelial cells into haematopoietic cells. Schematic of time-lapse microscopy of endothelial colonies derived from ES cells showing that some cells round up and begin to express haematopoietic antigens, such as CD45 (ref. 15). Using the same system, it was shown that mechanical shear stress enhances the formation of blood cells by inducing the formation of CD41, c-Kit-positive cells that produce increased levels of nitric oxide (NO) and upregulate Runx1 and Myb[16]. Similar observations were also made in zebrafish, demonstrating the need for blood flow and NO production for haematopoietic stem cell formation[17].

In this review we briefly chart the evolution of the transcription factor perturbation experiments and discuss how they have provided fundamental insights into the process of lineage specification. They identified lineage-instructive regulators, revealed the principle of transcription factor cross-antagonisms in binary lineage decisions and helped explain the dynamic behaviour of regulatory networks. We also discuss more generally what lineage reprogramming has shown about mechanisms of development (see also ref. 18) and place them in the context of emerging epigenetic landscape models. Finally, we compare transcription-factor-mediated lineage reprogramming to iPS cell reprogramming and discuss their potential for regenerative therapy.

Charting the beginnings

The instructive role of transcription factors in lineage specification was demonstrated in the 1980s, when Harold Weintraub's laboratory discovered that forced expression of MyoD can induce myotube formation in a fibroblast cell line[2] (Fig. 1a). Evidence for the reciprocal regulation of lineage-restricted genes came from the blood system, which, with its diversity of well-defined cell lineages and prospectively isolatable intermediate progenitors, is an ideal venue for lineage-conversion experiments. Thus, when ectopically expressed in cell lines of monocytes (macrophage precursors) at high levels, the erythroid-megakaryocyte-affiliated transcription factor GATA1 not only induced the expression of erythroid-megakaryocyte lineage markers, but also downregulated monocytic markers[3,19]. Lower levels of GATA1 induced the formation of eosinophils, in line with its levels in normal eosinophils[3] (Fig. 1a). The monocytic to erythroid switch could also be effected in the opposite direction: expression of PU.1 (also known as Sfpi1) in an erythroid-megakaryocytic cell line induced its conversion to the monocytic lineage, repressing GATA1 (ref. 20). A potential caveat of these studies was their reliance on cell lines, which may be more inherently plastic than their normal counterparts. This objection was dismissed when ectopic expression of GATA1 produced erythroid-megakaryocytic-eosinophil-basophil output from granulocyte-macrophage progenitors freshly isolated from normal bone marrow[21]. More recently it was shown that even fully differentiated cells can be switched: C/EBPα, a transcription factor required for the formation of granulocyte-macrophage precursors[22] can convert committed B- and T-cell progenitors into functional macrophages at frequencies approaching 100%[23,24]. Mature immunoglobulin-producing B cells could also be switched, although at lower frequencies[23] (Fig. 1d).

Mechanistic implications of the GATA1:PU.1 paradigm

The high efficiency of induced lineage reprogramming in the blood system indicates that ectopically expressed transcription factors interact with endogenous components of the recipient cells' transcriptional network. The switching mechanism may therefore encapsulate the principles of normal lineage specification. Indeed, the dominance of either GATA1 or PU.1 represents one of the earliest and most fundamental decisions during haematopoietic development, serving as a paradigm for cross antagonistic transcription factor interactions[25,26]. In Fig. 3 we have extrapolated the PU.1:GATA1 antagonism to normal lineage specification by assuming that basic gene expression programs of monocytic and erythroid cells are directly controlled by PU.1 and GATA1, respectively. PU.1, and possibly GATA1, also controls its own expression, forming an autoregulatory loop[27,28]. The model is reminiscent of a simpler genetic switch controlling the choice between lysogenic and lytic pathways in phage lambda by the cross-antagonistic and autoregulatory transcriptional regulators Cro and C1[29]. The central role of transcription

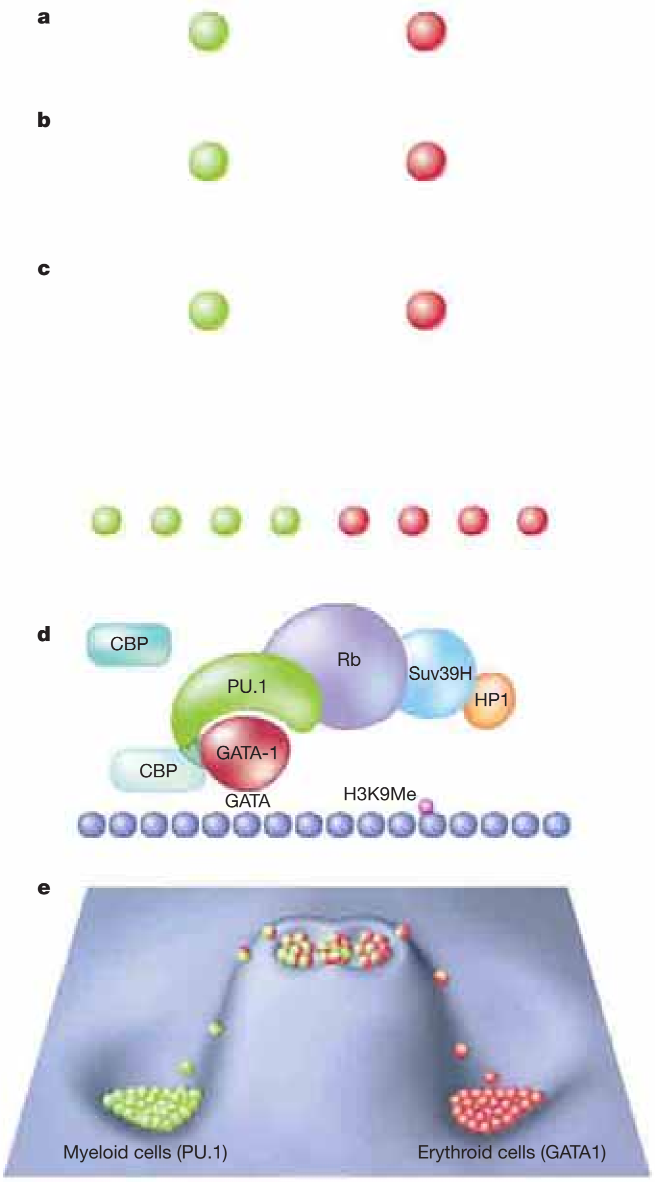


Figure 3 | Transcription factor cross-antagonism: the PU.1:GATA1 paradigm. **a**, In the simplest formulation of cross-antagonism, the two regulators (represented as green and red spheres, respectively) negatively influence each other. **b**, Representation of a cross-antagonistic motif in which the transcription factors also autoregulate. **c**, Here the two factors are shown to positively or negatively regulate the repertoire of their own and each other's target genes. **d**, Scheme of the biochemical mechanisms that underlie the GATA1 arm of the PU.1:GATA1 antagonism. To activate a target gene in erythroid cells GATA1 recruits the histone acetylase CREB-binding protein. Overexpressed PU.1 displaces CREB-binding protein (CBP) by binding to GATA1 and recruits Rb as well as Suv39H protein. This results in methylation of lysine 9 in histone H3 and recruitment of HP1a, causing repression of the target gene[34]. **e**, Representation of the PU.1:GATA1 antagonism as a binary attractor model in a modified Waddingtonian epigenetic landscape. Bicoloured marbles in the upper, shallow basin represent monocytic/erythroid progenitors that express different ratios of PU.1 and GATA1. These progenitors fluctuate between different states determined by the relative amount of PU.1 and GATA1. Cells at both ends of the spectrum are biased towards either monocytic or erythroid differentiation. During spontaneous or induced commitment they move out of the basin and roll into the attractor basins below. Green marbles represent monocytic cells expressing high levels of PU.1; red marbles erythroid cells expressing high levels of GATA1.

factor cross antagonisms in binary cell fate choices[30–32] is the single most important concept that has emerged from lineage reprogramming experiments.

Studies of GATA1-mediated myelomonocytic cell fate conversions showed that it directly binds to PU.1 protein (reviewed in ref. 31). High levels of GATA1 inhibit PU.1 by displacing c-Jun, a cofactor of PU.1, thus leading to the collapse of the monocytic program[33]. Conversely, PU.1 expressed in erythroid precursors interacts with GATA1 bound to promoters of target genes, including α- and β-globin as well as *EKLF* (also known as *KLF1*), and converts an activating into a repressive complex through displacement of the coactivator CREB-binding protein (CBP) and recruitment of the retinoblastoma protein[34] (Fig. 3d). Therefore, lineage-instructive transcription factors not only 'step on the accelerator' to induce a new gene expression program, but also 'put on the brakes' to inactivate key regulators of alternative cell types, leading to extinction of markers characteristic of the old phenotype. Once one of the two factors has become dominant the conflict is resolved and commitment ensues.

Transcription factor ablation and lineage re-specification

If forced resolutions of transcription factor cross antagonisms specify lineages it should also be possible to trigger differentiation by loss of transcription factor function. Evidence gathered in the haematopoietic system supports this prediction. The earliest haematopoietic cells in zebrafish arise anteriorly as macrophage precursors and posteriorly as erythroid precursors. Morpholino-mediated knockdown of PU.1 leads to the ectopic formation of haemoglobin-producing cells in the dorsal region[35] whereas inactivation of GATA1 induces the formation of monocytic cells in the posterior region[36]. This indicates that committed erythroid and monocytic progenitors can be re-specified when the opposing key regulator is ablated. However, recent work indicating pluripotency of the posterior population indicates a more complex PU.1/GATA1 balance in this region[37]. Inactivation of key regulators may also lead to the reactivation of earlier genetic programs in committed cells, resulting in their dedifferentiation and activation of multilineage potential. For example, ablation of Pax5 in B-cell precursors activates expression of genes from alternative haematopoietic lineages. Under appropriate culture conditions or after transplantation these cells can differentiate into granulocyte/macrophage, T-cell, dendritic, natural killer and osteoclast lineages[38]. Alternative lineage potentials can even be resuscitated in fully functional B cells: transplantation of Bcl2-stabilized Pax5-deficient cells into immunodeficient mice generates T cells, which contain immunoglobulin rearrangements[39] (Fig. 1e). This conversion does not appear to be direct as it entails the dedifferentiation to a lymphoid precursor.

Influence of cell-extrinsic signals

So far cross-antagonistic switches were presented as relatively simple circuits functioning in a broadly cell-intrinsic manner. In reality, most antagonistic circuits in metazoa are subject to graded external inputs[40,41]. An example that illustrates the interplay between cell intrinsic and extrinsic signals is relevant for the branching of CD4$^+$ T lineages into T_H17 and T_{reg} type helper cells[42]. Differentiation of T_H17 cells requires RORγt whereas T_{reg} cells require Foxp3, transcription factors that are coexpressed in naive CD4 cells. The differentiation of these two cell types is orchestrated by a transforming growth factor (TGF)-β gradient. Low TGF-β concentrations plus interleukin (IL)-6 and IL-21 upregulate RORγt and promote the formation of T_H17 cells. In contrast, high TGF-β concentrations upregulate Foxp3 and facilitate T_{reg} cell formation. In addition, Foxp3 inhibits RORγt function, probably through direct protein interaction[42]. Recent experiments have shed light on the long-standing debate as to whether or not haematopoietic cytokines have a lineage-instructive function or merely promote survival and proliferation of already committed cells. These experiments, conducted with isolated bipotent progenitors and followed by time-lapse microscopy, indicate that myeloid cytokines are capable of instructing lineage choice[43]. Similarly, experiments with mice deficient for the transcription factor MafB point to an instructive role for the macrophage colony stimulating factor (M-CSF) during myeloid commitment of haematopoietic stem cells; in the absence of

MafB haematopoietic stem cells become hyper-responsive to M-CSF through activation of the M-CSF receptor regulator PU.1, resulting in an enhanced myelomonocytic output after transplantation[44].

Interactions between external inputs and cell fate decisions are not restricted to blood cells. A classical example is the interplay between an activin gradient and the transcription factors brachyury, goosecoid and Mix during patterning of mesoderm. Brachyury, which autoregulates its own production, is activated by low and high levels of activin (nodal) signalling, leading to different developmental outcomes. Low levels lead to the activation of brachyury and repression of goosecoid through inactivation of Mix, resulting in the production of posterior mesoderm. High activin levels induce Mix expression, which in turn represses brachyury by activating goosecoid, resulting in endoderm and anterior mesoderm. As predicted, loss of goosecoid results in the production of posterior mesoderm at the expense of anterior mesoderm and endoderm[45].

Transcription factor network assembly and lineage outcome

Lineage switching experiments in the haematopoietic system have shown that the order in which two transcription factors become expressed in a progenitor can decide lineage outcome. Using prospectively isolated common lymphoid progenitors, sustained expression of C/EBPα generates granulocyte-macrophages whereas sustained expression of GATA2 generates mast cells. However, two entirely new cell types, eosinophils and basophils, are generated when CEBPα and GATA2 are sequentially expressed and in a different order[46] (Fig. 4). This shows that the same transcription factor pair can specify alternative cell types, probably because the separate expression of C/EBPα and GATA2 generates two distinct intermediate progenitors whose fates are further redirected by the incoming factor. It is possible that C/EBPα and GATA2 interact with different co-regulators in different cell types. Such a sequential participation of transcription factors in different protein complexes during differentiation has been likened to the changing interactions of guests at a cocktail party[47]. These observations underscore the importance of timing and cell context for the assembly of cell type specific transcription factor networks.

Gene regulatory networks and cell fate attractors

A popular framework for conceptualizing the specification of different cell types is that of the epigenetic landscapes proposed by Waddington[48]. Extrapolating from Waddington, different cell types may be seen as stable solutions of transcription factor networks—or 'attractors'—which occupy the basins of Waddington's landscape[49–51]. Within this framework developmental intermediates, such as multipotent progenitors, may be viewed as representing metastable states that are characterized by co-expression of cross-antagonistic regulatory factors driving alternative lineage-affiliated programs of gene expression. This arrangement affords structuring of lineage choice and ensures robustness of the differentiated state. Robustness may also underlie why direct reprogramming by an ectopic transcription factor works so well: it only requires destabilization of one stable network solution and the realization of another stable solution, a transition that can probably be achieved through multiple paths.

These views square well with experiments indicating that multipotent cells prime competing lineage-affiliated gene expression programs before commitment—a phenomenon dubbed 'lineage priming'[52–54]. Mostly on the basis of work with a bipotent haematopoietic cell line it has further been proposed that all cells within the multipotential compartment are not equivalently primed and may fluctuate between different lineage-biased states[55,56] (Fig. 3e). Whether these fluctuations, which have also been observed for Nanog expression within self-renewing ES cells[57], are driven by 'noise' or by other cell-intrinsic mechanisms is not known.

In complex differentiation hierarchies, like those of the blood system, one can envisage cell states cascading down the valleys of a mountain range, with each bifurcating decision heralded by activation of new cross-antagonistic pairings, which themselves result from the outcome of a prior 'bout' (Fig. 5). Although the combination of cross-antagonistic and autoregulatory circuits can in principle convert small initial asymmetries within cells into stable or metastable network states representing distinct cell types[58–60], the particular cross-antagonistic circuit used to select choice may still be amenable to resetting in the other direction. As cells cascade through bifurcating decisions, prior switches become less available, decreasing the probability of reversal and further restricting possible cell-type solutions going forward[58]. This 'passing of the baton' of networks from one cell-state to the next ensures forward momentum during lineage specification in development and explains the temporal and cellular profiles of transcription factor activity[61]. Some of the aforementioned antagonisms best exemplify such a mechanism. At the level of the common myeloid progenitor (CMP), resolution of the PU.1:GATA1 antagonism leads to the bifurcation into bipotential granulocyte/macrophage progenitors (GMP) and megakaryocyte/erythroid progenitors (MEP). In turn, at the level of these bipotent precursors resolution of the Gfi1:Nab2/Egr antagonism in granulocyte/macrophage progenitors and the EKLF:FLI1 antagonism in megakaryocyte/erythroid progenitors creates four distinct cell types: granulocytes, macrophages, erythrocytes and megakaryocytes[62,63]. Another example is the sequential cross-antagonisms during T-cell development, first involving GATA3 and T-bet at the level of T helper type 1/T helper type 2 (Th1/Th2) precursors and then RORγt and Foxp3 at the level of naive CD4 T helper cells[42,64]. This 'branching compartmentalization' allows decisions to be inherited from one progenitor to the next as well as effecting a separation of states, affording the re-use of a transcription factor in a different network context. Such a modus operandi accommodates the observations that the order of expression of transcription factors may affect cell fate outcomes and that the same final cell states can be reached via alternative routes, as exemplified by the different origins of granulocyte-macrophage precursors (Fig. 5).

Within the landscape/attractor-based conceptual framework, it is easy to see how cell fate transitions, such as those exhibited by lineage reprogramming events, may be favoured between cell types closely connected through shared regulatory switches. It also predicts that the efficiency by which transcription factors induce lineage conversions depends on the proximity of the cell type in question and that bridging greater distances may require additional factors acting at earlier common branch points. A possible example is the observation

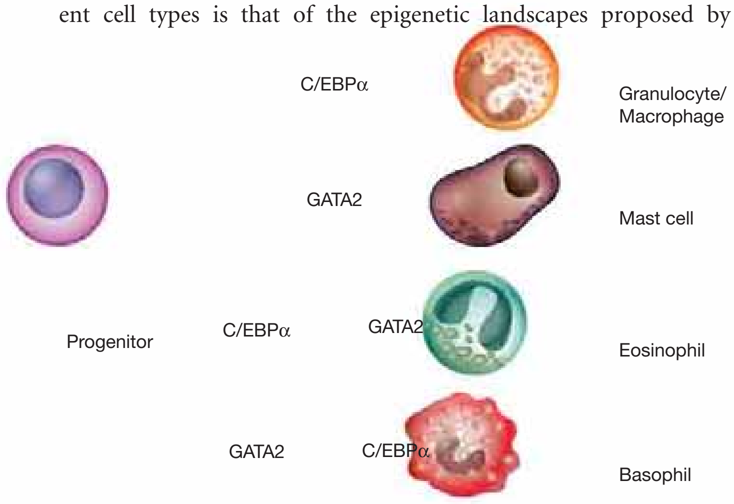


Figure 4 | Timing of transcription factor expression and lineage outcome. Forced expression of C/EBPα in common lymphoid progenitors induces the formation of granulocytes and macrophages, whereas GATA2 induces the formation of mast cells. If C/EBPα expression is followed by GATA2 the cells turn into eosinophils. If the order of expression is reversed they become basophils (after ref. 46). Similar rules apply to the physiological specification of the relevant cell types from the multipotent myeloid progenitor[26].

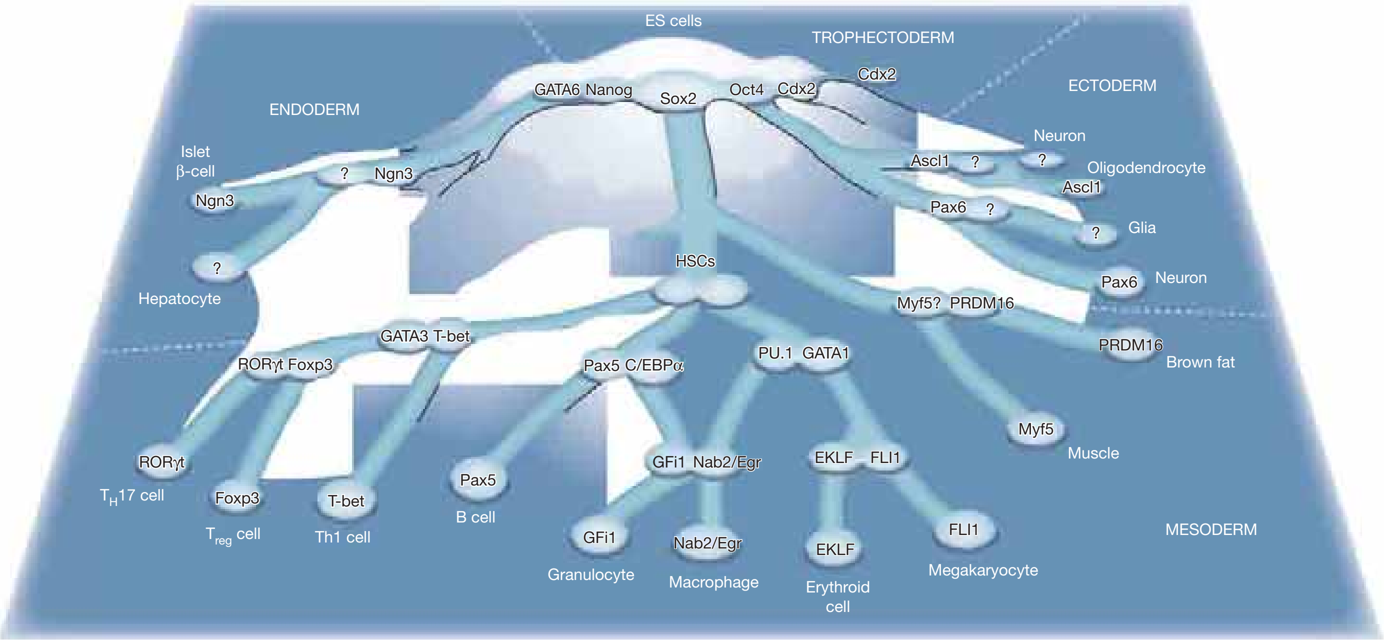


Figure 5 | Transcription factor cross-antagonisms in a cascading landscape of unstable and stable cell states. The territory, represented as a mountain range, depicts all possible solutions of a single regulatory network that specifies cell identity. Robust network states correspond to stably differentiated cell types (deep basins in the low-lying plains) whereas unstable solutions correspond to ridges and slopes in the landscape. The latter are only fleetingly occupied during development and thus unlikely to correspond to observable cell types. The route between pluripotent and fully differentiated network states is punctuated by a series of metastable states corresponding to progenitors characterized by the cross-antagonistic interaction of competing lineage-affiliated transcription factors. Within these goggle-shaped 'binary attractors' transcriptional networks fluctuate between lineage-biased states before exit either into a stable attractor corresponding to a developmental endpoint, or into a subsequent metastable attractor where a secondary lineage decision is taken. In this model the sequential establishment and resolution of transcription factor cross-antagonisms is a driving force in lineage specification. However, such a mechanism might not apply to earlier intermediates, which may only be partly restricted and do not necessarily commit through simple binary decisions. The intermediates and paths depicted may not be exclusive or obligatory transit points, but rather represent the most favoured possibilities. Although all the transcription factors shown have been experimentally demonstrated to possess lineage-instructive capacity, their precise mechanism of action or the identity of a presumed antagonistic partner (indicated with a question mark) is not known. ES cells, embryonic stem cells; HSCs, haematopoietic stem cells.

that switching into β-cell islets of hepatic progenitors only requires Ngn3 (ref. 65), whereas switching of exocrine pancreas cells requires in addition Pdx1 and MafA[66] (Fig. 1e). Transcription-factor-mediated lineage conversions may thus be achieved by effecting the same regulatory interactions that drive normal differentiation. However, the actual path taken by the cells is not clear. Here, two possibilities can be considered: (1) the ectopic transcription factor first resets the cell's regulatory network to an earlier branch point position and then directs it back along a physiological trajectory to the new cell type; (2) alternatively, reprogramming results in direct crossing of the 'ridge' that divides the two lineage-committed territories without reactivating progenitor programs.

Generalizing transcription factor cross-antagonisms

Many binary junctures during development seem to be governed by cross-antagonistic transcription factor interactions. As summarized in the epigenetic landscape in Fig. 5, the haematopoietic system offers the largest number of well-studied pairs. This may simply reflect the wealth of knowledge of developmental intermediates in which cross-antagonistic interactions may be studied. Alternatively, perhaps because of their largely free-floating nature, haematopoietic cells might be more weighted towards cell-autonomous decisions. In addition to the examples already discussed, the decision of erythroid against megakaryocytic cells is effected by the balance of EKLF:FLI1 (refs 63,67); granulocytes against macrophages by Gfi1:Nab2/Egr[62]; erythroid-megakaryocyte precursors against eosinophils by C/EBPb:FOG-1 (ref. 68); Th1 against Th2 cells by T-bet:GATA3 (ref. 64); and T_H17 against T_{reg} cells by RORγt:Foxp3 (ref. 42). A potential cross-antagonism outside the haematopoietic system is played out within skeletal muscle and brown fat precursors. Here, enforced expression of PRDM16 in Myf5-expressing mesenchymal progenitors induces their differentiation into brown fat cells. Conversely, inactivation of PRDM16 in these cells promotes muscle differentiation and causes a loss of brown fat characteristics[69]. This suggests that PRDM16 controls a bidirectional fate switch between skeletal myoblasts and brown fat cells. Other examples include the conversion of astrocytes into neurons[70]; neural precursors into oligodendrocytes by Ascl1 (ref. 71); neural precursors into inner ear sensory cells by Atoh1 (ref. 72); liver cells into islet β-cells by Pdx1-VP16 (ref. 73); and hepatocyte precursors into insulin producing islet-like β-cells by Ngn3 (ref. 65). Of note, for none of these factors has antagonistic partners been described, raising the possibility that in the absence of a lineage-instructive transcription factor the relevant precursors enter a default pathway. The very first developmental decisions in the pre-implantation embryo also seem to be guided by transcription factor cross-antagonisms. Here, the best-studied example is the pair Cdx2:Oct4, where forced expression of Cdx2 in ES cells induces the formation of trophectoderm cells by inhibiting Oct4 through direct protein interaction[74]. Finally, the predominance of either Nanog or GATA6 decides whether ES cells maintain their identity or differentiate into endoderm[75].

The new world order of iPS cells

So, how then does this framework of induced lineage reprogramming, normal forward differentiation and developmental transdifferentiation relate to the induced conversion of somatic into embryonic stem cells—the new world order of iPS cells? Before tackling this question, let us first consider the salient features of the iPS situation. Initial experiments demonstrated that the combination of Oct4, Sox2, Klf4 and Myc can induce the transition from fibroblasts into stable self-renewing cells closely resembling ES cells[8] (Fig. 1c). iPS reprogramming could subsequently also be achieved with a range of somatic cell types, including differentiated cells such as hepatic[76] or islet β-cells[77]. All of these cells express their own cell-type-specific repertoires of lineage-instructive transcription factors. How does reprogramming proceed in these cases? It seems unlikely that iPS reprogramming factors have evolved to interact with the large variety of lineage-affiliated transcription factors and thus divert regulatory networks within cells that are many branch points away from the pluripotent state. Instead, the low frequency and long duration of iPS reprogramming, in excess of a week (reviewed in refs 78, 79), suggests that stochastic mechanisms are involved and that several rounds of cell divisions are required.

As for directly induced lineage conversions, it would be predicted that reprogramming of cells that are developmentally closely related require fewer transcription factors. Indeed, neural progenitors, which already express Sox2, Klf4 and Myc, can be turned into iPS cells with only Oct4 (ref. 80). However, their reprogramming efficiency remains exceedingly low, suggesting that even here stochastic processes are at play. How reprogramming works remains unclear[79]. Oct4 and its partners might gradually gain access to hidden DNA binding sites through the dynamic 'breathing' of chromatin, eventually upregulating the corresponding endogenous factors and thus establishing transgene independence by activating autoregulatory loops. Another mechanism is the direct interaction with chromatin-remodelling proteins, leading to upregulation of critical ES cell regulators such as Nanog[81]. Repression of the resident cells' program in turn might be mediated by the capacity of ES cell regulators to actively silence differentiation-affiliated transcription factors, such as through recruitment of polycomb complexes and formation of bivalent chromatin domains[82].

No matter what the relevant mechanisms are, ES cells, and by inference iPS cells, are unique in that they represent a cellular ground state whose default configuration is that of self-renewal[83]. This ground state is particularly accessible to the transcription factor program that establishes the highly stable Oct4–Nanog–Sox2 network, where the three factors regulate each other's expression as well as their own, in an arrangement known as a fully connected triad[84–86]. The high proliferative potential of ES/iPS cells and the stability of the triad may explain why the relatively rare iPS reprogramming events can so easily be trapped in culture. Because of the heterogeneity of ES cells[57] the ES cell state can be seen as a broad attractor in which many pluripotent network configurations may co-exist and interconvert along the lines discussed for the blood system[56,87]. This level of tolerance in possible network configurations may additionally increase ease of access from many if not all somatic cell states.

Induced lineage reprogramming and regenerative medicine

The question then arises whether, given enough knowledge, it will be possible to directly reprogram any cell type into another and to custom-design cells for regenerative therapy from easily obtainable cell sources. Consider B cells for example: we know that transcription factor gain or loss of function can convert these cells into macrophages as well as into T cells and ES cells. But can B cells be induced directly to become, say, haematopoietic stem cells or islet β-cells at high frequencies, effecting 'long jumps' within the landscape of Fig. 5? Attempts to induce direct transitions between distantly related somatic cell types have been inconclusive so far. For example, forced expression of MyoD in a keratinocyte cell line did not induce myotube formation and only upregulated a few mesenchymal genes[88]. Co-expression of PU.1 and C/EBPα in fibroblasts converted them into macrophage-like cells, but the resulting cells were only partially functional[89]. Most encouraging are results describing the induction of the rapid and extensive reciprocal regulation of keratinocyte- and muscle-associated genes in heterokaryons between human keratinocytes and mouse muscle cells. Here, phenotypic dominance could be achieved by increasing the ratio of one cell type over the other[90]. It will now be interesting to see whether efficient long jumps can be achieved with defined genes into cells closely resembling their normal counterparts. However, for regenerative purposes a full equivalence to normal cells may not be necessary as long as the induced cells perform the desired functions *in vivo* and long term.

In conclusion, it may eventually be possible to generate cells 'a la carte' by forced transcription factor expression in cultured biopsies. However, because most progenitors and differentiated cells do not proliferate, the cells generated probably cannot be expanded, as is possible with iPS cells. For cell replacement therapy purposes it is therefore crucial that high frequency transitions can be achieved. This might necessitate, in addition to simple overexpression of transcription factor(s), a whole arsenal of tricks, including the inducible, sequential and graded expression of transcription factors[46,91], transcription factor knockdowns[38,92], modulation of microRNAs and chromatin remodelling factors[93–95], or treatment of the cells with chemicals[96]. If successful, such experiments, aside from their clinical potential, would provide valuable information about the regulatory networks that specify different cell types. A promising alternative for the directed induction of desired cell types are *in vivo* approaches. Perhaps the most spectacular example to date is the conversion of exocrine pancreas cells into fully functional islet β-cells in mice by Pdx1, Ngn3 and MafA[65,66]. In spite of these successes, custom-designing cells for cell therapy in humans is still a long way off. Only time will tell whether it will prevail over the application of cells derived from iPS or ES cell lines. But it seems safe to predict that transcription-factor-induced cell reprogramming will continue to reveal hidden secrets of cell differentiation for a long time to come.

Note added in proof: Two new examples of transcription factor cross antagonisms outside the blood cell system have recently been described. One describes the antagonism between the Ngn3 target Arx and Pax4 in pancreas development[97]. The other is an antagonism playing out during vascular and muscle specification in the dermomyotome[98].

1. Blau, H. M. How fixed is the differentiated state? Lessons from heterokaryons. *Trends Genet.* **5**, 268–272 (1989).
2. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).
3. Kulessa, H., Frampton, J. & Graf, T. GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. *Genes Dev.* **9**, 1250–1262 (1995).
   This paper, together with refs 19 and 20, established the principle of transcription factor cross-antagonisms.
4. Gurdon, J. B. & Byrne, J. A. The first half-century of nuclear transplantation. *Proc. Natl Acad. Sci. USA* **100**, 8048–8052 (2003).
5. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813 (1997).
6. Gurdon, J. B. & Melton, D. A. Nuclear reprogramming in cells. *Science* **322**, 1811–1815 (2008).
7. Hochedlinger, K. & Jaenisch, R. Monoclonal mice generated by nuclear transfer from mature B and T donor cells. *Nature* **415**, 1035–1038 (2002).
8. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
9. Slack, J. M. Metaplasia and transdifferentiation: from pure biology to the clinic. *Nature Rev. Mol. Cell Biol.* **8**, 369–378 (2007).
10. Yang, J. & Weinberg, R. A. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev. Cell* **14**, 818–829 (2008).
11. Kragl, M. *et al.* Cells keep a memory of their tissue origin during axolotl limb regeneration. *Nature* **460**, 60–65 (2009).
12. Chen, M. J., Yokomizo, T., Zeigler, B. M., Dzierzak, E. & Speck, N. A. Runx1 is required for the endothelial to haematopoietic cell transition but not thereafter. *Nature* **457**, 887–891 (2009).
13. Lancrin, C. *et al.* The haemangioblast generates haematopoietic cells through a haemogenic endothelium stage. *Nature* **457**, 892–895 (2009).
14. Dzierzak, E. & Speck, N. A. Of lineage and legacy: the development of mammalian hematopoietic stem cells. *Nature Immunol.* **9**, 129–136 (2008).

15. Eilken, H. M., Nishikawa, S. & Schroeder, T. Continuous single-cell imaging of blood generation from haemogenic endothelium. *Nature* **457,** 896–900 (2009).
An example of 'transdifferentiation' in the context of normal lineage progression; also highlights how real-time visualization may show cell fate conversions that are otherwise hard to document.
16. Adamo, L. *et al.* Biomechanical forces promote embryonic haematopoiesis. *Nature* **459,** 1131–1135 (2009).
17. North, T. E. *et al.* Hematopoietic stem cell development is dependent on blood flow. *Cell* **137,** 736–748 (2009).
18. Zhou, Q. & Melton, D. A. Extreme makeover: converting one cell into another. *Cell Stem Cell* **3,** 382–388 (2008).
19. Visvader, J. E., Elefanty, A. G., Strasser, A. & Adams, J. M. GATA-1 but not SCL induces megakaryocytic differentiation in an early myeloid line. *EMBO J.* **11,** 4557–4564 (1992).
20. Nerlov, C. & Graf, T. PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. *Genes Dev.* **12,** 2403–2412 (1998).
21. Heyworth, C., Pearson, S., May, G. & Enver, T. Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *EMBO J.* **21,** 3770–3781 (2002).
22. Zhang, P. *et al.* Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBPα. *Immunity* **21,** 853–863 (2004).
23. Xie, H., Ye, M., Feng, R. & Graf, T. Stepwise reprogramming of B cells into macrophages. *Cell* **117,** 663–676 (2004).
24. Laiosa, C. V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L. & Graf, T. Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBPα and PU.1 transcription factors. *Immunity* **25,** 731–744 (2006).
25. Arinobu, Y. *et al.* Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. *Cell Stem Cell* **1,** 416–427 (2007).
26. Iwasaki, H. & Akashi, K. Myeloid lineage commitment from the hematopoietic stem cell. *Immunity* **26,** 726–740 (2007).
27. Okuno, Y. *et al.* Potential autoregulation of transcription factor PU.1 by an upstream regulatory element. *Mol. Cell. Biol.* **25,** 2832–2845 (2005).
28. Yu, C. *et al.* Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage *in vivo*. *J. Exp. Med.* **195,** 1387–1395 (2002).
29. Ptashne, M. *A Genetic Switch. Phage Lambda Revisited* 3rd edn (Cold Spring Harbor Laboratory Press, 2004).
30. Cantor, A. B. & Orkin, S. H. Hematopoietic development: a balancing act. *Curr. Opin. Genet. Dev.* **11,** 513–519 (2001).
31. Graf, T. Differentiation plasticity of hematopoietic cells. *Blood* **99,** 3089–3101 (2002).
32. Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132,** 631–644 (2008).
33. Zhang, P. *et al.* Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proc. Natl Acad. Sci. USA* **96,** 8705–8710 (1999).
34. Stopka, T., Amanatullah, D. F., Papetti, M. & Skoultchi, A. I. PU.1 inhibits the erythroid program by binding to GATA-1 on DNA and creating a repressive chromatin structure. *EMBO J.* **24,** 3712–3723 (2005).
35. Rhodes, J. *et al.* Interplay of Pu.1 and Gata1 determines myelo-erythroid progenitor cell fate in zebrafish. *Dev. Cell* **8,** 97–108 (2005).
***In vivo* evidence for the importance of GATA1:PU.1 interplay in lineage specification.**
36. Galloway, J. L., Wingert, R. A., Thisse, C., Thisse, B. & Zon, L. I. Loss of Gata1 but not Gata2 converts erythropoiesis to myelopoiesis in zebrafish embryos. *Dev. Cell* **8,** 109–116 (2005).
37. Warga, R. M., Kane, D. A. & Ho, R. K. Fate mapping embryonic blood in zebrafish: multi- and unipotential lineages are segregated at gastrulation. *Dev. Cell* **16,** 744–755 (2009).
38. Nutt, S. L., Heavey, B., Rolink, A. G. & Busslinger, M. Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. *Nature* **401,** 556–562 (1999).
39. Cobaleda, C., Jochum, W. & Busslinger, M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* **449,** 473–477 (2007).
40. Rothenberg, E. V. Cell lineage regulators in B and T cell development. *Nature Immunol.* **8,** 441–444 (2007).
41. Davidson, E. H. & Levine, M. S. Properties of developmental gene regulatory networks. *Proc. Natl Acad. Sci. USA* **105,** 20063–20066 (2008).
42. Zhou, L. *et al.* TGF-β-induced Foxp3 inhibits T_H17 cell differentiation by antagonizing RORγt function. *Nature* **453,** 236–240 (2008).
43. Rieger, M. A., Hoppe, P. S., Smejkal, B. M., Eitelhuber, A. C. & Schroeder, T. Hematopoietic cytokines can instruct lineage choice. *Science* **325,** 217–218 (2009).
44. Sarrazin, S. *et al.* MafB restricts M-CSF-dependent myeloid commitment divisions of hematopoietic stem cells. *Cell* **138,** 300–313 (2009).
An example of how extrinsic signals may act through intrinsic regulators to specify lineage fates; ref. 57 addresses a similar issue from a mathematical modelling perspective.
45. Smith, J., Wardle, F., Loose, M., Stanley, E. & Patient, R. Germ layer induction in ESC–following the vertebrate roadmap. *Curr. Protocols Stem Cell Biol.* **1,** 1D.1.1–1D.1.22 (2007).
46. Iwasaki, H. *et al.* The order of expression of transcription factors directs hierarchical specification of hematopoietic lineages. *Genes Dev.* **20,** 3010–3021 (2006).
Showed that the order of transcription factor expression can induce different cell fates.
47. Sieweke, M. H. & Graf, T. A transcription factor party during blood cell differentiation. *Curr. Opin. Genet. Dev.* **8,** 545–551 (1998).
48. Waddington, C. H. *The Strategy of the Genes* (Allen & Unwin, 1957).
49. Kauffman, S. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* **22,** 437–467 (1969).
50. Kauffman, S. *Origins of Order: Self-organization and Selection in Evolution* (Oxford Univ. Press, 1993).
51. Enver, T., Pera, M., Peterson, C. & Andrews, P. W. Stem cell states, fates, and the rules of attraction. *Cell Stem Cell* **4,** 387–397 (2009).
52. Hu, M. *et al.* Multilineage gene expression precedes commitment in the hemopoietic system. *Genes Dev.* **11,** 774–785 (1997).
53. Miyamoto, T. *et al.* Myeloid or lymphoid promiscuity as a critical step in hematopoietic lineage commitment. *Dev. Cell* **3,** 137–147 (2002).
54. Månsson, R. *et al.* Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity* **26,** 407–419 (2007).
55. Enver, T., Heyworth, C. M. & Dexter, T. M. Do stem cells play dice? *Blood* **92,** 348–351,–352 (1998).
56. Graf, T. & Stadtfeld, M. Heterogeneity of embryonic and adult stem cells. *Cell Stem Cell* **3,** 480–483 (2008).
57. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–1234 (2007).
58. Chickarmane, V., Enver, T. & Peterson, C. Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility. *PLoS Comput. Biol.* **5,** e1000268 (2009).
59. Huang, S., Guo, Y. P., May, G. & Enver, T. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Dev. Biol.* **305,** 695–713 (2007).
Refs 57, 58 and 59 highlight how mathematical modelling of cross-antagonistic circuits illuminates their dynamic behaviour and capacity to effect stable lineage choice decisions.
60. Roeder, I. & Glauche, I. Towards an understanding of lineage specification in hematopoietic stem cells: a mathematical model for the interaction of transcription factors GATA-1 and PU.1. *J. Theor. Biol.* **241,** 852–865 (2006).
61. Swiers, G., Patient, R. & Loose, M. Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. *Dev. Biol.* **294,** 525–540 (2006).
62. Laslo, P. *et al.* Multilineage transcriptional priming and determination of alternate hematopoietic cell fates. *Cell* **126,** 755–766 (2006).
An example of sequential cross-antagonistic switches in the specification of cell lineage.
63. Frontelo, P. *et al.* Novel role for EKLF in megakaryocyte lineage commitment. *Blood* **110,** 3871–3880 (2007).
64. Hwang, E. S., Szabo, S. J., Schwartzberg, P. L. & Glimcher, L. H. T helper cell fate specified by kinase-mediated interaction of T-bet with GATA-3. *Science* **307,** 430–433 (2005).
65. Yechoor, V. *et al.* Neurogenin3 is sufficient for transdetermination of hepatic progenitor cells into neo-islets *in vivo* but not transdifferentiation of hepatocytes. *Dev. Cell* **16,** 358–373 (2009).
66. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D. A. *In vivo* reprogramming of adult pancreatic exocrine cells to β-cells. *Nature* **455,** 627–632 (2008).
Showed that expression in the pancreas of a combination of three key regulators re-specifies one somatic cell type into another functional cell type, *in vivo*.
67. Starck, J. *et al.* Functional cross-antagonism between transcription factors FLI-1 and EKLF. *Mol. Cell. Biol.* **23,** 1390–1402 (2003).
68. Querfurth, E. *et al.* Antagonism between C/EBPβ and FOG in eosinophil lineage commitment of multipotent hematopoietic progenitors. *Genes Dev.* **14,** 2515–2525 (2000).
69. Kajimura, S. *et al.* Regulation of the brown and white fat gene programs through a PRDM16/CtBP transcriptional complex. *Genes Dev.* **22,** 1397–1409 (2008).
70. Heins, N. *et al.* Glial cells generate neurons: the role of the transcription factor Pax6. *Nature Neurosci.* **5,** 308–315 (2002).
71. Jessberger, S., Toni, N., Clemenson, G. D. Jr, Ray, J. & Gage, F. H. Directed differentiation of hippocampal stem/progenitor cells in the adult brain. *Nature Neurosci.* **11,** 888–893 (2008).
72. Gubbels, S. P., Woessner, D. W., Mitchell, J. C., Ricci, A. J. & Brigande, J. V. Functional auditory hair cells produced in the mammalian cochlea by *in utero* gene transfer. *Nature* **455,** 537–541 (2008).
73. Horb, M. E., Shen, C. N., Tosh, D. & Slack, J. M. Experimental conversion of liver to pancreas. *Curr. Biol.* **13,** 105–115 (2003).
74. Niwa, H. *et al.* Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123,** 917–929 (2005).
75. Ralston, A. & Rossant, J. Genetic regulation of stem cell origins in the mouse embryo. *Clin. Genet.* **68,** 106–112 (2005).
76. Aoi, T. *et al.* Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* **321,** 699–702 (2008).
77. Stadtfeld, M., Brennand, K. & Hochedlinger, K. Reprogramming of pancreatic β cells into induced pluripotent stem cells. *Curr. Biol.* **18,** 890–894 (2008).

78. Hochedlinger, K. & Plath, K. Epigenetic reprogramming and induced pluripotency. *Development* **136,** 509–523 (2009).
79. Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460,** 49–52 (2009).
80. Kim, J. B. *et al.* Oct4-induced pluripotency in adult neural stem cells. *Cell* **136,** 411–419 (2009).
81. Loh, Y. H., Zhang, W., Chen, X., George, J. & Ng, H. H. Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.* **21,** 2545–2557 (2007).
82. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
83. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519–523 (2008).
84. Alon, U. *An Introduction to Systems Biology. Design Principles of Biological Circuits* (Chapman and Hall/CRC, 2006).
85. Chickarmane, V., Troein, C., Nuber, U. A., Sauro, H. M. & Peterson, C. Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput. Biol.* **2,** e123 (2006).
86. Chickarmane, V. & Peterson, C. A computational model for understanding stem cell, trophectoderm and endoderm lineage determination. *PLoS One* **3,** e3478 (2008).
87. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453,** 544–547 (2008).
88. Boukamp, P., Chen, J., Gonzales, F., Jones, P. A. & Fusenig, N. E. Progressive stages of ''transdifferentiation'' from epidermal to mesenchymal phenotype induced by MyoD1 transfection, 5-aza-2′-deoxycytidine treatment, and selection for reduced cell attachment in the human keratinocyte line HaCaT. *J. Cell Biol.* **116,** 1257–1271 (1992).
89. Feng, R. *et al.* PU.1 and C/EBPα/β convert fibroblasts into macrophage-like cells. *Proc. Natl Acad. Sci. USA* **105,** 6057–6062 (2008).
90. Palermo, A. *et al.* Nuclear reprogramming in heterokaryons is rapid, extensive, and bidirectional. *FASEB J.* **23,** 1431–1440 (2009).
91. Singh, H., Medina, K. L. & Pongubala, J. M. Contingent gene regulatory networks and B cell fate specification. *Proc. Natl Acad. Sci. USA* **102,** 4949–4953 (2005).
92. Kitajima, K., Zheng, J., Yen, H., Sugiyama, D. & Nakano, T. Multipotential differentiation ability of GATA-1-null erythroid-committed cells. *Genes Dev.* **20,** 654–659 (2006).
93. Judson, R. L., Babiarz, J. E., Venere, M. & Blelloch, R. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nature Biotechnol.* **27,** 459–461 (2009).
94. Takeuchi, J. K. & Bruneau, B. G. Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. *Nature* **459,** 708–711 (2009).
95. Viswanathan, S. R., Daley, G. Q. & Gregory, R. I. Selective blockade of microRNA processing by Lin28. *Science* **320,** 97–100 (2008).
96. Feng, B., Ng, J. H., Heng, J. C. & Ng, H. H. Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. *Cell Stem Cell* **4,** 301–312 (2009).
97. Collombat, P. *et al.* Opposing actions of Arx4 and Pax4 in endocrine pancreas development. *Genes Dev.* **15,** 2591–2603 (2003).
98. Lagha, M. *et al.* Pax3/7:Foxc2 reciprocal repression in the somite modulates multipotent cell fates. *Dev. Cell,* (in the press).

Acknowledgements We would like to thank J. Sharpe, C. Peterson, J. Brickman and D. Thieffry for feedback and suggestions. T.G. is an ICREA professor and T.E. is supported by an LRF specialist programme.

Author Contributions T.G. and T.E. together conceived the ideas encapsulated in the article and also drafted it jointly. Most of the figures were conceived by T.G. and modified by T.E.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence should be addressed to T.G. (thomas.graf@crg.es).

Vol 462|3 December 2009|**doi:10.1038/nature08592**

Direct cell reprogramming is a stochastic process amenable to acceleration

Jacob Hanna[1]*, Krishanu Saha[1]*, Bernardo Pando[2], Jeroen van Zon[2,3], Christopher J. Lengner[1],
Menno P. Creyghton[1], Alexander van Oudenaarden[2,3] & Rudolf Jaenisch[1,3]

Direct reprogramming of somatic cells into induced pluripotent stem (iPS) cells can be achieved by overexpression of Oct4, Sox2, Klf4 and c-Myc transcription factors, but only a minority of donor somatic cells can be reprogrammed to pluripotency. Here we demonstrate that reprogramming by these transcription factors is a continuous stochastic process where almost all mouse donor cells eventually give rise to iPS cells on continued growth and transcription factor expression. Additional inhibition of the p53/p21 pathway or overexpression of Lin28 increased the cell division rate and resulted in an accelerated kinetics of iPS cell formation that was directly proportional to the increase in cell proliferation. In contrast, Nanog overexpression accelerated reprogramming in a predominantly cell-division-rate-independent manner. Quantitative analyses define distinct cell-division-rate-dependent and -independent modes for accelerating the stochastic course of reprogramming, and suggest that the number of cell divisions is a key parameter driving epigenetic reprogramming to pluripotency.

Quantifying the efficiency and timescales of crucial events occurring during *in vitro* reprogramming to pluripotency[1–5] has been problematic due to the cellular and genetic heterogeneity of *de novo* infected somatic cells[6,7]. To circumvent the need for virus-mediated transduction and reduce the heterogeneity of reprogramming factor expression, a 'secondary' reprogramming transgenic system was devised where all somatic cells carry the same integration pattern of drug-inducible *Oct4*, *Sox2*, *Klf4* and *c-Myc* viral transgenes[8–11]. Although reprogramming of somatic cells from secondary mice was two orders of magnitude higher than in freshly infected somatic cells, only 1–20% of the induced cells generated iPS cells after 3–4 weeks of factor expression[9–13]. Furthermore, partially reprogrammed 'intermediate' cell lines have been derived in different experimental settings, some of which can give rise much later to fully reprogrammed iPS cells either spontaneously or upon additional manipulations[14,15]. Given that the timescale of several weeks and relatively low efficiencies persist even after controlling for adequate *Oct4*, *Sox2*, *Klf4* and *c-Myc* transgene expression, these studies left important questions relevant to the basic mechanisms of epigenetic reprogramming unresolved: how does the reprogramming process progress over time and what happens to the majority of the cells that do not become reprogrammed upon continued cell growth and expression of the reprogramming factors? Why do some somatic cells that circumvent senescence or apoptosis induced by *Oct4*, *Sox2*, *Klf4* and *c-Myc* convert into iPS cells earlier than others? Do all adult donor cells expressing *Oct4*, *Sox2*, *Klf4* and *c-Myc* reprogramming factors eventually give rise to iPS cells or would this be achieved only upon additional genetic or small molecule manipulation? Is high reprogramming efficiency restricted to non-lineage committed or adult stem cells[6,7,13,16,17]?

Models to account for the reprogramming process fall into two categories (models i–iv in Fig. 1). 'Deterministic' models posit that either 'all' (model i) or only a subset of 'elite' or 'stem-like' cells (model ii) within a donor population have the potential to generate iPS cells and are reprogrammed with a fixed latency. We define latency as the absolute time or the number of cell divisions that an individual donor cell undergoes until it gives rise to a daughter iPS cell. 'Stochastic' models posit that most if not all (model iii) or only a subset of 'elite' somatic cells (model iv) within a donor population have the potential to generate iPS cells, albeit with different latencies.

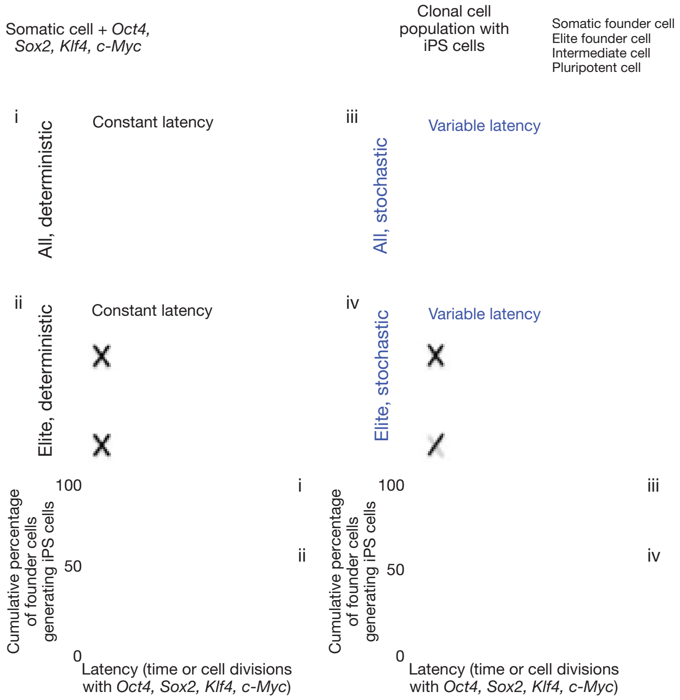


Figure 1 | Models of progressing to a pluripotent state during direct reprogramming. Four different models (i–iv) to account for the latency of donor somatic cells in progressing towards the iPS cell state following the expression of *Oct4*, *Sox2*, *Klf4* and *c-Myc* reprogramming factors. Latency can be measured in units of absolute time or cell divisions until the first iPS cell is generated from a monoclonal population. Graphs display the general shape of the reprogramming kinetics in the different models. Note that elite models do not necessarily reprogram more slowly as shown in the bottom plots.

---

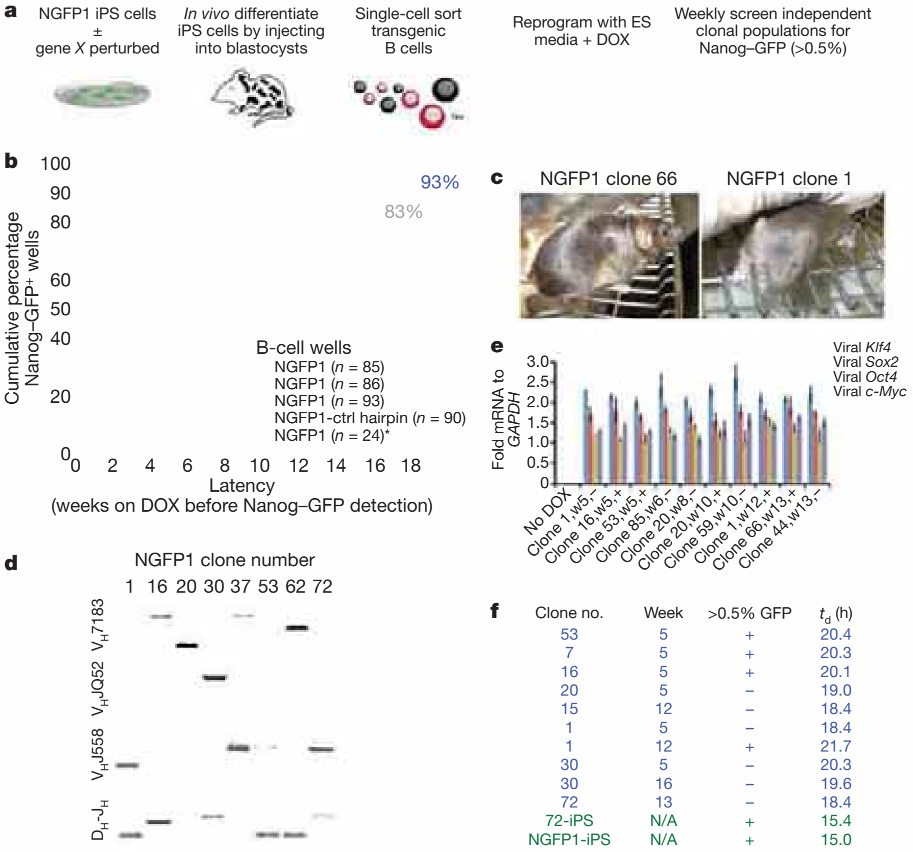
[1]The Whitehead Institute for Biomedical Research, [2]Department of Physics, [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA.
*These authors contributed equally to this work.

To gain insight into the mechanism of reprogramming and to address some of these questions, we have characterized the reprogramming efficiency and kinetics of over 1,000 somatic-cell-derived monoclonal populations expressing *Oct4*, *Sox2*, *Klf4* and *c-Myc* over an extended period of time and quantitatively defined distinct modes for changing the course of the reprogramming process upon additional genetic perturbations.

Reprogramming of monoclonal populations

We followed the reprogramming of individual somatic donor cells and studied their potential to generate iPS cells. Unlike fibroblasts, B-cell lineage-committed cells at the early pre-B-cell stage can be efficiently cloned as single cells immediately after isolation and were used as a defined, homogenous starting cell population for reprogramming into iPS cells. The NGFP1 iPS cell line was generated by infecting fibroblasts from Nanog–GFP reporter mice with doxycycline-inducible lentiviral vectors encoding the Oct4, Sox2, Klf4 and c-Myc transcription factors and injected into host blastocysts to generate secondary chimaeras[9,18,19] (Fig. 2a). NGFP1-derived secondary pre-B cells were single-cell sorted into individual wells and exhibited high cloning efficiency on doxycycline (>80%). Populations were serially passaged and monitored weekly for reactivation of the endogenous Nanog–GFP knock-in reporter (Supplementary Fig. 1), which represents one of the final events during reprogramming[4,20–22]. We defined reprogramming efficiency as the long-term potential of a cell to generate iPS daughter cells. A detection value of >0.5% for Nanog–GFP$^+$ cells per well reproducibly allowed for stable derivation of Nanog–GFP$^+$ iPS cells upon doxycycline withdrawal (Supplementary Fig. 2), and was set as the minimal threshold for defining positive detection of iPS cells in clonal populations.

Nanog–GFP$^+$ cells were detected after 2 weeks of doxycycline induction, with ~3–5% of the wells generating Nanog–GFP$^+$ cells at 2 weeks[10–12] (Fig. 2b). The remaining wells contained viable cells, could be propagated in the presence of doxycycline, and uniformly silenced somatic/haematopoietic surface markers[15,22] (Supplementary Fig. 3). Upon extended culture, the fraction of wells with iPS cells gradually increased and by 18 weeks after doxycycline addition >92% of the wells had produced Nanog–GFP$^+$ cells (Fig. 2b), demonstrating that most if not all donor cells have the potential to generate iPS cells. The reprogramming kinetics were reproducible in independent experiments (Fig. 2b) and were not restricted to B cells, as CD11b$^+$ monocytes that showed comparable transgene induction levels generated iPS cells with similar kinetics and efficiency (Supplementary Fig. 4). Nanog–GFP$^+$ cell populations selected at random gave rise to stable doxycycline-independent iPS cell lines. Furthermore, all tested iPS cell lines had normal karyotypes and generated teratomas and chimaeras irrespective of whether they were derived at early or late time points during the process (Fig. 2c and Supplementary Fig. 5). The iPS cell lines carried distinct genetic heavy-chain rearrangements verifying their independent clonal origin (Fig. 2d). We assessed whether transgene expression levels or increased proliferation rate could underlie the well-to-well differences seen in the latency of reprogramming. The



Clone no.	Week	>0.5% GFP	t_d (h)
53	5	+	20.4
7	5	+	20.3
16	5	+	20.1
20	5	–	19.0
15	12	–	18.4
1	5	–	18.4
1	12	+	21.7
30	5	–	20.3
30	16	–	19.6
72	13	–	18.4
72-iPS	N/A	+	15.4
NGFP1-iPS	N/A	+	15.0

Figure 2 | Long-term analysis of reprogramming monoclonal populations. **a**, Schematic of experiments. **b**, Reprogramming of pre-B-cell monoclonal populations measured as the cumulative number of wells that became Nanog–GFP$^+$. *n* indicates number of populations monitored. Asterisk indicates flow cytometry for GFP detection was performed every 4 weeks. **c**, Chimaeric mice with agouti coat colour from iPS cells derived after 12–13 weeks of doxycycline (DOX). **d**, Heavy-chain rearrangements in iPS cells. Genetic rearrangements were confirmed by sequencing to distinguish them from background signals, and are highlighted with red triangles. **e**, Relative transgene induction levels of monoclonal populations on doxycycline. Time is shown in weeks (w). Error bars indicate standard deviation (s.d.; $n = 3$). **f**, The population-averaged doubling time, t_d, for each clonal population. Boxes delineate cases where the same clonal population was measured at different times during doxycycline induction. The lower two rows (green) represent subcloned iPS cell lines. From left to right, columns show: clone number; weeks on doxycycline; Nanog–GFP >0.5% status (+/−); population-averaged doubling time (h).

population-averaged cell doubling times (t_d) and transgene induction levels during the reprogramming process were similar in NGFP1 clonal populations irrespective of time on doxycycline or whether these populations contained a Nanog–GFP$^+$ fraction (Fig. 2e, f).

These results suggest the following: (1) reprogramming of somatic cells is a continuous stochastic process where nearly all somatic donor cells have the ability to give rise to iPS cells upon continuous passaging and expression of *Oct4*, *Sox2*, *Klf4* and *c-Myc*. (2) Although reprogrammed cells do not appear before 8–10 days of *Oct4*, *Sox2*, *Klf4* and *c-Myc* expression[21,22], the time of doxycycline exposure or number of cell divisions achieved before a given clonal population generates iPS cells varies widely. (3) Our data are not consistent with an 'elite component' in reprogramming, as most if not all lineage-committed B cells or monocytes are able to generate iPS cells rather than only a small fraction of putative somatic stem cells present in the donor cell population[7,12,13]. (4) Finally, somatic cells reprogram with different latencies that cannot be predicted on the basis of time of doxycycline exposure or proliferation rate, consistent with undefined stochastic events driving the process (model iii in Fig. 1).

Cell division rate and reprogramming

We next characterized parameters of the reprogramming process by introducing defined genetic perturbations. Recently, *p53* (also called *Trp53*) inhibition has been shown to enhance the efficiency of iPS cell formation from fibroblasts by direct viral infection or transient transfection protocols by reducing apoptosis after initial transgene induction[16,23–27]. We tested whether and how *p53* inhibition would influence the reprogramming of secondary transgenic NGFP1 iPS-cell-derived pre-B cells that have a high single-cell cloning efficiency and stably grow in the presence of doxycycline without requiring additional immortalization and with only background apoptosis levels (Supplementary Fig. 6). NGFP1 iPS cells were infected with a constitutively expressed lentiviral vector encoding a short interfering RNA (siRNA) hairpin for *p53* (Supplementary Fig. 7)[28]. Infected cells were injected into host blastocysts and NGFP1-p53 knockdown (p53^{KD}) B cells were single-cell sorted and cultured in doxycycline. *p53* inhibition did not alter transgene expression levels or affect the already residual levels of apoptosis (Fig. 3a and Supplementary Figs 6 and 8), but rather shortened the cell-population-averaged doubling time of p53^{KD} cells in the presence of doxycycline by ~2-fold as compared to control NGFP1-derived cells (Fig. 3b and Supplementary Fig. 9). The kinetics of iPS cell formation proceeded with a significantly accelerated rate, with 93% of the wells producing Nanog–GFP$^+$ cells within 8 weeks of doxycycline, as compared to 17 weeks for the control cells (Fig. 3c). The iPS cell lines were doxycycline independent, expressed pluripotency markers and generated teratomas and mouse chimaeras (Supplementary Fig. 10).

To assess whether the enhanced reprogramming could be attributed to the effect of *p53* inhibition on proliferation rate, we estimated, based on the population-averaged doubling times measured throughout the process, how many cell divisions have occurred for each NGFP1 and NGFP1-p53^{KD} clonal population during the latency period. Upon rescaling of latency by the doubling time (Fig. 3c, d), the cumulative fraction of wells generating iPS cells collapsed to the same statistically significant distribution with cell division number for both the NGFP1 and NGFP1-p53^{KD} wells (log-rank test for dissimilarity, $P = 0.518$). A similar analysis on NGFP1 and NGFP1-p53^{KD} CD11b$^+$ cell-derived clonal populations showed a comparable distribution (log-rank test for dissimilarity, $P = 0.209$; Supplementary Fig. 11). Latencies were not normally distributed about the mean latency, but rather were better fit with a gamma distribution (Supplementary Figs 12 and 13). Knockdown of the *p21* gene, a downstream effector of p53 which regulates cell-cycle progression[29], recapitulated the change in cell division rate and acceleration of reprogramming dynamics upon *Oct4*, *Sox2*, *Klf4* and *c-Myc* expression (Fig. 3a–d and Supplementary Figs 6–9 and 14), further substantiating that change in cell division rate directly rescales the kinetics of reprogramming to distributions similar to Oct4, Sox2, Klf4 and c-Myc alone. Finally, we studied the effect of Lin28 overexpression on the reprogramming dynamics, as it enhances the reprogramming of human fibroblasts and acts as an oncogene by modulating the expression of cell-cycle regulators[5,30]. NGFP1-Lin28^{OE} (Lin28 overexpresser)-derived B-cell populations demonstrated accelerated reprogramming kinetics that directly correlated with the observed increase in cell division rate on doxycycline ($P = 0.327$; Fig. 3a–d). In summary, our data demonstrate that both p53/p21

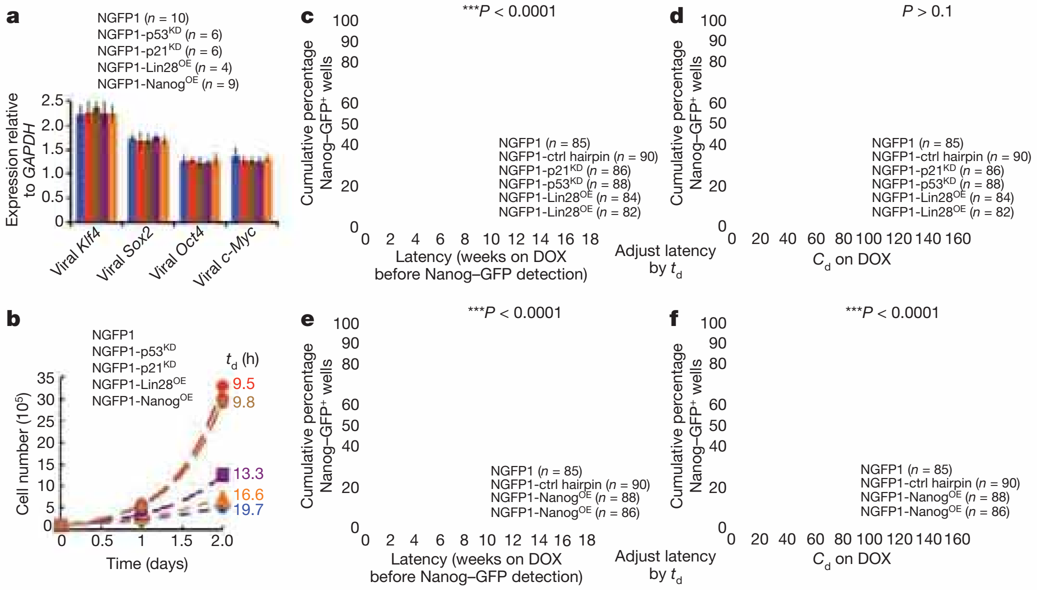


Figure 3 | Cell-division-rate-dependent and -independent acceleration of reprogramming. a, Average induction levels for transgenes in different NGFP1 cell populations. *n* indicates number of populations sampled per group, presented as mean ± s.d. **b**, Growth curves for cells on doxycycline. Exponential growth (dashed line) described the data well ($R^2 = 0.97$–1.0), and the population-averaged doubling times (t_d) were calculated from these fits (Supplementary Fig. 9). **c**, As in Fig. 2b, latencies for reprogramming various clonal B-cell-derived populations. NGFP1-p53^{KD}, NGFP1-p21^{KD} and NGFP1-Lin28^{OE} wells were statistically distinct from the NGFP1 and NGFP1-control hairpin wells ($P < 0.0001$, log-rank test for dissimilarity). **d**, Rescaling time by t_d provides an estimate for the number of cell divisions occurring during latency. C_d, population-averaged number of cell divisions on doxycycline before Nanog–GFP detection. No statistical difference between groups was observed after rescaling time by t_d ($P > 0.1$). **e**, **f**, As in **c**, **d**, but for NGFP1-Nanog^{OE} wells. *n* indicates number of populations monitored.

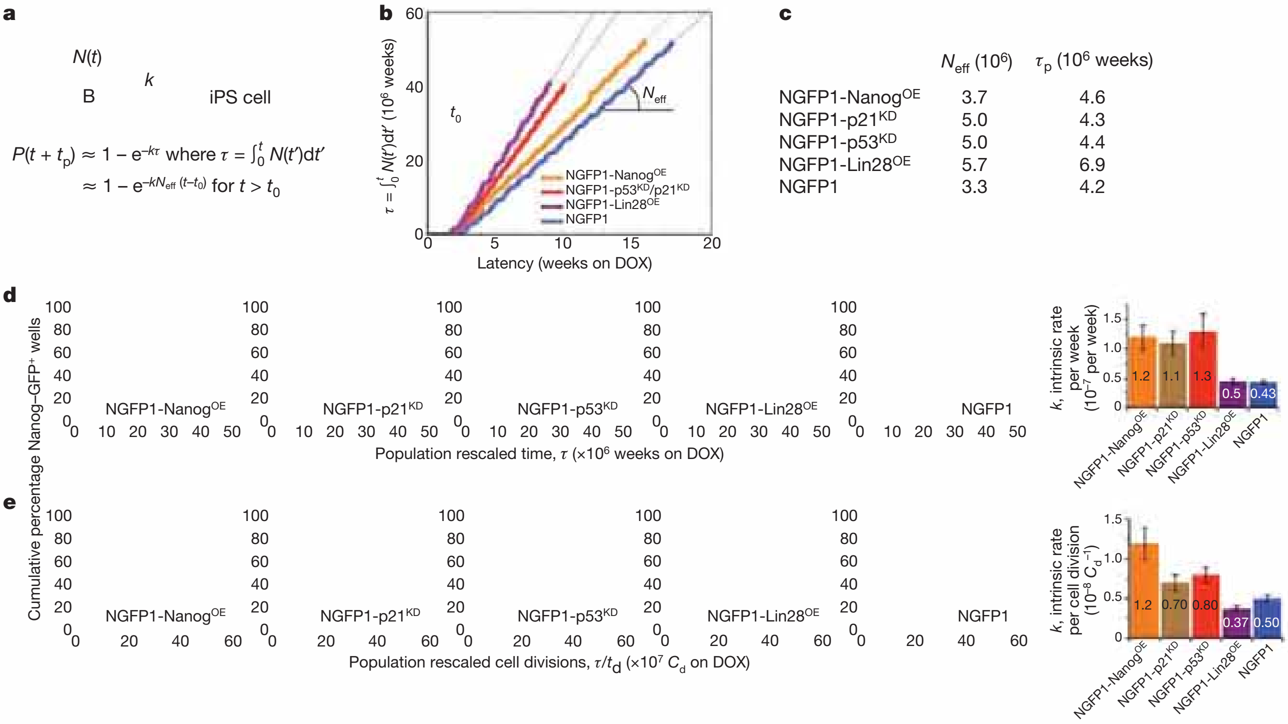
inhibition or Lin28 overexpression accelerate the reprogramming process rather than enhancing overall efficiency, as the cells divide more rapidly, resulting in an increased cumulative probability for the stochastic events to occur earlier in time (Fig. 3a–d).

Cell-division-rate-independent acceleration

We investigated whether reprogramming could be accelerated by mechanisms that are independent of cell proliferation rate. Nanog is a pluripotency factor expressed in the inner cell mass, and deriving embryonic stem cells and iPS cells requires the presence of functional endogenous *Nanog* alleles[31]. Moreover, Nanog enhances the transfer of pluripotency by cell fusion and facilitates direct reprogramming of human cells[5,32–35]. Therefore, we aimed to investigate the effect of additional ectopic Nanog expression on the reprogramming kinetics by Oct4, Sox2, Klf4 and c-Myc proteins. B cells were derived from an NGFP1-Nanog^{OE} (Nanog overexpresser) iPS cell line carrying a doxycycline-inducible Nanog-encoding transgene (Supplementary Fig. 7). The presence of the transgene did not alter the expression of *Oct4*, *Sox2*, *Klf4* and *c-Myc* (Fig. 3a). Overall efficiency of iPS cell formation was similar to that of the control cells but reprogramming occurred with a significantly accelerated kinetics, with 94% of the wells producing pluripotent Nanog–GFP$^+$ cells within 8 weeks of doxycycline (Fig. 3e and Supplementary Figs 15 and 16). Nanog overexpression slightly increased the cell-population-averaged doubling time of doxycycline-induced NGFP1-Nanog^{OE} cells compared to control NGFP1-derived reprogramming cells (Fig. 3b)[36,37]. However, even after rescaling time on doxycycline by the change in doubling time (Fig. 3f), the cumulative fraction of wells generating iPS cells did not collapse to the same distribution with cell division number (log-rank test for dissimilarity, $P < 0.0001$). The median number of cell division divisions before wells produced iPS cells was significantly reduced from about 70 cell divisions in the NGFP1, NGFP1-p21^{KD} and NGFP1-p53^{KD} wells to 50 cell divisions in the NGFP1-Nanog^{OE} wells (Fig. 3d, f and Supplementary Fig. 12d). These results suggest that Nanog overexpression accelerates the reprogramming kinetics by cell-intrinsic mechanisms that are independent of an altered cell proliferation rate.

Numerical modelling of reprogramming

We used our data sets to conduct simulations of the stochastic reprogramming process occurring in individual cells. To assess the intrinsic rate of reprogramming per cell, we tested whether a simple stochastic model would describe our observations. The model considers the reprogramming of B cells to occur as a one-step process with a constant cell-intrinsic rate k (Fig. 4a). Because iPS cells were adherent and selectively retained throughout the culturing procedures, the latency measured for each well was representative of the timing at which the first B cell in the population reprogrammed plus a time delay, t_p, during which the daughter cells needed to grow to reach the detection threshold. At any given time t, the number of cells in each well, $N(t)$, scales the rate at which the first reprogramming event takes place, and the cumulative probability distribution of reprogramming times is $P(t + t_p) \approx 1 - e^{-k\tau}$, where τ is the 'population-rescaled time', $\tau = \int_0^t N(t')dt'$. Interpreting the experimental results in terms of this rescaled version of time allowed one to separate the contributions of population size and cell-intrinsic reprogramming rate (Fig. 4). Cell division rate can control the population size, N, and thus re-scales time, affecting the observed rate of reprogramming. For example, if the cell division rate and culturing procedures were modified to an effective population size of 1,000 cells in each well rather than $\sim 10^6$ cells per well (Fig. 4b, c), longer times would have been necessary to reach >90% reprogrammed wells.



	N_{eff} (10^6)	τ_p (10^6 weeks)
NGFP1-Nanog^{OE}	3.7	4.6
NGFP1-p21^{KD}	5.0	4.3
NGFP1-p53^{KD}	5.0	4.4
NGFP1-Lin28^{OE}	5.7	6.9
NGFP1	3.3	4.2

Figure 4 | Quantitative analysis of cell reprogramming. a, Stochastic model summary. Sequential re-plating of individual wells during each experiment establishes that, after a time t_0 representative of the time at which the re-platings started, each experiment can be described in terms of a population of an effective size, N_{eff}. **b**, Estimate of the population rescaled time, τ, throughout each experiment. After t_0, population dynamics are effectively described by a fixed population of size, N_{eff}. NGFP1-p53^{KD} and NGFP1-p21^{KD} have similar dynamics. **c**, N_{eff} and the population-rescaled average proliferation times, τ_p, estimated as the population-rescaled time necessary for one iPS cell to reach the detection threshold ($\tau_p = t_{d,i}N_{eff}\log_2(\rho N_{eff})$, where ρ is the detection threshold and $t_{d,i}$ is the doubling time of iPS cells). **d**, Cumulative percentage of Nanog–GFP$^+$ wells as a function of τ, and best fits according to the proposed model. Modelling results are shown as thick lines; experiments are shown as dots. The far-right graph indicates best-fit estimates of the cell-intrinsic rate k expressed in terms of weeks. **e**, As in **d**, but per population-rescaled cell divisions, τ/t_d, instead of per τ units. t_d is the doubling time of the populations. Error bars indicate 95% confidence intervals.

Using cell population dynamics to account for the growth and culturing procedures used across the different NGFP1 lines (Fig. 4b, c), we calculated the population-rescaled time and fit the cell-intrinsic rate k using a likelihood maximization approach (Fig. 4d, e and Supplementary Fig. 17). Reasonable agreement was seen between best fits and the experiments. Furthermore, to take into account fluctuations in cell division times, number of cells and potential loss of iPS cells during culturing, in addition to the stochasticity in the cell-intrinsic reprogramming process, we implemented a detailed computer simulation of each experiment (Supplementary Fig. 18). Simulation results were consistent with those achieved using the analytical approach (Fig. 4d, e). The inferred intrinsic reprogramming rates per cell division for NGFP1-p53^{KD} and NGFP1-p21^{KD} cells were similar to NGFP1 cells (Fig. 4e and Supplementary Figs 17 and 18), suggesting that p53/p21 pathway inhibition in our system accelerates reprogramming through a predominately cell-division-rate-dependent mechanism. The slightly higher cell-intrinsic rates in some of our modelling results for p53/p21 inhibition (Fig. 4d, e and Supplementary Figs 17 and 18) may reflect a modest acceleration due to cell-division-rate-independent mechanisms and we cannot exclude that inhibition of the p53/p21 pathway slightly enhances reprogramming by additional mechanisms as well as in different experimental systems. The acceleration observed for NGFP1-Lin28^{OE} cells occurred predominantly because of a larger effective population size resulting from a faster cell division rate on doxycycline when following a culturing procedure similar to that used for NGFP1 cells (Fig. 4b, c). Only in the case of NGFP1-Nanog^{OE} cells, higher than 1.75-2 fold intrinsic reprogramming rates per cell division compared to NGFP1 cells was robustly obtained in all modelling approaches (Fig. 4e and Supplementary Figs 17 and 18), supporting the hypothesis that Nanog overexpression accelerates reprogramming through a predominately cell-division-rate-independent mechanism. Future characterization of how Nanog molecularly orchestrates the re-establishment of the core pluripotency circuitry is of great interest.

Discussion

We used a highly defined quantitative system for analysing the reprogramming process and suggest that yet-to-be-defined rate-limiting stochastic events occur as a function of cell division before the fully reprogrammed pluripotent state is attained. Upon *Oct4*, *Sox2*, *Klf4* and *c-Myc* induction in our system, most, if not all, monocytes and lineage-committed B cells harbouring genetic rearrangements have the potential to generate iPS cells, albeit with very different latencies (model iii, Fig. 1). These results are inconsistent with models in which iPS cells preferentially arise from a particular epigenetic state in the donor cell population, such as a progenitor or an adult stem cell (models ii and iv, Fig. 1). The dynamics of direct *in vitro* reprogramming to pluripotency by Oct4, Sox2, Klf4 and c-Myc are consistent with a continuous stochastic process, in which the conversion from a somatic cell to an iPS cell can be thought of as a drift in cell state. This cell state can be defined by a gene expression or epigenetic pattern, the fluctuations of which drive the conversion to an iPS cell. These fluctuations are probably due to cell-to-cell extrinsic heterogeneity or the inherent stochastic nature of gene expression or regulatory signalling processes[38–40]. In the simplest scenario, the process could be modelled as a one-step rate-limiting transition characterized by a cell-intrinsic rate, and we found that this model describes reasonably the observed dynamics across all of the different cell lines. Our model relies on inferring a single cell-intrinsic rate averaged over the entire course of reprogramming, although the model does not optimally describe the reprogramming behaviour at early times after transgene induction (Fig. 4). This deviation between modelling and experimental results may indicate that the cell-intrinsic rate changes over the course of reprogramming within clonal populations. Intraclonal monitoring of reprogramming dynamics at various points throughout the process could be used to add further mechanistic detail to the model, and to decipher the identity and sequence of rate-limiting step(s) and epigenetic changes that occur during reprogramming.

Quantitative analysis of the differences in reprogramming latency upon different genetic perturbations defined two distinct modes of accelerating the reprogramming process (Fig. 5 and Supplementary Fig. 19). In the 'cell-division-rate-dependent' mode, the cumulative probability for successful reprogramming is higher and can be achieved earlier in time and directly proportional to the augmentation in cell division rate, whereas in the 'cell-division-rate-independent' mode, reprogramming acceleration occurs over a lower average number of cell divisions. Notably, the different modes need not be mutually

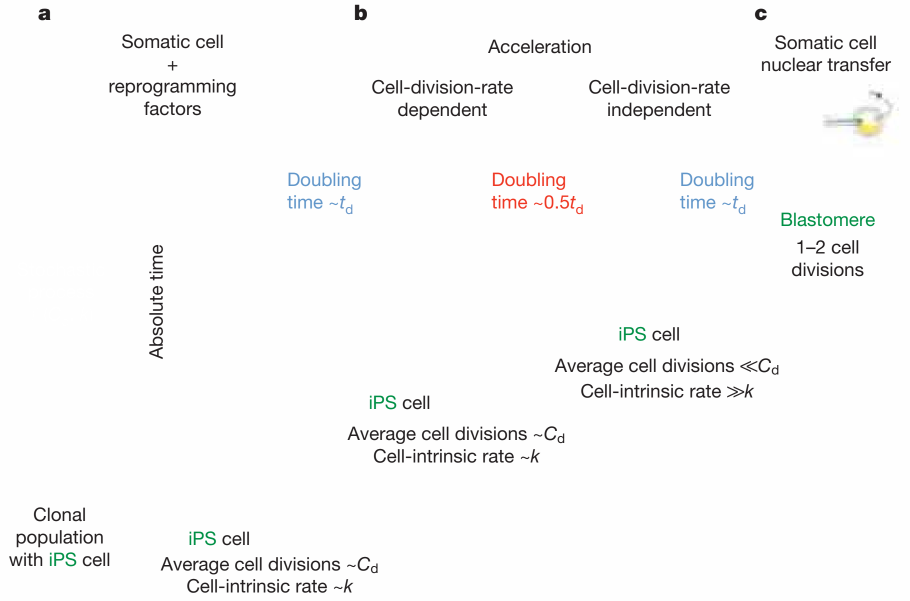


Figure 5 | Distinct modes for accelerating reprogramming to pluripotency. **a**, Nearly all donor cells can give rise to iPS cells via a stochastic process. Two parameters characterize the kinetics of the process: the average number of cell divisions required, C_d, and the cell-intrinsic reprogramming rate per cell division, k. **b**, Accelerating reprogramming in a cell-division-rate-dependent manner still requires as many divisions as the unaccelerated reference scenario (that is, still C_d on average) but occurs earlier in time because cells divide faster, whereas in the cell-division-rate-independent mode, the cell-intrinsic rate reflecting the occurrence of an unknown stochastic event(s) is enhanced ($\gg k$) and reprogramming is achieved within a lower average number of divisions ($\ll C_d$). **c**, In comparison, somatic cell nuclear transfer can reprogram within 1–2 cell divisions.

exclusive as certain perturbations could enhance or inhibit reprogramming via both cell-proliferation-dependent and -independent effects. Recent studies failed to attain correlation between cell division rate and reprogramming efficiency; however, these studies relied on measuring proliferation in the absence of *Oct4*, *Sox2*, *Klf4* and *c-Myc* transgenes[16] or by inducing proliferation with haematopoietic cytokines for which the receptors are rapidly silenced after transgene induction[13]. Our experimental and modelling data suggest that close monitoring of transgene induction, plating efficiency, cell proliferation and changes in population size throughout the experiments are needed to gain insight into the stochastic process of reprogramming. We consider two possibilities for the ability of increased cell division rate to accelerate and drive the kinetics of the reprogramming process: (1) cell division could amplify the number of daughter cells from partially reprogrammed cells where each resulting individual cell has an independent probability of progressing towards becoming an iPS cell; and (2) nuclear changes during cell division may facilitate the acquisition of epigenetic marks such as DNA and histone modifications that allow the re-establishment of the core transcriptional circuitry that stabilizes pluripotency[6,15,41,42].

iPS cells can be derived by various combinations of transcription factors and/or small molecules[3,5,43,44]. Our study identifies the average number of cell divisions required to give rise to an iPS cell at a particular efficiency as a key parameter that may be affected by different reprogramming strategies. After nuclear transfer, the pluripotency gene *Oct4* (also called *Pou5f1*) of the somatic nucleus is reactivated in the cloned embryo within 1–2 cell divisions[42,45,46] (Fig. 5c), indicating that the egg cytoplasm carries as-of-yet undefined determinants that accomplish robust reprogramming within very few cell divisions. These observations raise questions of whether direct reprogramming strategies devised so far can be robustly enhanced (for example, by optimizing stoichiometry, small molecule treatment or utilizing oocyte-derived transcript libraries) to enable complete *in vitro* reprogramming within only a few cell divisions.

METHODS SUMMARY

The NGFP1 iPS cell line was derived after culturing the infected mouse embryonic fibroblasts (MEFs) carrying ROSA26-M2rtTA mice and Nanog–GFP alleles with doxycycline-inducible lentiviruses encoding *Oct4*, *Klf4*, *c-Myc* and *Sox2* cDNA driven by the TetO/CMV promoter[9]. To generate the different genetically perturbed NGFP1 subclones, 50,000 NGFP1 cells were infected with the pSicoR vector[28] encoding a hairpin for the *p53* or *p21* tumour suppressor genes or with doxycycline-inducible lentiviruses encoding *Nanog* or *Lin28* cDNAs. To generate mouse chimaeras iPS cells were injected into diploid blastocysts that were initially harvested 94–98 h after hormone injection and placed in a drop of DMEM with 15% FCS under mineral oil. Pre-BCR$^+$ (Igμ^+IgK$^-$IgL$^-$) early-pre B cells were single-cell sorted from 3–5-week-old chimaeras and were plated in gelatinized and irradiated OP9-coated wells with ES media supplemented with doxycycline (4 μg ml^{-1}). IL-7 and SCF (10 ng ml^{-1} each, Peprotech) were supplemented to the media for the first 1–2 weeks to optimize cloning efficiency in the presence of doxycycline. After 2 weeks of doxycycline induction, cultures were trypsinized every week and passaged on gelatinized plates in ES cell medium plus doxycycline. Populations were defined positive for iPS cell generation when >0.5% of the 10,000 adherent/semi-adherent screened cells by FACS were GFP$^+$. NGFP1-p53^{KD} and p21^{KD} derived cells were passaged twice a week (versus once a week at the time of GFP assays for the other groups) to avoid over-confluence in the culture wells due to their accelerated proliferation rate. The reprogramming experiments were stopped at different time points and all non-reprogrammed populations were viable and demonstrated robust growth on doxycycline upon ending each experiment. For teratoma generation, 2×10^6 iPS cells were injected subcutaneously into both flanks of recipient SCID mice, and tumours were harvested for sectioning 3–6 weeks after initial injection.

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

Received 18 June; accepted 21 October 2009.

1. Hanna, J. *et al.* Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318,** 1920–1923 (2007).
2. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
3. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
4. Wernig, M. *et al. In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448,** 318–324 (2007).
5. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318,** 1917–1920 (2007).
6. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132,** 567–582 (2008).
7. Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. *Nature* **460,** 49–52 (2009).
8. Hanna, J., Carey, B. W. & Jaenisch, R. Reprogramming of somatic cell identity. *Cold Spring Harb. Symp. Quant. Biol.* doi:10.1101/sqb.2008.73.025 (2008).
9. Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133,** 250–264 (2008).
10. Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nature Biotechnol.* **26,** 916–924 (2008).
11. Markoulaki, S. *et al.* Transgenic mice with defined combinations of drug-inducible reprogramming factors. *Nature Biotechnol.* **27,** 169–171 (2009).
12. Hanna, J. *et al.* Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* **4,** 513–524 (2009).
13. Eminli, S. *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nature Genet.* **41,** 968–976 (2009).
14. Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136,** 364–377 (2009).
15. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).
16. Utikal, J. *et al.* Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature* **460,** 1145–1148 (2009).
17. Huangfu, D. *et al.* Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. *Nature Biotechnol.* **26,** 795–797 (2008).
18. Signer, R. A., Montecino-Rodriguez, E., Witte, O. N. & Dorshkind, K. Aging and cancer resistance in lymphoid progenitors are linked processes conferred by p16^{Ink4a} and Arf. *Genes Dev.* **22,** 3115–3120 (2008).
19. Chen, J., Lansford, R., Stewart, V., Young, F. & Alt, F. W. RAG-2-deficient blastocyst complementation: an assay of gene function in lymphocyte development. *Proc. Natl Acad. Sci. USA* **90,** 4528–4532 (1993).
20. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448,** 313–317 (2007).
21. Brambrink, T. *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2,** 151–159 (2008).
22. Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2,** 230–240 (2008).
23. Banito, A. *et al.* Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev.* **23,** 2134–2139 (2009).
24. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* **460,** 1132–1135 (2009).
25. Kawamura, T. *et al.* Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460,** 1140–1144 (2009).
26. Li, H. *et al.* The *Ink4/Arf* locus is a barrier for iPS cell reprogramming. *Nature* **460,** 1136–1139 (2009).
27. Marión, R. M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460,** 1149–1153 (2009).
28. Ventura, A. *et al.* Cre-lox-regulated conditional RNA interference from transgenes. *Proc. Natl Acad. Sci. USA* **101,** 10380–10385 (2004).
29. Rowland, B. D., Bernards, R. & Peeper, D. S. The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene. *Nature Cell Biol.* **7,** 1074–1082 (2005).
30. Xu, B., Zhang, K. & Huang, Y. Lin28 modulates cell growth and associates with a subset of cell cycle regulator mRNAs in mouse embryonic stem cells. *RNA* **15,** 357–361 (2009).
31. Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138,** 722–737 (2009).
32. Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230–1234 (2007).
33. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113,** 631–642 (2003).
34. Navarro, P. *et al.* Molecular coupling of Xist regulation and pluripotency. *Science* **321,** 1693–1695 (2008).
35. Silva, J., Chambers, I., Pollard, S. & Smith, A. Nanog promotes transfer of pluripotency after cell fusion. *Nature* **441,** 997–1001 (2006).
36. Darr, H., Mayshar, Y. & Benvenisty, N. Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. *Development* **133,** 1193–1201 (2006).
37. Niwa, H., Ogawa, K., Shimosato, D. & Adachi, K. A parallel circuit of LIF signalling pathways maintains pluripotency of mouse ES cells. *Nature* **460,** 118–122 (2009).
38. Huang, S. Reprogramming cell fates: reconciling rarity with robustness. *Bioessays* **31,** 546–560 (2009).
39. Kærn, M., Elston, T. C., Blake, W. J. & Collins, J. J. Stochasticity in gene expression: from theories to phenotypes. *Nature Rev. Genet.* **6,** 451–464 (2005).

40. Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135,** 216–226 (2008).
41. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122,** 947–956 (2005).
42. Egli, D., Birkhoff, G. & Eggan, K. Mediators of reprogramming: transcription factors and transitions through mitosis. *Nature Rev. Mol. Cell Biol.* **9,** 505–516 (2008).
43. Kim, J. B. *et al.* Oct4-induced pluripotency in adult neural stem cells. *Cell* **136,** 411–419 (2009).
44. Shi, Y. *et al.* Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. *Cell Stem Cell* **3,** 568–574 (2008).
45. Boiani, M. *et al.* Variable reprogramming of the pluripotent stem cell marker Oct4 in mouse clones: distinct developmental potentials in different culture environments. *Stem Cells* **23,** 1089–1104 (2005).
46. Egli, D., Rosains, J., Birkhoff, G. & Eggan, K. Developmental reprogramming after chromosome transfer into mitotic mouse zygotes. *Nature* **447,** 679–685 (2007).

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

Acknowledgements We would like to thank N. Maheshri, K. Wood, Y. Bugnanim, B. Carey, J. Cassady and F. Soldner for commenting on the manuscript; J. Dausmann, S. Markoulaki, P. Schorderet and M. Pawlak for conducting blastocysts injections; and members of the Jaenisch laboratory for discussions. R.J. is supported by grants from the NIH (RO1-HDO45022, R37-CA084198, RO1-CA087869). J.H. is supported by a postdoctoral fellowship from Helen Hay Whitney Foundation. K.S. is supported by the Society in Science: The Brano-Weiss fellowship. J.v.Z. is supported by the Human Frontiers Science Program. A.v.O. is supported by a National Institutes of Health (NIH) Director's Pioneer award and an award from the National Cancer Institute (U54CA143874).

Author Contributions J.H., K.S. and R.J. conceived the ideas for this study and designed the experiments. J.H. conducted the reprogramming experiments and data collection. M.P.C. and C.J.L. assisted in construct preparation, transgene induction and blastocyst injections. K.S., B.P., J.v.Z, J.H., and A.v.O. performed statistical and numerical analyses, and also constructed the stochastic models. J.H., K.S. and R.J. wrote the manuscript with contributions from all other authors.

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

METHODS

Pluripotent lines and viral infections. Mouse embryonic fibroblasts (MEFs) used to derive primary induced pluripotent stem (iPS) cell lines by infections with inducible lentiviruses were harvested at 13.5 days post coitum from F_1 mating between ROSA26-M2rtTA and Nanog–GFP mice[9]. Lentiviral preparation and infection with doxycycline-inducible lentiviruses encoding *Oct4*, *Klf4*, *c-Myc* and *Sox2* cDNAs driven by the TetO/CMV promoter were previously described[21]. The NGFP1 iPS cell line used in this study was derived after culturing the infected MEFs with doxycycline, and grew stably in culture independent of doxycycline. To generate NGFP1-p53$^{\mathrm{KD}}$ and p21$^{\mathrm{KD}}$ clonal cell lines, 50,000 NGFP1 cells were infected with the pSicoR-PGK-puro vector encoding a hairpin for mouse *p21* (targeting sequence: GCAGATTGGTCTTCTGCAA) and a previously described specific hairpin for the mouse *p53* tumour suppressor gene[28]. Hairpin against the CD8 sequence was used as control where indicated. A total of 10 μg of lentiviral vector and packaging vectors was co-transfected in 293T cells by using the FuGENE 6 reagent (Roche Diagnostics). Supernatants were collected 36–48 h after transfection, filtered through a 0.45-μm filter, and used to infect the iPS cells for 48 h. Afterwards, the cells were trypsinized, plated at low densities and subcloned to test for knockdown specificity (NGFP1 already contains a puromycin selection cassette in the ROSA26 locus restricting our ability to use puromycin as a marker for infected cells). To verify integration of PsicoR lentiviral vectors, the following oligonucleotides were used: forward, CCCGGTTAATTTGCATATAATATTTC; reverse, CATGATACAAAGGCATTAAAGCAG. cDNAs encoding *Nanog* and *Lin28* transcripts where cloned into the EcoRI site of TetO-FUW lentiviral vectors and were used to infect NGFP1 iPS cells. Integration for TetO-Nanog and TetO-Lin28 lentiviral vectors was verified by PCR and Southern blot analysis on genomic DNA using the EcoRI-digested insert as a probe (data not shown). A NGFP1 subclone that demonstrated >95% constitutive knockdown of *p53* (*Trp53*) or *p21* gene products, termed NGFP1-p53$^{\mathrm{KD}}$ or NGFP1-p21$^{\mathrm{KD}}$, was used in the subsequent experiments. iPS cells were cultured in DME containing 15% fetal calf serum (FCS), leukaemia inhibiting factor (LIF), penicillin/streptomycin, L-glutamine, β-mercaptoethanol and non-essential amino acids. Chromosomal karyotyping of iPS cell lines was performed by Cell Line Genetics on 20 G-banded metaphase cells from each line tested.

Chimaera and teratoma formation. iPS cells were injected in C57B6 × 129Sv F_1 *Rag2*$^{-/-}$ or BDF2 diploid blastocysts that were initially harvested 94–98 h after hormone injection and placed in a drop of DMEM with 15% FCS under mineral oil. A flat-tip microinjection pipette with an internal diameter of 1.2–1.5 mm was used for iPS cell injection (using a Piezo micromanipulator). A controlled number of cells was injected into the blastocyst cavity. After injection, blastocysts were returned to KSOM media (Invitrogen) and placed at 37 °C until transferred to recipient females. Ten to fifteen injected blastocysts were transferred to each uterine horn of 2.5 days post coitum pseudo-pregnant females. To recover full-term pups, recipient mothers were killed at 19.5 days post coitum. For teratoma generation, 2×10^6 iPS cells were injected subcutaneously into both flanks of recipient SCID mice, and tumours were harvested for sectioning 3–6 weeks after initial injection.

Reprogramming into iPS cells. Transgenic pre-BCR^{+} ($\mathrm{Ig}\mu^{+}\mathrm{IgK}^{-}\mathrm{IgL}^{-}$) early pre-B cells, which are not generated in *Rag2*$^{-/-}$ mice owing to their inability to undergo heavy- and light-chain rearrangements, were isolated from bone marrow of 3–5-week-old *Rag2*$^{-/-}$ NGFP1 chimaeras and single-cell sorted into 96-well plates. This choice of host ensured that any isolated B cells were derived from the injected iPS cells, and not from the host blastocysts. In certain experiments, we labelled the NGFP1-iPS cell clone with a lentivirus constitutively expressing dTomato (a gift from K. Hochedlinger) and injected cells into BDF2 blastocysts to produce chimaeric mice, and used the dTomato as a marker of transgenic cells. One of the experimental replicates was performed on B220^{+}CD25^{+} pro-B cells isolated from reprogrammable mice carrying the same set of transgenes[11]. Cells were plated in gelatinized and irradiated (3,000 rad) OP9-coated wells in ES medium containing doxycycline (4 μg ml^{-1}). IL-7 and SCF (10 ng ml^{-1} each, Peprotech) were added to the medium for the first 1–2 weeks to optimize cloning efficiency in the presence of doxycycline (the plating efficiency of pre-B cells to grow as single cells was >80%). Cells on doxycycline did not exhibit contact inhibition and grew both on gelatin and in suspension. After 2 weeks of doxycycline induction, cultures were trypsinized every week and passaged on gelatinized plates in ES medium plus doxycycline. Populations were defined as being positive for iPS cell generation when >0.5% of the adherent/semi-adherent screened cells were GFP^{+}. Upon detection of a GFP^{+} fraction, further flow cytometric (FACS) assaying of the clonal population was discontinued and, in all cases, stable doxycycline-independent iPS cell lines were derived by growing the cells in the absence of doxycycline (which allows the iPS cells to overgrow transgene-dependent, partially reprogrammed GFP^{-} cells in the population). A fraction of ~250,000 cells, including one-quarter of all fully adherent cells, was replated at the end of each week (beginning at week 2) for further follow-up analysis. By including adherent cells in the culture plate upon passaging, we consistently retained all iPS cells generated during a given culture period, as no iPS cells were detected by FACS in the non-adherent fraction (Supplementary Fig. 1). NGFP1-p53$^{\mathrm{KD}}$ and NGFP1-p21$^{\mathrm{KD}}$ derived cells were passaged and replated twice a week (versus once a week at the time of GFP assays for other cells) to avoid over-confluence in the culture wells due to their higher proliferation rate. Population-averaged doubling time for reprogramming and iPS cell populations was determined by plating 10^5 cells, and counting total cell number after 24 and 48 h in duplicates. Numbers of cells as a function of time were fit using exponential growth regression in Excel (R^2 ranging from 0.97 to 1.00). The CD11b^{+} myeloid fraction was sorted from spleen and cells were plated in gelatinized and irradiated OP9-coated 96-well plates with ES medium supplemented with doxycycline (4 μg ml^{-1}). M-CSF, Flt3L, LPS and SCF (10 ng ml^{-1}, Peprotech) was added to the medium for the first 2 weeks to boost plating efficiency. Wild-type CD11b^{+} cells were obtained from 'reprogrammable' transgenic mice carrying identical doxycycline-inducible copies of the reprogramming factors *Oct4*, *Sox2*, *Klf4* and *c-Myc*, Rosa26-M2rtTA and a Nanog–GFP knock-in reporter. Unlike iPS cell chimaeras, all cells in these reprogrammable mice carried the same set of transgenes[11]. However, unlike for pre-B cells, which are not found in *Rag2*$^{-/-}$ hosts and thus all isolated cells were of transgenic origin, CD11b^{+} cells were isolated from reprogrammable mice carrying the same set of transgenes[11] or from dTomato-labelled chimaeric donor mice. V(D)J-IgH rearrangements were amplified from genomic DNA samples by PCR using degenerate primer sets as described[9].

Immunofluorescence. Cells were fixed in 4% paraformaldehyde for 20 min at 25 °C, washed three times with PBS and blocked for 15 min with 5% FBS in PBS containing 0.1% Triton X-100. After incubation with primary antibodies against Oct4 (Santa Cruz), Nanog (polyclonal rabbit, Bethyl) and SSEA1 (monoclonal mouse, Developmental Studies Hybridoma Bank) for 1 h in 1% FBS in PBS containing 0.1% Triton X-100, cells were washed three times with PBS and incubated with fluorophore-labelled appropriate secondary antibodies purchased from Jackson Immunoresearch. Specimens were analysed on an Olympus fluorescence microscope, and images were acquired with a Zeiss Axiocam camera.

Quantitative RT–PCR. Total RNA was isolated using the RNeasy Kit (Qiagen). Three micrograms of total RNA was treated with DNase I to remove potential contamination of genomic DNA using a DNA Free RNA kit (Zymo Research). One microgram of DNase-I-treated RNA was reverse transcribed using a First Strand Synthesis kit (Invitrogen) and ultimately re-suspended in 100 μl of water. Quantitative PCR analysis was performed in triplicate using 1/50 of the reverse transcription reaction in an ABI Prism 7000 (Applied Biosystems) with Platinum SYBR green qPCR SuperMix-UDG with ROX (Invitrogen). Error bars indicate standard deviation of triplicate measurements for each clone. Primers used for transgene-encoded amplification were as follows: *c-Myc*, forward 5′-ACCTAACTCGAGGAGGAGCTGG-3′, reverse 5′-TCCACATAGCGTAAAAGGAGC-3′; *Klf4*, forward 5′-ACACTGTCTTCCCACGAGGG-3′, reverse 5′-GGCATTAAAGCAGCGTATCCA-3′; *Sox2*, forward 5′-CATTAACGGCACACTGCCC-3′, reverse 5′-GGCATTAAAGCAGCGTATCCA-3′; *Oct4*, forward 5′-AGCCTGGCCTGTCTGTCACTC-3′, reverse 5′-GGCATTAAAGCAGCGTATCCA-3′; *Nanog*, forward 5′-ACATGCAACCTGAAGACGTG-3′, reverse 5′-CACATAGCGTAAAAGGAGCAA-3′. To ensure equal loading of cDNA into RT reactions, *GAPDH* mRNA was amplified using the following primers: forward 5′-TTCACCACCATGGAGAAGGC-3′, reverse 5′-CCCTTTTGGCTCCACCCT-3′. Data were extracted from the linear range of amplification.

Antibodies and apoptosis measurement assays. Fluorescently conjugated antibodies (phycoerythrin, FITC, Cy-Chrome or APC labelled) were used for flow cytometric analysis, and for cell sorting: anti-CD11b^{+}, pre-BCR, IgK, IgL, CD19, B220, CD45.2 and IL-7R antibodies (BD-Biosciences) were used. Enrichment for CD11b$^{\mathrm{high}}$ cells by using a CD11b magnetic bead isolation kit was carried out before sorting. Cell sorting was performed by using FACS-Aria (BD-Biosciences), and consistently achieved cell sorting purity of >99%. Antibodies for western blot analysis: anti-mouse p53 (BAF1355, R&D systems), anti-β actin (ab8226, Abcam) and anti-p21$^{\mathrm{cip}}$ (clone C-19, Santa Cruz Biotechnology). To determine the rate of apoptosis, samples were stained with the annexin V staining kit (BD Biosciences) and propidium iodide according to the manufacturer's recommendations and analysed by flow cytometry. In addition, samples were subjected to the Tdt-mediated dUTP-biotin nick end labelling (TUNEL) test for apoptotic cells by using the flow-cytometry-based MEBSTAIN apoptosis kit (MBL).

Non-parametric statistical analysis. The latency of iPS cell appearance of each well was used to generate a survival curve for each experiment. Latency was considered either as a function of absolute time (that is, the weeks on doxycycline until GFP was detected) or as a function of the population-averaged number of cell divisions during latency (that is, latency in time divided by doubling time). Any wells that failed to generate Nanog–GFP^{+} cells at the end of each experiment were considered to be censored, and are shown with solid geometric shapes in Figs 2 and 3. Histograms of the fraction of cells at initial GFP detection were performed using

Prism 5 (version 5.0b; Graphpad Software). Using the censored latency data sets, the log-rank (Mantel–Cox) test, a non-parametric statistical test appropriate for right-censored data, was used to test the null hypothesis that survival functions do not differ across groups. In contrast to parametric analyses, this test does not require knowledge about the shape of the survival curve or the distribution of survival times. Analysis was performed using Prism 5 (version 5.0b; Graphpad Software) and two-tailed *P*-values. *P*-values above 0.05 indicate that the latencies between the two groups were similar (that is, the treatment did not change survival; accept the null hypothesis at a 95% confidence level), whereas *P*-values less than 0.05 indicate that latencies between groups were not similar (that is, survival functions differ; reject the null hypothesis at a 95% confidence level).

Parametric statistical analysis. The censored latency data sets were also fit to several univariate probability distributions using maximum likelihood estimation via the 'dfittool' in Matlab (The Mathworks). For optimal univariate distribution fits, the chi-squared goodness-of-fit statistic, χ^2, was used to assess the quality of each fit. Gamma distributions had the lowest χ^2 of any fit distribution (see Supplementary Fig. 12).

Computational simulation. Simulations of the proposed model were generated using a hybrid scheme in which the size of each subpopulation (either B-cells or iPS cells) was evolved by considering (1) stochastic Gillespie-like evolution of Poissonian growth dynamics for small population sizes; (2) deterministic evolution of population sizes using a time step of $\Delta t = 0.001$ weeks whenever the probability of generating a new cell in the time Δt exceeded 0.1; (3) Gillespie-like evolution of the reprogramming transition; (4) periodic re-plating, selection of iPS cells and observation of the fraction of reprogrammed cells according to the details of the experimental protocol followed in each experiment.

nature Vol 462|3 December 2009|doi:10.1038/nature08613

ARTICLES

A gate–latch–lock mechanism for hormone signalling by abscisic acid receptors

Karsten Melcher[1]*, Ley-Moy Ng[1,2]*, X. Edward Zhou[1]*, Fen-Fen Soon[1,2]*, Yong Xu[1], Kelly M. Suino-Powell[1], Sang-Youl Park[3], Joshua J. Weiner[4], Hiroaki Fujii[3,5], Viswanathan Chinnusamy[3,5], Amanda Kovach[1], Jun Li[1,2], Yonghong Wang[6], Jiayang Li[6], Francis C. Peterson[4], Davin R. Jensen[4], Eu-Leong Yong[2], Brian F. Volkman[4], Sean R. Cutler[3], Jian-Kang Zhu[3,5] & H. Eric Xu[1]

Abscisic acid (ABA) is a ubiquitous hormone that regulates plant growth, development and responses to environmental stresses. Its action is mediated by the PYR/PYL/RCAR family of START proteins, but it remains unclear how these receptors bind ABA and, in turn, how hormone binding leads to inhibition of the downstream type 2C protein phosphatase (PP2C) effectors. Here we report crystal structures of apo and ABA-bound receptors as well as a ternary PYL2–ABA–PP2C complex. The apo receptors contain an open ligand-binding pocket flanked by a gate that closes in response to ABA by way of conformational changes in two highly conserved β-loops that serve as a gate and latch. Moreover, ABA-induced closure of the gate creates a surface that enables the receptor to dock into and competitively inhibit the PP2C active site. A conserved tryptophan in the PP2C inserts directly between the gate and latch, which functions to further lock the receptor in a closed conformation. Together, our results identify a conserved gate–latch–lock mechanism underlying ABA signalling.

Abscisic acid is a vital hormone used by plants in their responses to environmental stresses, such as drought, cold and salinity. Genetic studies have established a canonical signalling pathway mediated by PP2Cs and subfamily 2 of SNF1-related kinases (SnRK2 kinases), which act downstream of ABA sensing[1–5]. However, identification of bona fide ABA receptors has been a daunting task, full of controversy and frustration[6]. A new class of START protein receptors called the PYR/PYL/RCARs has recently been described[7–9]. ABA binding to members of this protein family enables them to bind to and inhibit PP2Cs such as HAB1, leading to the activation of the SnRK2 kinases that phosphorylate downstream effectors, such as the basic leucine-zipper transcription factors called ABFs/AREBs, thus switching on stress response programs[1,10–13]. The mechanisms of ABA binding and PP2C inhibition by the PYR/PYL/RCARs are currently unknown. In addition, the binding of ABA to RCAR1 (PYL9) and PYL5 is stimulated >10 fold by the presence of a PP2C, and many of the PYR/PYL/RCAR proteins show constitutive (that is, unliganded) binding to PP2Cs[7,8]. These observations have led to an important but currently unresolved question: does ABA bind to a PYR/PYL/RCAR-PP2C co-receptor, or does ABA bind first to the PYR/PYL/RCAR receptors followed by secondary PP2C interactions that stabilize ligand binding? Many of the most critical questions regarding how ABA receptors function will require structural elucidation of ligand-bound and ligand-free receptors, and a receptor–PP2C complex. Here we report the crystal structures of apo- and ABA-bound forms of PYL2, and the ABA–PYL2–HAB1 complex at resolutions of 1.95 Å, 1.85 Å and 1.95 Å, respectively, as well as the apo-PYL1 structure at a resolution of 2.40 Å. Comparison of the apo- and ABA-bound structures shows that ABA binding induces closure of two highly conserved β-loops that function as a gate and latch to sense ligand entry. In the apo form of the receptor, the gate and latch reside in an open conformation, which permits access of ABA to the ligand-binding pocket. ABA binding induces the closure of the gate and formation of a gate/latch interface, which in turn binds to the PP2C active site and converts the ABA-bound PYL2 into a competitive PP2C inhibitor. The interactions between PYL2 and HAB1 further induce conformational changes in both proteins that lock the gate into the closed position, which is primarily mediated by a conserved tryptophan of HAB1. Importantly, this residue also makes a water-mediated interaction with ABA, thus allowing HAB1 to sense the presence of ABA in the PYL2 ligand-binding pocket. The critical residues in ligand binding and the PYL2/HAB1 interface are highly conserved, suggesting that the gate–latch–lock mechanism of ligand sensing and signal transduction is likely to be used by the entire PYR/PYL/RCAR receptor family.

To prepare proteins for crystallization, the entire PYR/PYL family was expressed in *Escherichia coli* and screened for suitability in crystallization trials. PYR1 and PYL1–6 were soluble and yielded high levels of purified proteins (Supplementary Fig. 1a). Each of these family members was capable of interacting with the ABA-signalling PP2Cs HAB1, ABI1 and ABI2 (Supplementary Fig. 1b). However, as suggested previously by yeast two hybrid assays[7,8], the PYR/PYL proteins and PP2Cs showed differential interactions in response to the presence or absence of ABA, the particular stereoisomer used, and the particular PP2C probed. These biochemical observations point to possible functional variations between family members in PP2C binding and signalling. Among the proteins expressed, PYL1 and PYL2 produced high-quality crystals of apo receptors, and PYL2 crystals diffracted better than 2.0 Å both in the absence and presence of ABA (Supplementary Fig. 1c). PYL2 interacted with the PP2C

---

[1]Laboratory of Structural Sciences, Van Andel Research Institute, 333 Bostwick Avenue, N.E., Grand Rapids, Michigan 49503, USA. [2]Department of Obstetrics & Gynecology, National University Hospital, Yong Loo Lin School of Medicine, Graduate School for Integrative Sciences & Engineering, National University of Singapore, Singapore 119074, Republic of Singapore. [3]Department of Botany and Plant Sciences, University of California at Riverside, Riverside, California 92521, USA. [4]Department of Biochemistry, Medical College of Wisconsin, Milwaukee, Wisconsin 53226, USA. [5]Center for Plant Stress Genomics and Technology, King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia. [6]State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, National Center for Plant Gene Research, Beijing 100101, China.
*These authors contributed equally to this work.

HAB1 in an ABA-dependent manner, and showed clear preference for the natural (+)-stereoisomer (Fig. 1a). PYL2 was also able to inhibit the phosphatase activity of HAB1 with an IC_{50} of 0.15 μM with (+)-ABA, but with an IC_{50} of 1.7 μM with the biologically less active (−)-isomer (Fig. 1b). This 11-fold difference in selectivity is comparable to the 19-fold selectivity shown for PYL5, another member of the PYR/PYL family[9]. The PYL1 and PYL2 structures were solved by molecular replacement starting from a model of the birch pollen allergen Bet VI (PDB code 1bv1)[14], with the statistics of the structure refinement shown in Supplementary Table 1.

The structures of apo- and ABA-bound PYL2 have an α-β-α2-β6-α topology and exhibit nearly identical helix-grip folds, which is a hallmark of the START protein superfamily[15] (Fig. 1c). A dominant feature of the structure is the long carboxy-terminal α-helix enfolded by an anti-parallel β-sheet of seven strands and α2 and α3 helices. The packing interactions of these elements form the hydrophobic ligand-binding cores of START proteins. Between the C-terminal helix and the β-sheet is a large pocket of 543 Å^3 in the apo structure (Fig. 1c), and 480 Å^3 in the ABA-bound structure (Fig. 1d). This pocket is surrounded by 23 residues, which are mostly hydrophobic and highly conserved (Fig. 1f). PYL1 and PYL2 share 51% sequence identity (Supplementary Table 2), and the apo-PYL1 structure resembles the apo-PYL2 structure (Fig. 1e and Supplementary Fig. 2b). The major differences between the apo-PYL1 and apo-PYL2 structures are seen in the relative position of their amino-terminal helices and their loop structure between β-strands 6 and 7, which is invisible in the apo-PYL1 structure. In contrast, the C-terminal α-helix and the β-strands surrounding the ligand binding pocket are highly similar.

The overall structure of the PYL2–ABA complex reveals a monomeric receptor–ligand complex with 1:1 stoichiometry. The binding mode of ABA is clearly defined by a high resolution electron density map (Fig. 2a and Supplementary Fig. 3a). ABA is centred in the pocket with its hydrocarbon chain and cyclohexene ring fitting snugly into the hydrophobic pocket of PYL2. The intermolecular interactions between ABA and PYL2 are shown in diagram form in Fig. 2b. Each of the three polar groups of ABA forms direct or water-mediated hydrogen bonds with the receptor. The acid head group forms a charged interaction with the side chain of K64 (Fig. 2a), and also forms a complex network of water-mediated hydrogen bonds with residues E98, N173 and E147 (Fig. 2a, b, and Supplementary Fig. 3a). The hydroxyl group from the chiral carbon also forms two water-mediated hydrogen bonds with E98 and N173. Mutations in these polar residues and several other pocket residues in PYL2 reduced its ability to bind to radio-labelled ABA, to interact with HAB1 in response to ABA, and to inhibit HAB1's PP2C activity (Fig. 2d–f).

The ABA-bound PYL2 structure also reveals the basis for its stereoselectivity. Figure 2c shows that the monomethyl group from the cyclohexene ring fits into a narrow pocket formed by residues F66, V87, L91, P92, F165 and V169, while the dimethyl group fits into a larger pocket formed by A93, S96 and V114. The dimethyl group flipped in the (−)-isomer would cause stereo collision between the dimethyl group and the narrow pocket that accommodates the

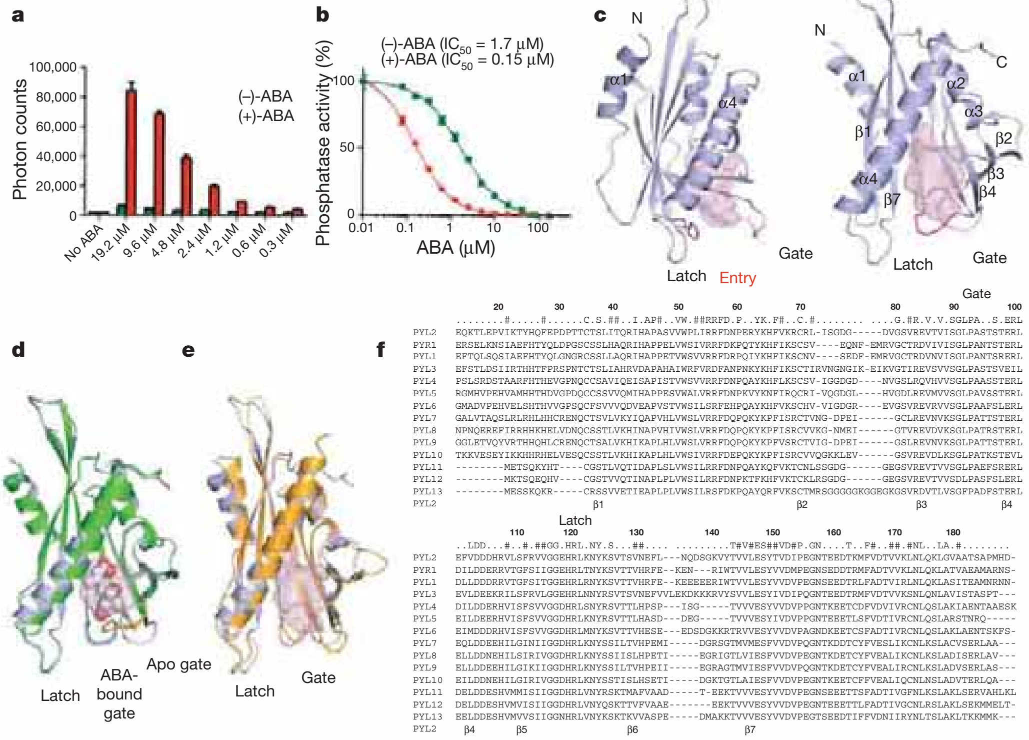


Figure 1 | Structures of apo-PYL2, ABA–PYL2 and apo-PYL1. **a**, ABA-dependent interaction of PYL2 with HAB1. Binding of recombinant H6GST–PYL2 to recombinant biotin–HAB1 at the indicated concentrations of (+)- and (−)-ABA was determined by AlphaScreen assays as described in Methods and in Supplementary Fig. 4a (*n* = 3; error bars, s.d.). **b**, ABA-dependent inhibition of HAB1 phosphatase activity by PYL2. Phosphate release from ^{P}S175 of 100 μM SnRK2.6 phosphopeptide was determined by colorimetric assay in the presence of recombinant HAB1 and PYL2 (*n* = 3; error bars, s.d.). The IC_{50} values were derived from curve fitting based on a competitive inhibitor model for the binding of ABA-bound PYL2 to HAB1 ($R^2 = 0.998$ for (+)-ABA and 0.997 for (−)-ABA). **c**, Two views of the apo-PYL2 structure with its empty ligand binding pocket shown as mesh. The gate (pink) and latch (magenta) for the ligand entry are indicated by arrows. **d**, Overlay of the apo (grey) and ABA-bound (green) PYL2 structures. ABA is shown as a ball model with its surrounding ligand binding pocket as mesh. The latch (magenta for apo- and cyan for ABA-bound PYL2) and the conformational change in the gate (pink and yellow in apo- and ABA-bound structures, respectively) are indicated by arrows. **e**, Overlay of the apo structures of PYL2 (grey) and PYL1 (yellow). **f**, Sequence alignment of the PYR/PYL family with secondary structure elements of PYL2 below the sequences and amino acid position of full length PYL2 above the alignment. Note that PYL2 in the crystal structure starts at position 14 relative to the full-length protein. Conserved residues are highlighted, residues within a 5 Å range of the ligand are boxed, and the gate and latch residues are noted.

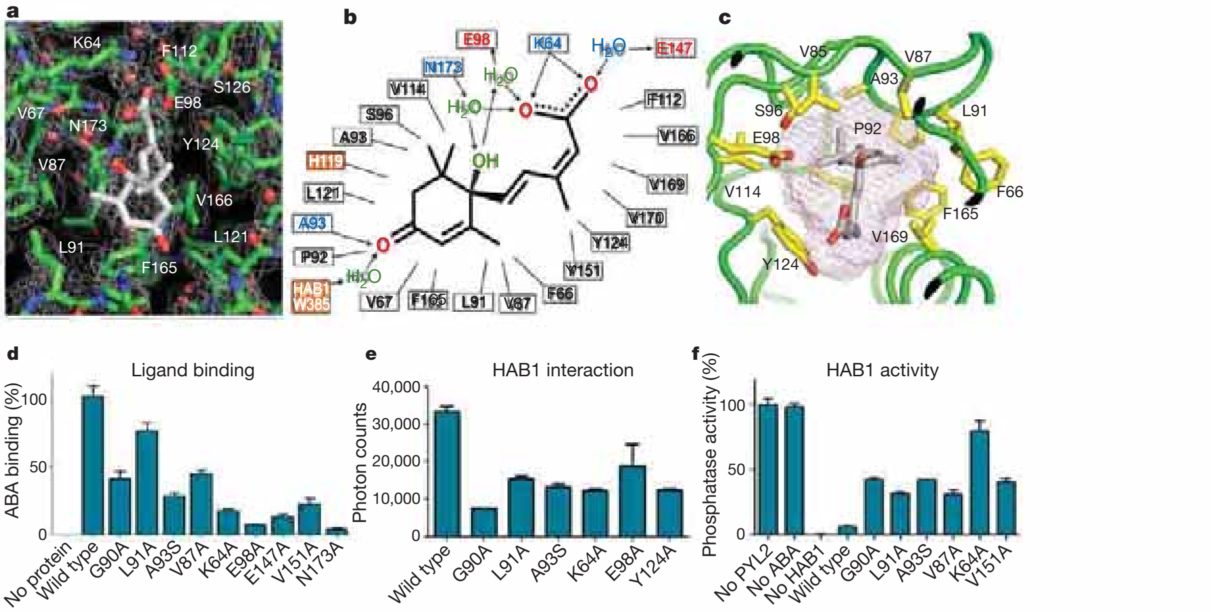


Figure 2 | Structural, functional and mutational analysis of the ABA binding pocket. **a**, A $2F_o - F_c$ electron density map of bound ABA and its surrounding residues contoured at 1.0σ. The charge interaction between the ABA acid group and K64 is indicated by two arrows. **b**, Schematic presentation of the interactions between PYL2 binding pocket residues and the bound (+)-ABA. Charged interactions and hydrogen bonds are indicated by arrows, hydrophobic interactions by solid lines with hydrogen-bond donors in blue and acceptors in red. The two orange-boxed residues, PYL2 H119 and HAB1 W385, contact ABA only on formation of the trimeric PYL2–ABA–HAB1 complex. **c**, The pocket topology of PYL2 surrounding the monomethyl and the dimethyl groups of ABA shown with key PYL2 residues and the ligand binding pocket (mesh). **d**–**f**, Mutations in key ligand binding pocket residues of PYL2 compromise **d**, ligand binding as determined by scintillation proximity assay with ^{3}H-labelled ABA ($n = 3$; error bars, s.d.), **e**, HAB1 interaction determined by AlphaScreen assays ($n = 3$; error bars, s.d.), and **f**, inhibition of HAB1 phosphatase activity ($n = 3$; error bars, s.d.).

monomethyl group (Fig. 2c). Thus, the size and shape of PYL2's pocket that accommodates the cyclohexene ring, together with the specific hydrogen bonds between PYL2 and ABA, contribute to its stereo-selectivity.

Comparison of the apo- and ABA-bound PYL2 structures suggests that a gate and latch mechanism is employed in hormone binding and signal transduction. The entry of the ligand-binding pocket is surrounded by a gate-like loop between β-strands 3 and 4 (residues 89–93: SGLPA) on one side, and on the other side, a latch-like region formed by a loop between β-strands 5 and 6 (residues 119–121: HRL) (Fig. 3a). The positions of H119 and L121 (latch) do not change substantially between apo- and ABA-bound structures. In contrast, ABA binding induces a dramatic conformational change in the SGLPA loop, which shifts 3–9 Å away from the corresponding loop of the apo-structure (Fig. 3a, b), a swing that closes the ligand entry gate and seals the bound ABA molecule from exposure to the solvent.

Outside the SGLPA gate, the most dramatic change of conformation occurs at residue E118, which is the residue immediately preceding the HRL latch motif (Figs 3a and 1f). In the apo-structure, the E118 side chain points into the pocket, a position that would prevent the closure of the SGLPA gate. In the ABA-bound structure, the E118 side chain is flipped by ~150°, so that it points outside the pocket. This creates the space required for the SGLPA gate to close onto the HRL latch, with the distance between LPA (gate) and HRL (latch) changing from 11–14 Å to 3.5–3.7 Å (Fig. 3a–c). On closing the gate, the SGLPA loop together with the HRL residues form a new interface, which we predicted could fit into the active site of the PP2C by using molecular modelling (see Methods and Supplementary Fig. 3b).

To validate our prediction and unravel the mechanism of PP2C inhibition by PYR/PYL proteins, we prepared a ternary complex for crystallization by mixing the purified proteins of PYL2 and HAB1 in the presence of (+)-ABA (Supplementary Fig. 1d). Because of the high binding affinity of PYL2 to HAB1 in the presence of (+)-ABA, (estimated to be 0.15 μM, see Supplementary Fig. 4), the ternary complex was stable and readily formed crystals (Supplementary Fig. 1c)

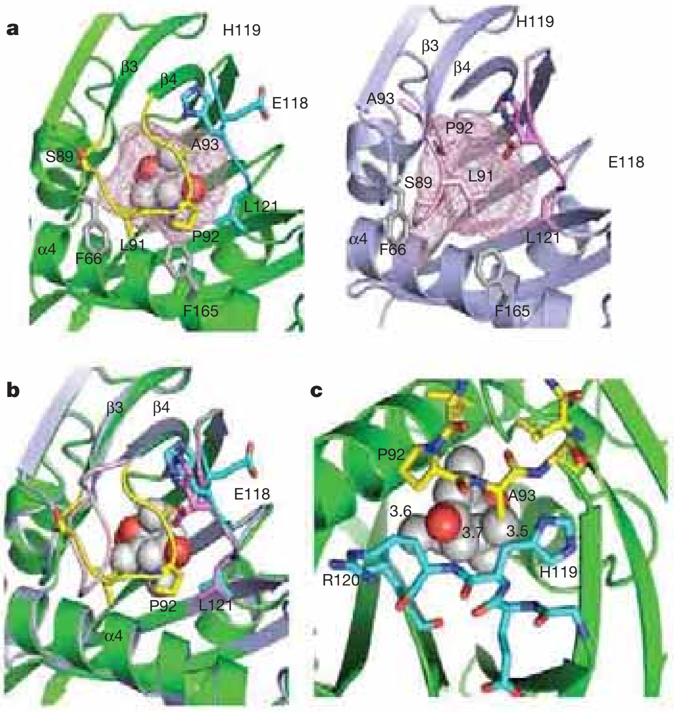


Figure 3 | A gate and latch mechanism seals ABA in the binding pocket. **a**, Close view of the ligand entry site in the apo- and ABA-bound PYL2 structure with the pocket shown as mesh. Side chains in the entry site are marked. The gate and latch are shown in yellow and cyan in the ABA-bound structure and in pink and magenta in the apo structure. **b**, Overlap of the ligand binding pocket in the ABA-bound and apo form with the gate and latch in the same colour code as in **a**. **c**, Interactions between gate (yellow) and latch (cyan) residues in the PYL2–ABA complex. Distances between residues are indicated in Å. ABA is shown as ball model.

that diffracted better than 1.95 Å. The structure was determined by molecular replacement with the ABA–PYL2 structure and a HAB1 model built from the PPM1B structure, a human PP2C[16]. The statistics of the structure refinement are shown in Supplementary Table 1.

The overall structure of the PYL2–ABA–HAB1 complex reveals a monomeric receptor–ligand–PP2C complex with 1:1:1 stoichiometry (Fig. 4a). The catalytic domain of HAB1 adopts a typical PP2C fold[17], containing two five-stranded β-sheets that are sandwiched by two pairs of anti-parallel α-helices. Between the two central β-sheets, there is a small domain of one short α-helix and two anti-parallel β-strands (residues 361–412) that forms the major part of the binding interface with PYL2 (Fig. 4a–c). The catalytic site, which is located at the top edge of the two central β-sheets with three Mg^{2+} ions, is partially occupied by the bound PYL2 (Fig. 4c, d).

The complex contains a large complementary interface of 1,820 Å^2 between the ABA-bound PYL2 and HAB1, whose intermolecular interactions are summarized in Supplementary Table 3. Consistent with our modelling, ABA-bound PYL2 docks into the catalytic active site of HAB1 (Fig. 4d), with the SGLPA gate packing closely with residues D243 to G246, which forms the active site loop of the PP2C. The G246D mutation in HAB1 and the corresponding residues in other group A PP2Cs are well documented to cause dominant ABA-insensitivity[18]. The distance of G246 to the SGLPA gate is only 2.8–4.0 Å, which probably explains why G246D can disrupt the PP2C/PYR1 interaction[7–9,19].

Importantly, the ternary complex structure reveals several critical insights of ABA binding and signal transduction that were not predicted by our modelling studies or apparent from the PYL2–ABA complex. The most striking observation is that the indole ring of W385 of HAB1 inserts between SGPLA gate and the HRL latch. In so doing, it points into the ABA-binding pocket and makes a water-mediated hydrogen bond to the ketone group of ABA (Fig. 4e, f). This observation therefore provides direct proof that HAB1 can serve as a co-receptor that senses the binding of ABA into the PYR/PYL receptors.

In response to the insertion of W385 of HAB1 into the ABA-binding pocket of PYL2, the SGPLA gate and the HRL latch undergo several unexpected conformational changes between the PYL2–ABA and ternary PYL2–ABA–HAB1 structures (Fig. 4g). The most drastic changes are seen in the latch residues E118 and H119, whose Cα atoms move 4.5 Å and 1.9 Å, respectively. In both apo and ABA-bound PYL2, the side chain of H119 is outside of the pocket. However, in the HAB1 complex, PYL2's H119 side chain points into the pocket and makes contacts with the dimethyl groups of ABA's cyclohexene ring. Correspondingly, the position of ABA's cyclohexene ring shifts 1.2–1.4 Å, allowing its ketone group to face directly towards the water molecule that makes a hydrogen bond with W385 of HAB1(Fig. 4e, f). In this configuration, W385 of HAB1 functions like a molecular lock that keeps the gate and the latch in the closed positions. This locking mechanism probably explains the higher ABA binding affinities measured for PYR/PYLs in the presence of PP2Cs[7,8]. Thus, PP2C binding modifies the PYL2–ABA contact surface and creates a water-mediated contact between ABA and PP2C.

The importance of the gate and latch loops is supported by extensive mutational data (Fig. 2d–f). Mutations in both the gate and latch compromised ABA-dependent HAB1 interaction and PP2C activity (H119A and R120A in Fig. 5a, b). Additionally, it has been noted that mutations in the PP2C active site (mutations that cause the dominant ABA-insensitive phenotype, for example, G180D in ABI1, G168D in

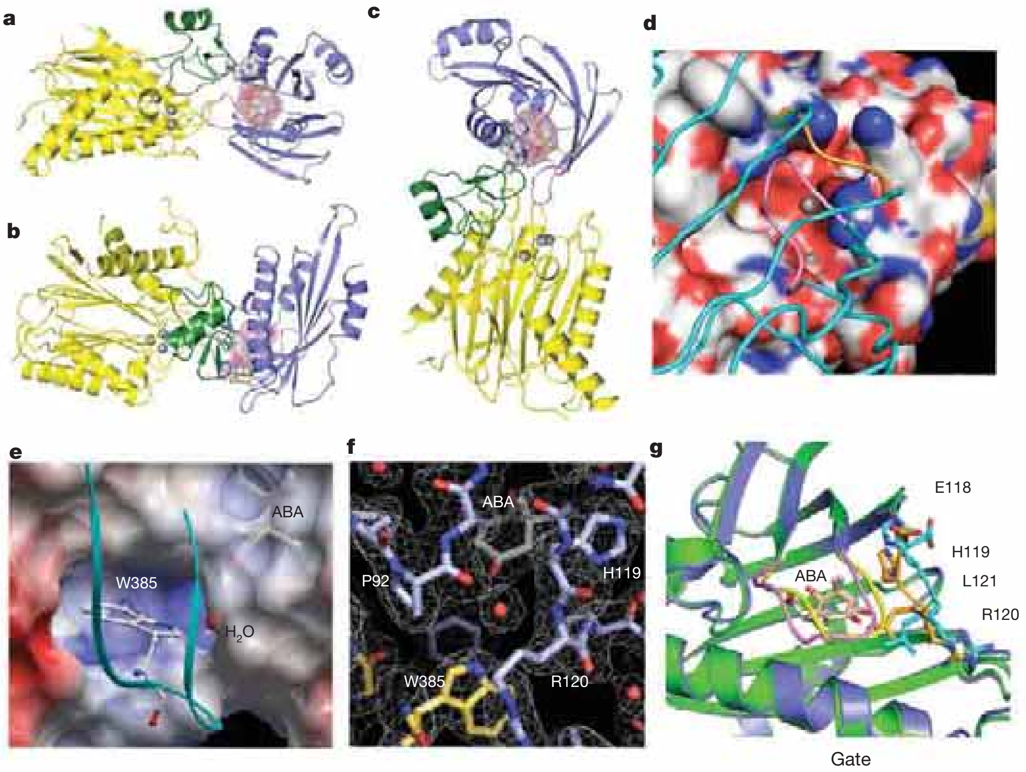


Figure 4 | Structure of the PYL2–ABA–HAB1 co-receptor complex. **a–c**, Three views of the PYL2–ABA–HAB1 complex. HAB1 is shown with its catalytic domain in yellow (Mg^{2+} ions as balls) and the ABA-interacting domain in green (W385 lock as a ball model). PYL2 is shown in blue and ABA as a ball model with its surrounding ligand binding pocket as mesh. **d**, HAB1 active site as space-filling presentation with Mg^{2+} ions as balls and the docking PYL2 shown as carbon backbone (gate in pink and latch in gold colour). **e**, HAB1 W385 docking into a PYL2 cavity enables it to make a water-mediated interaction with the ketone group of ABA. The loop containing W385 is shown as ribbon in cyan. **f**, Network of water-mediated interactions with the ketone group of ABA in the PYL2–ABA–HAB1 complex. ABA and key residues are shown with the surrounding electron density map ($2F_o - F_c$) contoured at 1σ, W385 of HAB1 is highlighted in yellow. **g**, HAB1 docking induces conformational changes in the PYL2 gate and latch. Close-up overlay view of PYL2–ABA in the presence (blue) and absence (green) of HAB1. The latch is shown in cyan (−HAB1) and gold (+HAB1), the gate in yellow (−HAB1) and pink (+HAB1), and ABA in grey (−HAB1) and peach (+HAB1).

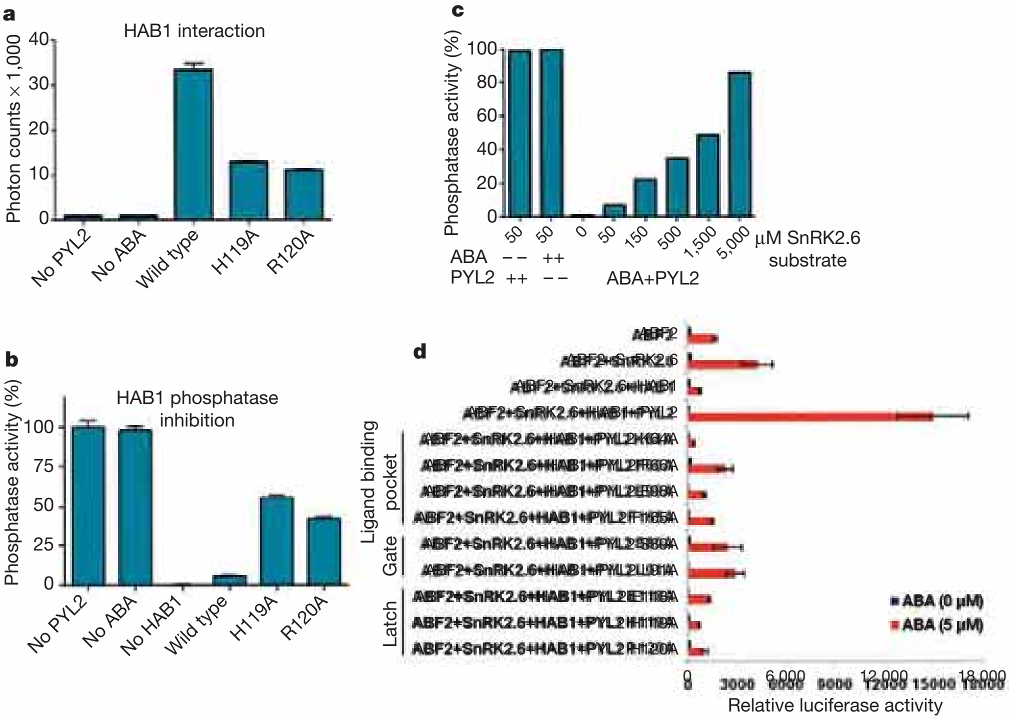


Figure 5 | Mutational analysis of PYL2 and HAB1 interface residues. **a**, **b**, Functional analysis of mutations in PYL2 latch residues by **a**, HAB1 interaction as determined by AlphaScreen ($n = 3$; error bars, s.d.), and by **b**, HAB1 phosphatase activity ($n = 3$; error bars, s.d.). **c**, Increasing concentrations of the SnRK2.6 substrate overcomes HAB1 inhibition by ABA-bound PYL2. Uninhibited HAB1 phosphatase activity (absence of either ABA or PYL2) is shown as control. **d**, Reconstituted ABA signalling pathway for gene expression in *Arabidopsis* mesophyll protoplasts. PYL2, HAB1, SnRK2.6 and ABF2 expression vectors were cotransfected into protoplasts together with a luciferase reporter plasmid driven from the promoter of the ABA-responsive *RD29B* gene. Mutations in the PYL2 ligand binding pocket (K64A, F66A, E98A, F165A), the SGLPA gate (S89A and L91S) and the latch (E118A, H119A, R120A) are defective in the reconstituted signalling pathway ($n = 3$; error bars, s.e.m.).

ABI2, and G246D in HAB1) disrupt PP2C's interaction with PYR1 (refs 7–9, 19), which is consistent with our model, as the larger side chain of the mutated residues would collide with the SGPLA gate (Supplementary Fig. 3c). Furthermore, the direct docking of PYL2's SGPLA loop into the active site of HAB1 indicates that the ABA-bound PYL2 functions as a competitive inhibitor of PP2Cs. To test this prediction, we investigated whether increasing concentrations of a PP2C substrate, a SnRK2.6 peptide containing residues 170–180 and a phosphorylated serine at position 175, could overcome the inhibition of HAB1 PP2C activity by ABA-bound PYL2 (Fig. 5c). SnRK2.6 is a natural substrate of PP2C, and phosphorylation at S175 is required for SnRK2.6 kinase activation and its role in ABA signalling[5]. Our results are consistent with a competitive inhibitor mechanism between ABA-bound PYL2 and the PP2C active site (Figs 4d and 5c). Thus, our structures illuminate both the mode of ligand binding and PP2C inhibition.

To validate the functional roles of the PYL2/HAB1 interface in ABA signalling, we mutated key residues that form the gate and latch of PYL2, and analysed their function in protoplasts[20], where ABA-induction of gene expression can be reconstituted by coexpression of a PYR/PYL protein, PP2Cs, SnRK2 kinases, the ABF2 transcription factor and an ABA-responsive promoter reporter construct[11]. Using this system, ABA induces reporter gene activation in the presence of PYL2, HAB1, SnRK2.6 and ABF2 (Fig. 5d). However, mutations in PYL2 residues involving ligand binding or formation of the gate and latch compromise the ability of PYL2 to activate the reporter in response to ABA. These mutations also reduced the ability of PYL2 function through ABI1 (Supplementary Fig. 5). Thus, our mutational and structural data indicate that the gate and latch residues involved in conformational changes are critical components of the information transfer mechanism.

The 14 members of the PYR/PYL family share 38–84% sequence identity over the entire length of the proteins (Supplementary Table 2), but the SGLPA and HRL sequences are invariant among the PYR/PYL family members (Fig. 1f), suggesting that the gate and latch mechanism is likely to be a common feature of these receptors. To further investigate this mechanism and ascertain if it is used by other family members, we studied the consequences of ABA binding on PYR1 using the HSQC (heteronuclear single quantum coherence) NMR technique, which is a complementary method because it enables measurements of dynamics directly in solution. PYR1 was chosen as a model because its HSQC NMR spectra have been previously investigated[7]. ABA binding induced either exchange broadening or chemical shift perturbation in residues surrounding the ligand binding pocket (Supplementary Fig. 6a). In contrast, the vast majority (66 of 71, 93%) of unperturbed residues in the PYR1 spectrum are located outside the ABA binding pocket (Supplementary Fig. 6b). Notably, the SGLPA gate and the HRL latch residues exhibit broadening or large shift perturbation, consistent with a large conformational change on ligand binding (Supplementary Fig. 6c, d). Thus, the chemical shift perturbations induced by ABA binding to PYR1, collected by solution NMR, correlate exceedingly well with the molecular interactions observed in the crystal structures of apo- and ABA-bound PYL2. These observations strongly argue that conformational dynamics accompany ligand binding in multiple PYR/PYL family members, and strengthen our conclusion that they are central to the mechanism of signal transduction.

We also performed transgenic plant studies of PYR1 with a mutated latch residue (H115A). The homologous residue (H119) in PYL2 is pointed outward to the solvent in both the apo- and ABA-bound PYL2 structures (Fig. 3a, b), thus the H115A mutation is predicted not to interfere with ABA binding in the absence of HAB1. Indeed, NMR spectra of ^{15}N-labelled PYR1H115A show that the mutated protein forms a stably folded protein that responds to ABA similarly to the wild-type PYR1 (Fig. 6a), indicating that the H115A PYR1 has comparable ligand binding capability to the wild type in the absence of HAB1. However, consistent with the important

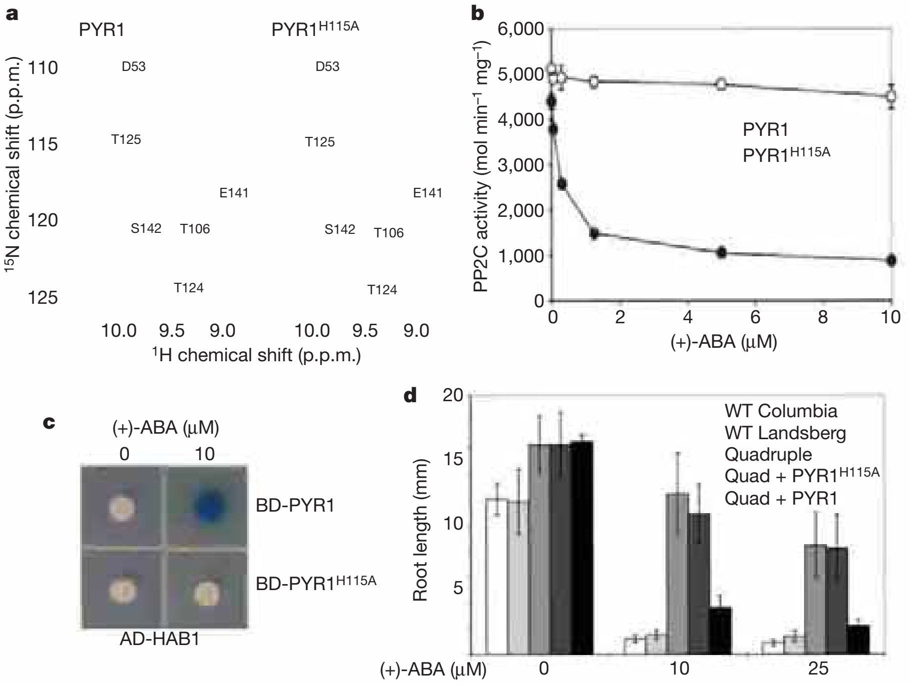


Figure 6 | Mutations in the PYR1 latch and gate affect ABA signalling *in vitro* and *in vivo*. a, ABA binding is unaltered by the H115A mutation in the PYR1 latch. Comparisons of ^{1}H–^{15}N HSQC spectra for PYR1 (left) and PYR1(H115A) (right) in the presence (red contours) and absence (black contours) of (+)-ABA indicate that residues surrounding the binding site observed in the ABA–PYL2 crystal structure are similarly perturbed in both proteins. **b**, HAB1 phosphatase activity is abolished in H115A PYR1 ($n = 3$, error bars, s.d.). **c**, The H115A PYR1 mutant is defective in ABA-mediated HAB1 interaction, as shown by yeast two-hybrid assay. AD, activation domain; BD, DNA binding domain. **d**, The H115A PYR1 mutant is defective in inhibiting the root growth of transgenic plants in response to ABA. A quadruple *pyr1/pyl1/pyl2/pyl4* mutant (Q) defective in root growth inhibition can be complemented by wild-type *PYR1*, but not by *PYR1 (H115A)* ($n = 6$–10; error bars, s.d.).

role of this latch residue in ABA binding in the PYL2/PP2C complex, the H115A PYR1 protein was defective in ABA-mediated PP2C inhibition *in vitro* (Fig. 6b) and was not able to interact with HAB1 in response to (+)-ABA (Fig. 6c). Furthermore, H115A PYR1 failed to rescue the ABA-response defect of a quadruple *pyr1/pyl1/pyl2/pyl4* mutant in transgenic plants, whereas the wild-type PYR1 did (Fig. 6d). Together, these results demonstrate that the latch residue H115 is not important for ligand binding in the context of PYR1 itself, but is critical for the receptor to relay the ABA binding signal to PP2Cs and downstream signalling effectors.

In this paper, we present four crystal structures involving two ABA receptors, PYL1 and PYL2, in apo-forms, and the ABA-bound forms of PYL2, either alone or in complex with the HAB1 PP2C. In combination with biochemical, mutagenesis, NMR, and transgenic plant studies, our structures highlight a gate–latch–lock mechanism of ABA binding and signalling by the PYR/PYL family and PP2Cs. The ternary complex structure reveals that PP2Cs can function as co-receptors to facilitate the interaction of PYR/PYL proteins with ABA, and establish that these proteins together are bona fide ABA receptors that transduce plant stress signals. Elucidation of the structures and mechanisms of these core components of ABA signalling pathways will help construct a rational framework for understanding plant stress biology as well as for engineering plants that are resistant to salt and drought environments.

The START domain was originally identified in proteins involved in steroidogenesis and metabolism in metazoans[21–23]. Although members of the START domain family continue to be identified[15,24], the cellular functions of most START proteins still remain elusive beyond the established roles in transfer of lipids and cholesterol for a few members[25]. Identification of the gate–latch–lock mechanism for hormone binding and signalling by ABA receptors suggests that other members in the large family of START domain proteins may also function as ligand sensing and signal transduction molecules like ABA receptors.

METHODS SUMMARY

PYL2, PYL1 and HAB1 were expressed as 6×His-GST (H6GST) or H6SUMO fusion proteins in *E. coli*. Proteins were purified by Ni-NTA chromatography, followed by proteolytic release of tags and size-exclusion chromatography. For formation of PYL2–ABA and HAB1–PYL2–ABA complexes, ABA was mixed with PYL2 and HAB1–PYL2 at 5:1 ratios. Crystals were grown by vapour diffusion and diffraction data were collected from cryo-protected crystals at beamlines 21-ID-D and 21-ID-F at the Advanced Photon Source at Argonne National Laboratories. Structures were solved by molecular replacement in PHASER[26] using the structure of the plant START protein Bet v 1 as model for PYL2 and the structure of the human PP2C PPM1B as model for HAB1. Models were manually fitted using O and Coot[27,28] and further refined using CNS and Refmac5[29,30].

Mutant proteins were expressed as H6GST-fusion proteins and purified by glutathione sepharose chromatography. Protein–protein interactions were determined by luminescence proximity AlphaScreen assay and by yeast two-hybrid assay. Biotinylated HAB1 for the luminescence proximity assay was generated by *in vivo* biotinylation of an avitag–HAB1 fusion protein. ABA binding was determined by scintillation proximity assay using ^{3}H-labelled ABA. HAB1 phosphatase activity was measured by phosphate release from a SnRK2.6 phosphoprotein (Figs 1–5) or from a generic pNPP phosphate substrate (Fig. 6b).

For transgenic studies, wildtype and mutant *35S::GFP–PYR1* constructs were transformed by the floral dip method into *pyr1/pyl1/pyl2/pyl4* quadruple mutants. Mutant complementation of GFP^{+} seedlings was assayed by root length measurements. The ABA signal transduction pathway was reconstituted in protoplasts by transient transfection of *PYL2*, *PP2C*, *SnRK2.6* and *ABF2* expression plasmids. Activation of an ABA-inducible *RD29B* promoter-*LUC* reporter by PYL2 mutant proteins was determined by luciferase assays normalized for β-glucuronidase activity from a *UQ10–GUS* reporter.

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

Received 16 October; accepted 27 October 2009.

1. Yoshida, T. *et al.* ABA-hypersensitive germination3 encodes a protein phosphatase 2C (AtPP2CA) that strongly regulates abscisic acid signaling during germination among *Arabidopsis* protein phosphatase 2Cs. *Plant Physiol.* **140**, 115–126 (2006).

2. Mustilli, A. C., Merlot, S., Vavasseur, A., Fenzi, F. & Giraudat, J. *Arabidopsis* OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. *Plant Cell* **14,** 3089–3099 (2002).
3. Chinnusamy, V., Gong, Z. & Zhu, J. K. Abscisic acid-mediated epigenetic processes in plant development and stress responses. *J. Integr. Plant Biol.* **50,** 1187–1195 (2008).
4. Yoshida, R. *et al.* ABA-activated SnRK2 protein kinase is required for dehydration stress signaling in *Arabidopsis. Plant Cell Physiol.* **43,** 1473–1483 (2002).
5. Fujii, H., Verslues, P. E. & Zhu, J. K. Identification of two protein kinases required for abscisic acid regulation of seed germination, root growth, and gene expression in *Arabidopsis. Plant Cell* **19,** 485–494 (2007).
6. Pennisi, E. Plant biology. Stressed out over a stress hormone. *Science* **324,** 1012–1013 (2009).
7. Park, S. Y. *et al.* Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324,** 1068–1071 (2009).
8. Ma, Y. *et al.* Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324,** 1064–1068 (2009).
9. Santiago, J. *et al.* Modulation of drought resistance by the abscisic acid receptor PYL5 through inhibition of clade A PP2Cs. *Plant J.* doi:10.1111/j.1365-313X.2009.03981.x (published online 16 July 2009).
10. Kobayashi, Y. *et al.* Abscisic acid-activated SNRK2 protein kinases function in the gene-regulation pathway of ABA signal transduction by phosphorylating ABA response element-binding factors. *Plant J.* **44,** 939–949 (2005).
11. Furihata, T. *et al.* Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. *Proc. Natl Acad. Sci. USA* **103,** 1988–1993 (2006).
12. Zhu, J. K. Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* **53,** 247–273 (2002).
13. Yamaguchi-Shinozaki, K. & Shinozaki, K. Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.* **57,** 781–803 (2006).
14. Gajhede, M. *et al.* X-ray and NMR structure of Bet v 1, the origin of birch pollen allergy. *Nature Struct. Biol.* **3,** 1040–1045 (1996).
15. Iyer, L. M., Koonin, E. V. & Aravind, L. Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. *Proteins* **43,** 134–144 (2001).
16. Almo, S. C. *et al.* Structural genomics of protein phosphatases. *J. Struct. Funct. Genomics* **8,** 121–140 (2007).
17. Das, A. K., Helps, N. R., Cohen, P. T. & Barford, D. Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 Å resolution. *EMBO J.* **15,** 6798–6809 (1996).
18. Bertauche, N., Leung, J. & Giraudat, J. Protein phosphatase activity of abscisic acid insensitive 1 (ABI1) protein from *Arabidopsis thaliana. Eur. J. Biochem.* **241,** 193–200 (1996).
19. Sheen, J. Mutational analysis of protein phosphatase 2C involved in abscisic acid signal transduction in higher plants. *Proc. Natl Acad. Sci. USA* **95,** 975–980 (1998).
20. Yoo, S. D., Cho, Y. H. & Sheen, J. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols* **2,** 1565–1572 (2007).
21. Allen-Baume, V., Segui, B. & Cockcroft, S. Current thoughts on the phosphatidylinositol transfer protein family. *FEBS Lett.* **531,** 74–80 (2002).
22. Seedorf, U., Ellinghaus, P. & Roch Nofer, J. Sterol carrier protein-2. *Biochim. Biophys. Acta* **1486,** 45–54 (2000).
23. Wirtz, K. W. Phospholipid transfer proteins in perspective. *FEBS Lett.* **580,** 5436–5441 (2006).
24. Schrick, K., Nguyen, D., Karlowski, W. M. & Mayer, K. F. START lipid/sterol-binding domains are amplified in plants and are predominantly associated with homeodomain transcription factors. *Genome Biol.* **5,** R41 (2004).
25. Olayioye, M. A. *et al.* StarD10, a START domain protein overexpressed in breast cancer, functions as a phospholipid transfer protein. *J. Biol. Chem.* **280,** 27436–27442 (2005).
26. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
27. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
28. Kleywegt, G. J. & Jones, T. A. Efficient rebuilding of protein structures. *Acta Crystallogr. D* **52,** 829–832 (1996).
29. Brunger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54,** 905–921 (1998).
30. Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D* **55,** 247–255 (1999).

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

Acknowledgements We thank the staff of LS-CAT for assistance in data collection at the beam lines of sector 21, which is in part funded by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor. Use of the Advanced Photon Source was supported by the Office of Science of the US Department of Energy. This work was supported by the Jay and Betty Van Andel Foundation (H.E.X.), the National Institutes of Health (H.E.X., B.F.V. and J-K. Z.), and the National Science Foundation (S.R.C.). L.-M.N. and F.-F.S. were supported by an overseas PhD scholarship from the NUS Graduate School for Integrative Sciences & Engineering (NGS).

Author Contributions K.M., Jun Li, Jiayang Li, E.-L.Y., B.F.V., S.R.C., J.-K.Z. and H.E.X. conceived the project and designed research; K.M., L.-M.N., X.E.Z., F.-F.S., Y.X., K.M.S.-P., S.-Y.P., J.J.W., H.F., V.C., A.K., Y.W., F.C.P., D.R.J. and H.E.X. performed research; K.M., L.-M.N., X.E.Z., F.-F.S., Y.X., K.M.S.-P., S.-Y.P., J.J.W., H.F., V.C., B.F.V., S.R.C., J.-K.Z. and H.E.X. analysed data; and K.M., S.R.C. and H.E.X. wrote the paper with contributions from all authors.

Author Information The structure factors and atomic coordinates discussed in this work have been deposited in the Protein Data Bank. The accession codes are: 3KAY for apo-PYL1; 3KAZ for apo-PYL2; 3KB0 for the ABA-bound PYL2 complex; and 3KB3 for the ternary complex of PYL2–ABA–HAB1. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to H.E.X. (eric.xu@vai.org).

METHODS

Protein preparation. PYL2 (residues 14–188) was expressed as a 6 × His-GST (H6GST) fusion protein from the expression vector pET24a (Novagen). The modified fusion protein contains a H6-tag (MKKGHHHHHHG) at the N terminus and a thrombin protease site between GST and PYL2. BL21(DE3) cells transformed with the expression plasmid were grown in LB broth at 16 °C to an A_{600} of ~1.0 and induced with 0.1 mM IPTG for 16 h. Cells were harvested, resuspended in 100 ml extract buffer (20 mM Tris, pH 8.0, 200 mM NaCl and 10% glycerol) per 6 l of cells, and passed three times through a French press with pressure set at 1,000 Pa. The lysate was centrifuged at 16,000 r.p.m. (30,800*g*) in a Sorvall SS34 rotor for 30 min, and the supernatant was loaded on a 50 ml Nickel HP column. The column was washed with 10% buffer B (20 mM Tris, pH 8.0, 200 mM NaCl, 500 mM imidazole and 10% glycerol) for 600 ml and eluted in two steps with 50% buffer B for 200 ml, then 100% buffer B for 100 ml. The eluted H6GST–PYL2 was dialysed against extract buffer and cleaved overnight with thrombin at a protease/protein ratio of 1:250 in the cold room. The cleaved H6GST tag was removed by passing through a Nickel HP column, and the protein was further purified by chromatography through a HiLoad 26/60 Superdex 200 gel filtration column in 25 mM Tris, pH 8.0, 200 mM ammonium acetate, 1 mM dithiothreitol and 1 mM EDTA. Apo-PYL2 eluted as a sharp single peak with an estimated molecular mass of 23 kDa, suggesting that apo-PYL2 is a monomer in solution (predicted molecular mass of 21 kDa).

PYL1 (residues 36–211) was expressed as a H6SUMO fusion protein from the expression vector pSUMO (LifeSensors). The expression and purification of PYL1 followed the same method as for PYL2 except that the H6SUMO tag was cleaved overnight with SUMO protease at a protease/protein ratio of 1:1,000 in the cold room. A typical yield of the purified protein is ~6 mg for PYL1 or ~1 mg for PYL2 from each litre of cells. The apo-PYL1 is eluted from the gel filtration column as a broad peak that corresponds to an estimated molecular mass of 31 kDa, suggesting the apo-PYL1 may exhibit a monomer–dimer equilibrium in solution. To prepare the protein–ligand complex, we added (+)-ABA to the purified proteins at a 5:1 molar ratio.

HAB1 (residues 172–511) was expressed as a H6GST fusion protein from the expression vector pET24a (Novagen). Expression and purification of HAB1 followed the same method as for PYL2 except that all buffers contained 5 mM $MgCl_2$. A typical yield of the purified protein was about 5 mg from each litre of cells. To prepare the protein–ligand–PP2C ternary complex, we added (+)-ABA and purified PYL2 to purified HAB1 at a 5:1:1 molar ratio in the presence of 5 mM $MgCl_2$.

Small scale purifications of H6GST-tagged PYL/PYR proteins for expression screening (Supplementary Fig. 1a) and for binding studies of wild-type and mutant proteins were performed by standard glutathione sepharose chromatography. To generate biotinylated proteins for luminescence proximity assays (AlphaScreen), PP2C open reading frames (ORFs) were expressed in *E. coli* BL21(DE3) cells from a pETDuet (Novagen) derivative vector. The first T7 polymerase-driven expression unit of this vector contains the PP2C ORF as H6-thioredoxin-thrombin cleavage site-avitag fusion, the second site the *E. coli* biotin-ligase gene *BirA*. The 14 amino acid avitag functions as a defined *in vivo* biotinylation site in *E. coli*[31]. Cells grown in the presence of 40 μM biotin were lysed and fusion protein purified over Nickel HiTrap columns as above. Following thrombin release of the H6-thioredoxin tag, PP2C proteins with biotinylated avitags were purified over monomeric avidin columns (Pierce) according to the manufacturer's instructions. Residual biotin was removed by extensive dialysis before their use in luminescence proximity assays.

Crystallization. The apo-PYL1 crystals were grown at room temperature in hanging drops containing 1.0 μl of the purified PYL1 protein at 9.1 mg ml^{-1} and 1.0 μl of well solution containing 0.1 M ammonium sulphate, 0.1 M Na-HEPES, pH 7.5, 10% w/v PEG 4000 and 30% glycerol. Crystals of about 100 μm in length appeared within 1–2 days. Crystals were flash frozen in liquid nitrogen before data collection. The ABA–PYL1 complex did not yield crystals.

Apo-PYL2 crystals were grown at room temperature in hanging drops containing 1.0 μl of the purified PYL2 protein at a concentration of 6.6 mg ml^{-1} and 1.0 μl of well solution containing 0.056 M sodium phosphate monobasic monohydrate/ 1.344 M potassium phosphate dibasic, pH 8.2, and 5% butanediol. Crystals appeared within 1–2 days and grew to a dimension of 200–250 μm in length over a period of 4 days. Crystals were serially transferred to well buffer with increasing glycerol concentration (20% v/v final) before flash freezing in liquid nitrogen.

The PYL2–ABA complex crystals were grown at room temperature in hanging drops containing 1.0 μl of the above protein–ligand solutions at a concentration of 6.6 mg ml^{-1}, and 1.0 μl of well solution containing 0.1 M HEPES, pH 7.5, 2 M ammonium sulphate and 20% sorbitol. Crystals of about 250 μm in length appeared the following day. Crystals were soaked in well solution with a final concentration of 30% sorbitol before flash freezing in liquid nitrogen.

The PYL2–ABA–HAB1 complex crystals were grown at room temperature in hanging drops containing 1.0 μl of the above protein–ligand–PP2C solutions and 1.0 μl of well solution containing 0.2 M ammonium sulphate, 0.1 M BIS-Tris pH 6.5, and 15% PEG 3350. Crystals appeared within 1–2 days and grew to a dimension of 450–600 μm in length on the third day. Crystals were serially transferred to well buffer with increasing PEG3350 concentration (35% v/v final) before flash freezing in liquid nitrogen.

Data collection and structure determination. The data sets were collected with MAR300 and MAR225 CCD detectors (MAR Research) at the ID-D and ID-F lines of sector-21 (LS-CAT) at the Advanced Photon Source at Argonne National Laboratory. The observed reflections were reduced, merged, and scaled with DENZO and SCALEPACK in the HKL2000 package[32]. Crystals of apo-PYL1, apo-PYL2, ABA–PYL2, and ABA–PYL2–HAB1 diffracted to resolutions of 2.40 Å, 1.85 Å, 1.95 Å and 1.95 Å, respectively. The apo-PYL1 crystals formed in the $P6_5$ space group with two receptors in each asymmetry unit. The weak dimer interface observed in the apo-PYL1 crystals is consistent with monomer–dimer behaviour of the protein in solution. The apo-PYL2 crystals formed in the $C222_1$ space group with three receptors in each asymmetry unit. The ABA-bound PYL2 crystals formed in the $P6_122$ space group with one receptor in each asymmetry unit and the ABA–PYL2–HAB1 crystals formed in the $P2_12_12_1$ space group with one complex in each asymmetric unit. The intermolecular packing in both apo- and ABA-bound crystals is less extensive than the apo-PYL1 crystals, consistent with the monomeric behaviour of the protein in solution.

Molecular replacement was performed by using the Collaborative Computational Project 4 (CCP4) program Phaser[26]. Programs O and Coot were used to manually fit the protein model[27,28]. Model refinement was performed with CNS and the CCP4 program Refmac5[29,30]. The volumes of the ligand binding pocket were calculated with the program Voidoo by using program default parameters and a probe with a radius of 1.4 Å (ref. 33). All structure figures were prepared by using PyMOL (DeLano Scientific). The statistics of data collection and the model refinement are summarized in Supplementary Table 1.

Molecular modelling. The PYL2 model was built with the Prime module (Schrödinger) using the structure of Bet V I, a birch pollen allergen that contains a START domain (PDB code: 1bv1), and this PYL2 model was used as template for molecular replacement. The HAB1 model was built with the Prime module (Schrödinger) using the structure of PPM1B, a human PP2C that shares 38% homology with the HAB1 PP2C core domain (PDB code: 2P8E)[16]. The HAB1–PYL2–ABA complex model was built with ZDOCK[34] via the protein–protein docking web server (http://zdock.bu.edu/) by uploading the crystal structure of the ABA-bound PYL2 and the homology model of HAB1. The top docking models were retrieved from this server and visually inspected. Structure-based sequence alignments were performed with the ICM program package (Molsoft).

Assays for the interactions between PYR/PYL and PP2Cs. Interactions between PYR/PYL and PP2Cs were assessed by luminescence-based AlphaScreen technology (Perkin Elmer) that our group has used extensively to determine ligand-dependent protein–protein interactions for nuclear receptors[35–37]. The scheme of AlphaScreen is illustrated in Supplementary Fig. 4a. Briefly, biotinylated PP2Cs were attached to streptavidin-coated donor beads, and H6-tagged PYR/PYL receptors were attached to nickel-chelated acceptor beads. The donor and acceptor beads were brought into proximity by the interactions between PYR/PYLs and PP2Cs, which were measured with and without abscisic acid, the (−)-isoform (Sigma) or the (+)-isoform (A.G. Scientific). When excited by a laser beam of 680 nm, the donor beam emits singlet oxygen that activates thioxene derivatives in the acceptor beads, which releases photons of 520–620 nm as the binding signal. The experiments were conducted with 100 nM of PP2Cs and PYR/PYL proteins in the presence of 5 μg ml^{-1} donor and acceptor beads in a buffer of 50 mM MOPS, pH 7.4, 50 mM NaF, 50 mM CHAPS, and 0.1 mg ml^{-1} bovine serum albumin. The results were based on an average of three experiments with standard errors typically less than 10% of the measurements.

Radio-ligand binding assay. Two micromolar H6GST–PYL2 were incubated with 250 μg of yttrium silicate copper-chelating scintillation proximity assay (SPA) beads (GE Healthcare) in a buffer of 50 mM MOPS, pH 7.4, 50 mM NaF, 50 mM CHAPS, and 0.1 mg ml^{-1} bovine serum albumin for 40 min shaking on ice. H6GST–PYL2 bound to SPA was separated from free H6GST–PYL2 by centrifugation at 5,200*g* for 30 s. Bead pellets were washed with 1 ml of the same buffer, then resuspended in 50 μl of the buffer supplemented with 45 nM ^{3}H-labelled ABA (GE Healthcare; mixture of (+)-/(−)- and *cis*-/*trans*-isomers), 1 μM unlabelled (+)-ABA and 10 μM HAB1 and incubated shaking for 1 h at room temperature. ^{3}H-ABA–PYL2 binding brings the radioactive ABA into the immediate proximity of the scintillant embedded in the SPA beads resulting in the generation of light, which was quantified in a liquid scintillation counter.

Assays of HAB1 phosphatase activity. Biotin–HAB1 (100 nM) and 500 nM PYL2 were pre-incubated in 50 mM imidazole, pH 7.2, 5 mM $MgCl_2$, 0.1% β-mercaptoethanol and 0.5 μg ml^{-1} BSA for 30 min at room temperature.

Reactions were started by addition of 100 µM of a phosphopeptide corresponding to amino acids 170–180 of SnRK2.6 (HSQPK^{P}STVGTP). This peptide is phosphorylated at a single residue corresponding to Ser 175 in SnRK2.6, whose phosphorylation is required for SnRK2.6 kinase activity[2]. Phosphate release from ^{P}S175 from the phosphopeptide was determined by colorimetric assay (BioVision). Phosphatase inhibition assays in Fig. 6b were performed using the generic pNPP phosphatase substrate, as described previously[7], with the minor modification that the assay buffer used 10 mM Mn^{2+} instead of Mg^{2+}, which we found to enhance HAB1's specific activity ~10-fold.

Mutagenesis. Site-directed mutagenesis was carried out using the QuickChange method (Stratagene) or the GeneTailor System (Invitrogen). Mutations and all plasmid constructs were confirmed by sequencing before both protein expression and yeast hybrid assays. Expression and purification of GST–HAB1, H6PYR1 and H6PYR1H115A were performed as described previously[7].

PYR1 transgenic studies. To probe function of mutant PYR1 proteins *in planta*, we cloned mutant cDNAs into pEGAD[38] to create the 35S::GFP–PYRH115A construct, which was then transformed into the *pyr1/pyl1/pyl2/pyl4* quadruple mutant line as described[7] using floral dip transformation. For complementation tests, T1 seeds were germinated on MS plates after 4 days stratification and two day old seedlings were screened using a fluorescence-dissecting microscope to identify GFP^{+} seedlings. Six to ten GFP^{+} seedling were transferred to MS plates (at least 6 seedlings per plate) containing differing ABA concentrations, and grown vertically under fluorescent lighting. Root lengths were measured 4 days post transfer. For analysis of 35S::GFP–PYR1, a homozygous single insert line in the quadruple mutant background was used, which was made using a previously described DNA construct[7].

Protoplast transient assays. An ABA-inducible luciferase reporter was constructed by replacing the *CBF3* promoter in the *CBF3–LUC* construct with the promoter region of *RD29B*, a highly ABA-inducible gene[8]. *CBF3–LUC, UQ10–GUS* and *ABI1* protoplast expression plasmid vectors were kindly provided by J. Sheen (Massachusetts General Hospital). The coding regions of His–PYR1 and mutants, *ABI1–myc, ABF2–HA, and SnRK2.6–Flag* were used to replace *ABI1* in the *ABI1* protoplast expression plasmid vectors.

Protoplasts were isolated from *Arabidopsis thaliana* ecotype Columbia-0 plants, which were grown on Jiffy7 soil (Jiffy Products) in an environment-controlled chamber at 22 °C with a photosynthetically active radiation of 75 µmol m^{-2} s^{-1} and a day/night cycle of 13 h light/11 h dark. Transient activity assays were performed in *Arabidopsis* mesophyll protoplasts from Columbia-0 wild-type plants as described by the Sheen laboratory (http://genetics.mgh.harvard.edu/sheenweb)[20]. Transfected protoplasts were incubated for 5 h in light in the presence of 0 or 5 µM (+)-ABA, and then used for measuring LUC and GUS activity. The *UQ10–GUS* reporter plasmid (β-glucuronidase) was used as internal control to normalize transfection efficiency in protoplast assays.

NMR studies of the wild-type and H115A mutated PYR1. The wild type and H115A mutated PYR1 proteins for NMR were expressed in *E. coli* and purified as described previously[7]. ABA–PYR1 complexes were obtained by addition of a twofold excess of (+)-ABA to the purified proteins. NMR experiments were performed at 37 °C on a Bruker Avance 600MHz spectrometer equipped with a 5 mm TCI CryoProbe and processed with NMRPipe[39]. Backbone ^{1}H, ^{15}N and ^{13}C chemical shifts of ABA–PYR1 were obtained using automated assignment software[40,41] and manual analysis of 3D HNCO, HNCOCA, HNCA, HNCOCACB, HNCA, HNCACB, HNCACO, CCONH, HBHACONH, HCCONH, and ^{15}N-edited NOESY-HSQC spectra. For apo-PYR1, only 3D HNCA, 3D HNCO, and 3D ^{15}N NOESY experiments were performed before sample degradation prevented further data collection, but these spectra were sufficient to permit identification of all signals assigned in the ABA–PYR1complex. Assignments were confirmed for 120 of the 183 non-proline residues of PYR1. Portions of the PYR1 sequence could not be assigned due to extreme line broadening in the 2D and 3D spectra, including residues 1–7, 54–64, 81–88, 99–106, 117–123, and 152–166. Many of these residues participate in a dimer interface observed in the crystal packing of apo-PYL2, suggesting that line broadening associated with monomer–dimer exchange is responsible for the majority of missing signals.

31. Smith, P. A. *et al.* A plasmid expression system for quantitative *in vivo* biotinylation of thioredoxin fusion proteins in *Escherichia coli*. *Nucleic Acids Res.* **26,** 1414–1420 (1998).
32. Otwinowski, Z., Borek, D., Majewski, W. & Minor, W. Multiparametric scaling of diffraction intensities. *Acta Crystallogr. A* **59,** 228–234 (2003).
33. Kleywegt, G. J. & Jones, T. A. Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr. D* **50,** 178–185 (1994).
34. Wiehe, K. *et al.* The performance of ZDOCK and ZRANK in rounds 6–11 of CAPRI. *Proteins* **69,** 719–725 (2007).
35. Li, Y., Kovach, A., Suino-Powell, K., Martynowski, D. & Xu, H. E. Structural and biochemical basis for the binding selectivity of peroxisome proliferator-activated receptor gamma to PGC-1alpha. *J. Biol. Chem.* **283,** 19132–19139 (2008).
36. Suino, K. *et al.* The nuclear xenobiotic receptor CAR; structural determinants of constitutive activation and heterodimerization. *Mol. Cell* **16,** 893–905 (2004).
37. Xu, H. E. *et al.* Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha. *Nature* **415,** 813–817 (2002).
38. Cutler, S. R., Ehrhardt, D. W., Griffitts, J. S. & Somerville, C. R. Random GFP:cDNA fusions enable visualization of subcellular structures in cells of *Arabidopsis* at a high frequency. *Proc. Natl Acad. Sci. USA* **97,** 3718–3723 (2000).
39. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
40. Bartels, C., Günterts, P., Billeter, M. & Wüthrich, K. GARANT — a general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra. *J. Comput. Chem.* **18,** 139–149 (1997).
41. Bahrami, A., Assadi, A. H., Markley, J. L. & Eghbalnia, H. R. Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy. *PLOS Comput. Biol.* **5,** e1000307 (2009).

Vol 462|3 December 2009|doi:10.1038/nature08583

Structural basis of abscisic acid signalling

Ken-ichi Miyazono[1]*, Takuya Miyakawa[1]*, Yoriko Sawano[1]*, Keiko Kubota[1]*, Hee-Jin Kang[1], Atsuko Asano[1], Yumiko Miyauchi[1], Mihoko Takahashi[1], Yuehua Zhi[1], Yasunari Fujita[2], Takuya Yoshida[2,3], Ken-Suke Kodaira[2,3], Kazuko Yamaguchi-Shinozaki[2,3] & Masaru Tanokura[1]

The phytohormone abscisic acid (ABA) mediates the adaptation of plants to environmental stresses such as drought and regulates developmental signals such as seed maturation. Within plants, the PYR/PYL/RCAR family of START proteins receives ABA to inhibit the phosphatase activity of the group-A protein phosphatases 2C (PP2Cs), which are major negative regulators in ABA signalling. Here we present the crystal structures of the ABA receptor PYL1 bound with (+)-ABA, and the complex formed by the further binding of (+)-ABA-bound PYL1 with the PP2C protein ABI1. PYL1 binds (+)-ABA using the START-protein-specific ligand-binding site, thereby forming a hydrophobic pocket on the surface of the closed lid. (+)-ABA-bound PYL1 tightly interacts with a PP2C domain of ABI1 by using the hydrophobic pocket to cover the active site of ABI1 like a plug. Our results reveal the structural basis of the mechanism of (+)-ABA-dependent inhibition of ABI1 by PYL1 in ABA signalling.

The phytohormone ABA has central roles in the adaptation of vegetative tissues to water stress, as well as in seed maturation and dormancy. ABA promotes stomatal closure in guard cells, which is mediated by solute efflux, and regulates the expression of many genes, the products of which may function in dehydration tolerance in both vegetative tissues and seeds.

Phosphorylation/dephosphorylation-mediated regulation has been reported to have key roles in ABA signalling (Supplementary Fig. 1). Subclass III protein kinases of the SNF1-related protein kinase 2 (SnRK2) family have been shown to be activated by ABA[1,2], to mediate the regulation of stomatal apertures[3,4], and to function upstream of ABA-responsive gene expression[5,6]. They phosphorylate downstream substrates, including AREB/ABF-type bZIP transcription factors, to activate ABA-responsive gene expression or other ABA-related responses[5,7,8]. Group-A PP2Cs have also been shown to be very important components of ABA signalling[9]. They act as negative regulators of ABA signalling in both seeds and vegetative tissues[9], and inactivate subclass III SnRK2s by interacting with them[10]. Thus, the positive and negative regulators of ABA signalling and their downstream regulation of ABA-responsive gene expression have been well characterized. However, the findings concerning ABA receptors are less clear. FCA, ABAR (also called CHLH or GUN5), GCR2, GTG1 and GTG2 were initially reported as ABA receptors, but some of these identifications were retracted or questioned, and none of these receptors is involved in the previously identified positive and negative regulators of ABA signalling[11].

Recently, notable progress in ABA receptor research was reported by two independent research groups[10,12]. These groups identified the *Arabidopsis* PYR/PYL/RCAR family of START proteins as ABA receptors that inhibit group-A PP2C activity in response to ABA. These ABA receptor proteins are connected to the downstream ABA signalling pathway according to the following model. In the absence of ABA, PP2Cs repress the ABA signalling pathway through inactivation of SnRK2s. In the presence of ABA, PYR/PYL/RCARs recognize and bind to PP2Cs, thereby releasing SnRK2s from PP2C-dependent negative regulation, which allows SnRK2s to phosphorylate downstream substrates such as AREB/ABFs and activate ABA-responsive gene expression or other ABA-related responses.

Here we report a series of crystallographic and biochemical studies of (+)-ABA perception by the (+)-ABA receptor PYL1 and (+)-ABA-dependent inhibition of the PP2C protein ABI1 by the receptor. Our results reveal the structural mechanism of ABI1 inhibition that underlies the regulation of phosphorylation signalling in plant stress responses and seed developmental processes by ABA sensing of PYL1.

PYL1 acts as a (+)-ABA-specific receptor

The binding of (+)-ABA by PYL1 and the (+)-ABA-dependent intermolecular interaction of PYL1 with ABI1 were confirmed by nuclear magnetic resonance (NMR), isothermal titration calorimetry (ITC), surface plasmon resonance (SPR) and glutathione *S*-transferase (GST) pull-down assays. The NMR measurements showed that PYL1 binds a biologically active phytohormone (+)-ABA but not a non-natural stereoisomer (−)-ABA (Fig. 1a, b). In the ITC and SPR measurements, PYL1 was indicated to specifically bind (+)-ABA with dissociation constant (K_d) values of 52 μM and 340 μM, respectively (Fig. 1c and Supplementary Table 1). This difference between the K_d values may be tolerated by considering the accuracy of the curve fitting and the effects of immobilization of the protein in the SPR measurements. The reaction is extraordinarily endothermic in spite of the rather high affinity. Moreover, the pull-down assays and ABI1 inhibition assays revealed that PYL1 interacts with the PP2C domain of ABI1 and inhibits the PP2C activity in a (+)-ABA-dependent manner (Fig. 1d, e). These observations show that PYL1 acts as an (+)-ABA receptor in a manner similar to PYR1[10].

Structure of the (+)-ABA-bound PYL1

The structure of (+)-ABA-bound PYL1 (PYL1–(+)-ABA) adopts the typical START protein fold consisting of three helices and a seven-stranded antiparallel β-sheet with the following topology: $\beta1-\alpha2-\eta1-\beta2-\beta3-\beta4-\beta5-\beta6-\beta7-\alpha3$ (where η represents a 3_{10}-helix; Fig. 2a, b and Supplementary Fig. 2). The highly curved β-sheet surrounds the three helices on one side and faces an additional

---

[1]Department of Applied Biological Chemistry, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan. [2]Biological Resources Division, Japan International Research Center for Agricultural Sciences, Ibaraki 305-8686, Japan. [3]Laboratory of Plant Molecular Physiology, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo 113-8657, Japan.
*These authors contributed equally to this work.

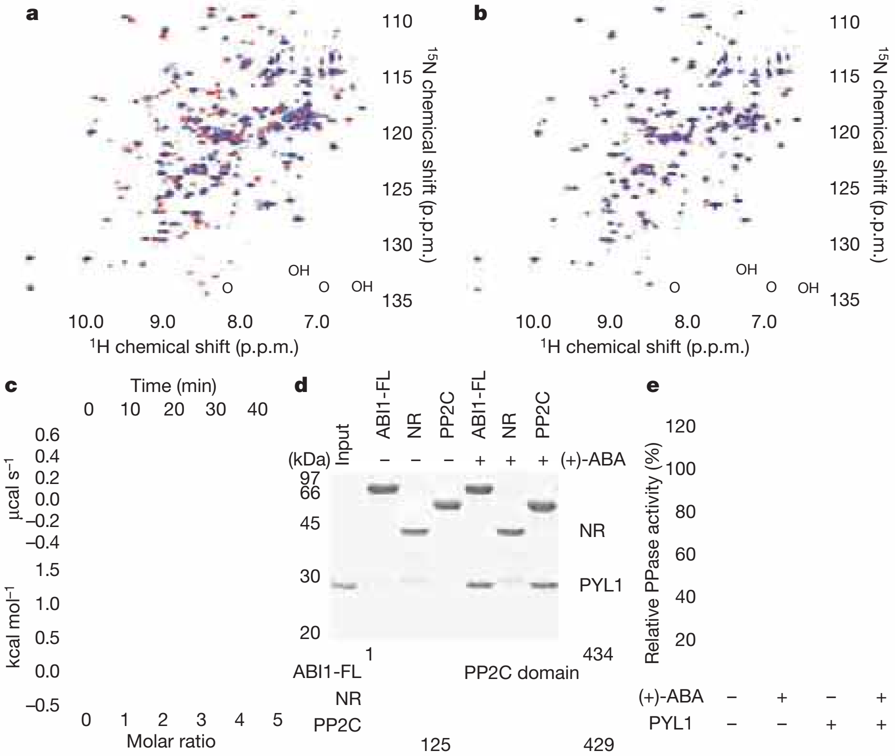


Figure 1 | (+)-ABA-binding specificity and (+)-ABA-dependent interaction of PYL1 with ABI1. **a**, **b**, Chemical-shift perturbation by the addition of (+)-ABA (**a**) and (−)-ABA (**b**) to PYL1. The ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) spectra of 0.2 mM PYL1(28–210) were recorded in the absence (red) and presence (blue) of 0.5 mM ABA. The structure of the ABA molecule added to the NMR sample is shown in the corresponding spectrum. p.p.m. parts per million. **c**, ITC of PYL1 with (+)-ABA. The raw and integrated ITC data are shown in the upper and lower panels, respectively. The K_d, ΔH and ΔS values were 52 μM, 1.4 kcal mol^{-1} and 24 cal mol^{-1} deg^{-1}, respectively. **d**, ABA-dependent interaction of PYL1 with several constructs of ABI1. The proteins after GST pull-down assay in the presence (+) or absence (−) of 10 μM (+)-ABA are separated by SDS–PAGE and visualized by Coomassie blue staining. The sizes of the protein markers are indicated on the left side of the panel. ABI1 constructs are illustrated at the bottom of the SDS–PAGE panel. FL, full length; NR, N-terminal region. **e**, Effect of protein phosphatase (PPase) activity of ABI1 by (+)-ABA and PYL1. The initial reaction velocities of ABI1(125–429) were detected by using the system designed with RPA(pT)VA phosphopeptide as a substrate ($n = 3$). (+)-ABA was added in the reaction solution with a final concentration of 10 μM. The relative activities in the absence of (+)-ABA are defined as 100%. Error bars represent standard deviation (s.d.).

amino-terminal α-helix (α1) on the other side. The ligand-binding site of START proteins is formed by the three helices (α2, η1 and α3 of PYL1) and the inner side of the highly curved β-sheet, and (+)-ABA binds to this site. The entrance to the ligand-binding site of PYL1 is covered by two loops, the β3–β4 and β5–β6 loops, which trap one (+)-ABA molecule in the site (Fig. 2a, b). The surface model of the PYL1 structure shows that the entrance of the ligand-binding site is too small for (+)-ABA to enter the site (Fig. 2d). The solvent-accessible area of (+)-ABA in the bound form is only 12.0 Å^2, which corresponds to 2.6% of the total (calculated in AREAIMOL[13]). Although the (+)-ABA free structure of PYL1 is still unclear, the structure of the loops covering the ligand-binding site would take an open conformation in a ligand-free form and become closed during (+)-ABA binding in the manner of a 'lid'. The conformational change to the closed form gives rise to a large hydrophobic area on the surface of the lid. The formation of a large hydrophobic surface area is observed as an endothermic reaction, which is also seen in the case of Ca^{2+} binding to the regulatory protein calmodulin[14,15].

Mechanism of ABA recognition by PYL1

The structure of PYL1–(+)-ABA shows that PYL1 recognizes the characteristic shape and functional groups of (+)-ABA with high specificity by means of an ion pair, hydrogen-bond networks, and hydrophobic contacts (Fig. 3 and Supplementary Fig. 3). The carboxyl group of (+)-ABA binds to the inner side of the ligand-binding site of PYL1 and is recognized by formation of an ion pair with Lys 86 and three water-mediated hydrogen-bond networks (Fig. 3a, c). In addition to the ion-pair formation, one oxygen atom of the carboxyl group (O11) is recognized by formation of hydrogen bonds with two water molecules that are stabilized by the side chains of Tyr 147, Ser 149, Glu 171 and Asn 197. The other oxygen atom of the carboxyl group (O12) is recognized by formation of a water-mediated hydrogen-bond network with Glu 121. This residue is also involved in the recognition of the hydroxyl group of the cyclohexene ring of (+)-ABA by a water-mediated hydrogen-bond network. The cyclohexene group of (+)-ABA is surrounded by the outer side of the ligand-binding site of PYL1 and its carbonyl group is recognized by formation of a water-mediated hydrogen bond with Arg 143 on the β5–β6 loop that covers the ligand-binding site. There are also a number of hydrophobic contacts between PYL1 and (+)-ABA (Fig. 3b, c). The pentadienoic acid group of (+)-ABA is mainly recognized by hydrophobic interaction with Phe 135, Leu 144 and Val 193. Among these residues, Val 193 on the α3 helix forms the most significant contacts with (+)-ABA, with a buried surface area of 48.6 Å^2, to recognize the C2, C3, C4 and C6 atoms of (+)-ABA. The cyclohexene group of (+)-ABA is surrounded by and interacts with several hydrophobic residues forming ligand-binding sites (that is, Phe 88, Val 110, Ala 116, Ile 137 and Phe 189) to recognize the shape of the ligand. These (+)-ABA recognition residues are highly conserved among those homologous proteins that are predicted to be ABA-specific receptor proteins (Supplementary Fig. 4).

Structure of the PYL1–(+)-ABA–ABI1 complex

In the complex between (+)-ABA-bound PYL1 and ABI1 (PYL1–(+)-ABA–ABI1) the subunits form a heterodimer (Fig. 4a). Although the structure of PYL1–(+)-ABA is almost identical before and after interacting with ABI1 (a root mean squared deviation

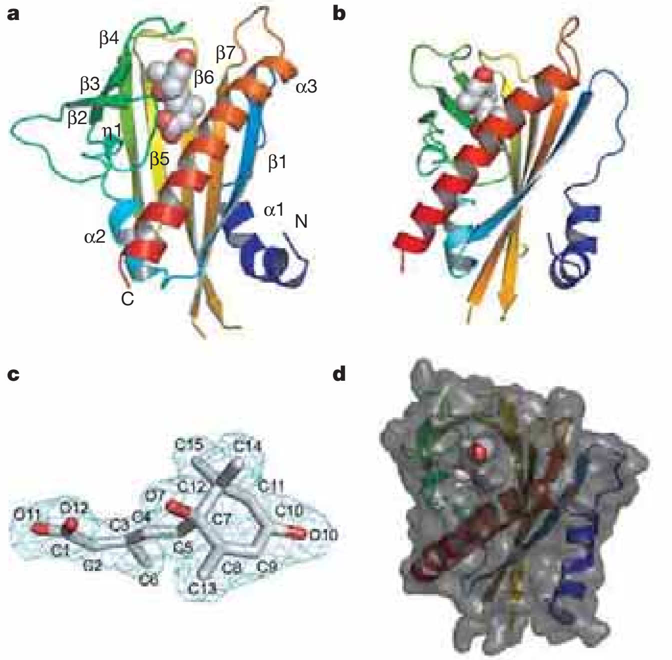


Figure 2 | Overall structure of (+)-ABA-bound PYL1. a, Ribbon diagram of the overall structure with annotation of secondary structure elements. The ribbon is coloured blue (in the N terminus) to red (in the C terminus). (+)-ABA is indicated by a sphere model coloured grey (carbon atoms) and red (oxygen atoms). **b**, As in **a**, but rotated 90° around a vertical axis. **c**, 2$F_o - F_c$ map around (+)-ABA observed in the structure of PYL1–(+)-ABA. The electron density of (+)-ABA is contoured by 1σ (cyan mesh). (+)-ABA is shown using a stick model coloured grey (carbon atoms) and red (oxygen atoms). **d**, Surface model of the structure. The molecular surface of PYL1 is coloured dark grey.

(r.m.s.d.) value of 0.8 Å for 172 superposed Cα atoms), the structures of three loops—α1–β1, β3–β4 and β7–α3—are slightly modified by the formation of the complex. In particular, the loops β3–β4 and β7–α3 are shifted to (+)-ABA to form a narrow hydrophobic pocket consisting of Pro 115, Arg 143, Leu 144, Pro 178, Asn 181, Asp 185, Thr 186 and Phe 189 with approximate dimensions of 8 × 5 × 7 Å^3 on the outside of the (+)-ABA binding site (Fig. 4c).

Meanwhile, the ABI1 structure in the complex adopts a similar folding pattern to the PP2C family proteins, with a central β-sandwich surrounded by α-helices[16]. ABI1 is composed of two five-stranded antiparallel β-sheets (the order of the strands is β2, β3, β5, β6, β12 and β1, β14, β13, β7, β8) forming a β-sandwich in its central region, five α-helices surrounding the β-sandwich (α1 and α2 on one side and α4, α5 and α6 on the other side), and an additional small domain with a two-stranded antiparallel β-sheet and helix α3, which sticks out from the core structure. The structure of ABI1 is well superposed with the human PP2C, with an r.m.s.d. value of 1.8 Å for 257 superposed Cα atoms (Supplementary Fig. 5a). The sequential and structural comparison of ABI1 with the pre-existing PP2C family phosphatases revealed that the active-site residues of ABI1 are highly conserved (residues Arg 138, Glu 142, Asp 143, Asp 177, Gly 178, His 179, Asp 347 and Asp 413; see Supplementary Figs 5b and 6).

In the structure of the PYL1–(+)-ABA–ABI1 complex, PYL1 contacts ABI1 using the loops covering the (+)-ABA-binding site (β3–β4, β5–β6 and β7–α3) and the N-terminal part of the helix α3. On the other hand, ABI1 interacts with PYL1 using the region around its active site and the small protruding domain described above. The interface is almost hydrophobic and several hydrogen bonds stabilize the interaction (Fig. 4d and Supplementary Fig. 7). The buried surface areas of ABI1 and PYL1 in the complex are 7.4% (883 Å^2) and 8.9% (851 Å^2) of the total surface area of each molecule, respectively. The most striking feature of the intermolecular interaction between PYL1 and ABI1 is that the active-site cleft of ABI1 is covered by the β3–β4 loop of PYL1 (Fig. 4b). In addition, Trp 300 of the small protruding domain of ABI1, the side chain of which sticks out from the surface, also undergoes a significant interaction with the narrow hydrophobic pocket of (+)-ABA-bound PYL1 (Fig. 4c).

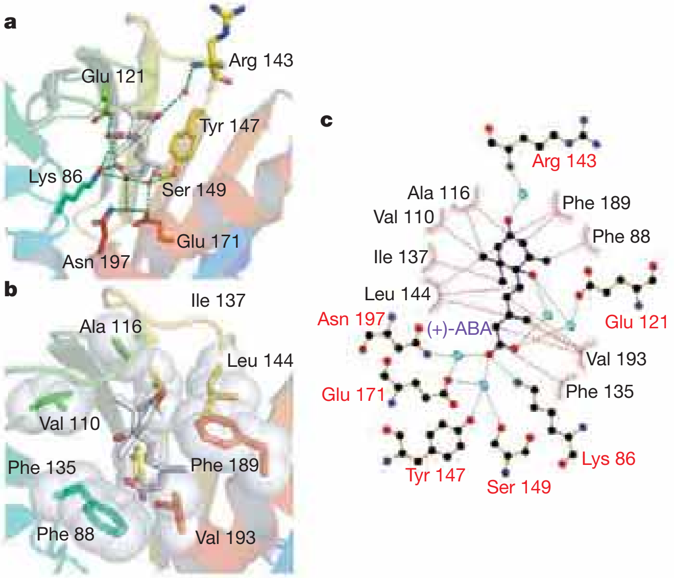


Figure 3 | (+)-ABA recognition by PYL1. a, Hydrogen-bond networks used for the recognition of (+)-ABA. Residues involved in the hydrogen-bond networks are shown using a stick model. Water molecules and hydrogen bonds are indicated by red spheres and green dotted lines, respectively. **b**, Hydrophobic contacts between PYL1 and (+)-ABA. Residues involved in the recognition of (+)-ABA (Phe 88, Val 110, Ala 116, Phe 135, Ile 137, Leu 144, Phe 189 and Val 193 of PYL1) are shown by using stick and sphere models. **c**, Schematic diagram of the recognition mechanism of (+)-ABA by PYL1. Water molecules are shown as cyan spheres. Hydrogen bonds and hydrophobic contacts are indicated by green and red dotted lines, respectively.

In the active site of ABI1, Ser 112 of PYL1 contacts Glu 142 of ABI1, which is one of the active-site residues of ABI1, by direct and water-mediated hydrogen bonds at the side-chain oxygen and main-chain nitrogen atoms of Ser 112 of PYL1 and the side chain of Glu 142 of ABI1 (Fig. 4b). The main-chain oxygen atom of Ser 112 of PYL1 also forms a hydrogen bond with the main-chain nitrogen atom of Gly 180 of ABI1. Other residues of the β3–β4 loop of PYL1 are located near the active-site cleft of ABI1. The accessible surface area of the active-site residues of ABI1 is reduced from 293.1 Å^2 to 170.5 Å^2 by the interaction with PYL1. These facts would suggest that the enzymatic activity of ABI1 is inhibited by the binding of PYL1, because PYL1 covers the active-site cleft of ABI1 to prevent access to the substrates of ABI1, phosphoserine and phosphothreonine. The dominant *abi1-1* mutant of *Arabidopsis* that converts Gly 180 of ABI1 to Asp shows a decrease in ABA responsiveness[17–19]. In our structure, a substitution of Gly 180 to Asp seems to reduce the intermolecular interaction at this site due to the bulky acidic side chain of Asp. The ABA-insensitive phenotype of the *abi1-1* mutant may be caused by the inhibitory effect of Asp 180 on the intermolecular interaction between PYL1 and ABI1.

The PYL1–(+)-ABA–ABI1 complex is highly stabilized by the interaction between Trp 300 of ABI1 and the narrow hydrophobic pocket of PYL1 on the outside of the (+)-ABA-binding site; the side chain of Trp 300 of ABI1 is buried within the hydrophobic pocket of PYL1 to form hydrophobic contacts with the residues Pro 115, Arg 143, Leu 144, Pro 178, Asn 181, Asp 185, Thr 186 and Phe 189 of PYL1 (Fig. 4c). In the hydrophobic pocket, the side-chain imino group of Trp 300 of ABI1 forms a water-mediated hydrogen-bond network with Pro 115 and Arg 143 of PYL1, and with the O10 atom of

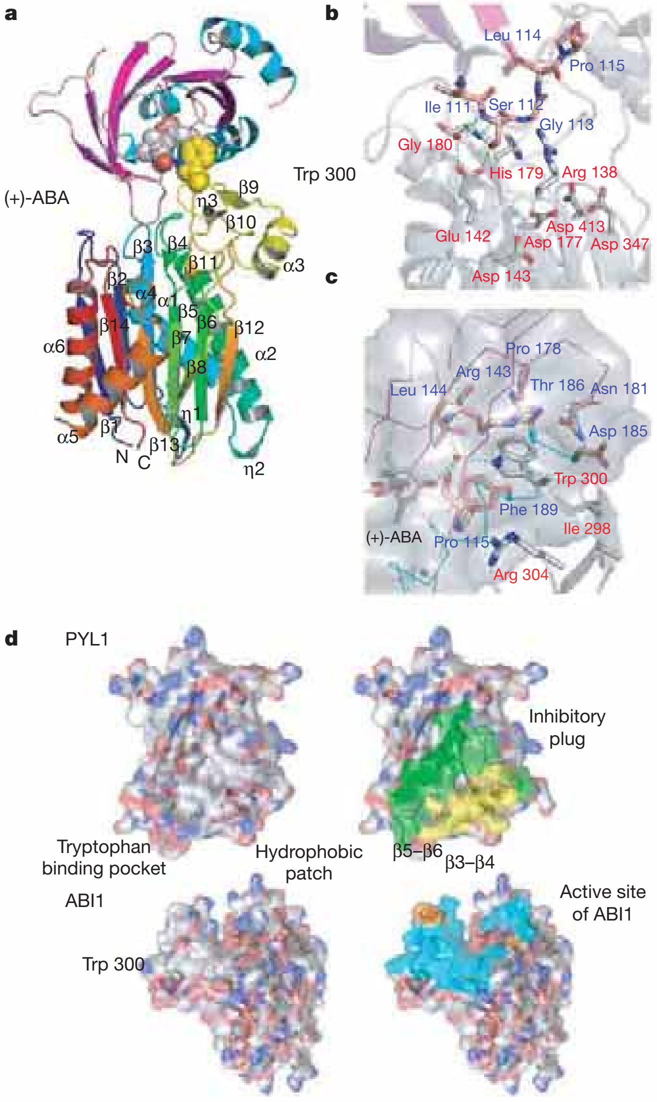


Figure 4 | Overall structure of the PYL1–(+)-ABA–ABI1 complex. **a**, Ribbon diagram of the overall structure of the PYL1–(+)-ABA–ABI1 complex. The ribbon of ABI1 is coloured blue (in the N terminus) to red (in the C terminus) and that of PYL1 is coloured magenta (strands) and cyan (helices). (+)-ABA and Trp 300 of ABI1 are shown using a sphere model. **b**, Intermolecular interaction between PYL1 and ABI1 around the active site of ABI1. Side chains of the active residues of ABI1 are shown by a stick model and are labelled red. Residues of the β3–β4 loop of PYL1 are shown by a stick model and are labelled blue. A water molecule and hydrogen bonds are indicated by red spheres and green dotted lines, respectively. **c**, Intermolecular interaction between PYL1 and ABI1 around Trp 300 of ABI1. (+)-ABA, Trp 300 and the residues of PYL1 forming a hydrophobic pocket that binds Trp 300 of ABI1 are shown by a stick model. The molecular surface of PYL1 is coloured grey. **d**, Surface models of PYL1 and ABI1. In the left panel, the molecular surfaces are coloured grey (carbon atom), blue (nitrogen atom) and red (oxygen atom). In the right panel, interface residues are coloured cyan (interface residues of ABI1), orange (Glu 142 and Trp 300 of ABI1) green (interface residues of PYL1) and yellow (interface residues from (+)-ABA covering loops β3–β4 and β5–β6 of PYL1).

(+)-ABA to recognize the presence of (+)-ABA in the (+)-ABA-binding pocket of PYL1.

Key contacts between PYL1 and ABI1

To confirm the structural observations for the (+)-ABA-dependent PYL1–ABI1 interaction and the inhibition of ABI1 PP2C activity by PYL1, we performed pull-down binding assays and ABI1-inhibition assays using several mutants of PYL1 and ABI1. According to the derived structure, Ser 112 on the β3–β4 loop of PYL1 directly interacts with the catalytic residue Glu 142 of ABI1 (Fig. 4b). The S112A mutant of PYL1 and the E142A mutant of ABI1 showed a lower affinity than the wild-type proteins (Fig. 5a, b). In addition, the E142A mutant almost completely lost the phosphatase activity even in the absence of PYL1–(+)-ABA (Fig. 5c), and the S112A mutant of PYL1 also showed a decrease in inhibitory activity against ABI1 (Fig. 5d). These findings support the idea that the β3–β4 loop of PYL1 covers the active site of ABI1 so as to inhibit the PP2C activity of ABI1.

Among the loops surrounding the ABA-binding site of PYL1, the β5–β6 loop constituted by His 142, Arg 143 and Leu 144 in particular contributes to the PYL1–ABI1 interaction (Fig. 5a). The β5–β6 loop is arranged to form the hydrophobic pocket of PYL1 that entraps the side chain of Trp 300 of ABI1 (Fig. 4c), and the W300A mutant of ABI1 shows a marked loss of affinity for PYL1 (Fig. 5b). In addition, ABI1 scarcely binds to PYL1 by each point mutation of the residues Ile 298 and Arg 304 (Fig. 5b), which make several contacts with the residues forming the hydrophobic Trp 300-binding pocket of PYL1 (Fig. 4c). Another hydrophobic region is also formed on the outer side of the η1–β2 loop, the β3–β4 loop and the α3 helix around the ABA-binding site of PYL1 (Fig. 3b) and interacts with the residues Phe 306, Val 308 and Tyr 319 of ABI1 (Fig. 4d and Supplementary Fig. 7). The contribution of these residues to the PYL1–ABI1 interaction is considerably smaller than that of all the ABI1 residues contacting the hydrophobic pocket of PYL1 (Fig. 5b). Therefore, the pocket seems to be the key architecture for the interaction between PYL1 and ABI1. Moreover, each point mutation of Leu 144 of PYL1 and Trp 300 of ABI1 was confirmed to cause a marked level of inhibitory activity of PYL1 against ABI1 (Fig. 5d), although the tryptophan residue was not involved in the phosphatase activity of ABI1 directly (Fig. 5c). These results indicate that the docking of ABI1 Trp 300 into the PYL1 hydrophobic pocket is an essential step for the inhibition of ABI1 by PYL1 and is probably necessary for the β3–β4 loop of PYL1 to be properly located into the active site of ABI1.

Our alanine scanning analysis indicates that the ABA-dependent PYL1–ABI1 interaction is hardly affected by the point mutation of Pro 115 compared with those of the residues Arg 143, Leu 144 and Phe 189 of the PYL1 hydrophobic pocket (Fig. 5a). It has, however, been reported that the point mutation of Pro 88 of PYR1—which corresponds to Pro 115 of PYL1—to Ser causes the immediate loss of ABA-dependent inhibitory activity against ABI1[10]. This difference may be explained by the rise in the hydrophilicity of the hydrophobic Trp 300-binding pocket, because the proline residue is a component of the pocket.

Mechanism of ABA-dependent inhibition

The present study enables us to outline the molecular mechanism of the (+)-ABA-dependent regulation of ABI1 PP2C activity by PYL1 as follows (Fig. 6). The crystal structure of PYL1–(+)-ABA has indicated that (+)-ABA is an allosteric regulator of the ABA receptor PYL1. When a (+)-ABA molecule binds to PYL1, the β3–β4 and β5–β6 loops of PYL1 (the lids) make significant conformational changes and completely cover the (+)-ABA-specific ABA-binding site that is responsible for the entrapment. The resulting PYL1–(+)-ABA complex gives rise to the ABI1 PP2C-domain-binding site, which consists of the inhibitory plug and a large hydrophobic surface area with the pocket for ABI1 Trp 300 binding. The β5–β6 loop forms a key component of the lid and of the hydrophobic binding pocket for ABI1 Trp 300, which is the only residue in ABI1 that interacts with a (+)-ABA molecule entrapped in PYL1 by using a hydrogen-bond network through a water molecule. On the other hand, the β3–β4 loop of the lid comes into contact with the active site of ABI1 to inhibit the phosphatase activity of ABI1 in a (+)-ABA-dependent manner (Fig. 5d). The conformations of these loops do not seem to take a fully stable state only by binding of a (+)-ABA molecule. The comparison of the PYL1–(+)-ABA and PYL1–(+)-ABA–ABI1 complex structures indicates the significant structural rearrangement in the

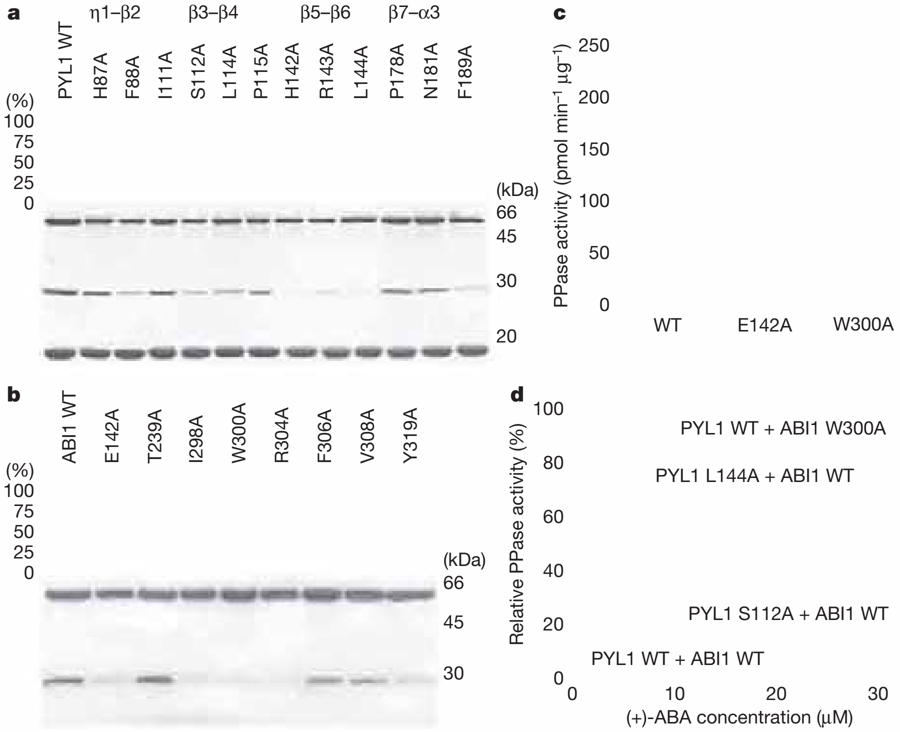


Figure 5 | Recognition and inhibition of ABI1 by PYL1. **a**, **b**, Effect of point mutations of PYL1 (**a**) and ABI1 (**b**) on the interaction between PYL1 and ABI1. The top panel of **a** and upper panel of **b** show the relative affinities of PYL1 mutants and ABI1 mutants, respectively. The relative affinities of the wild-type proteins are defined as 100%. The middle panel of **a** and lower panel of **b** show the electrophoretic patterns of PYL1 and ABI1 after GST pull-down assays in the presence of 10 µM (+)-ABA. The proteins are separated by SDS–PAGE and stained with Coomassie blue. The sizes of protein markers are indicated on the right side of the panel. The protein bands located at the size of ~50 kDa and ~25 kDa are GST-fused ABI1(125–429) and PYL1(8–211), respectively. The bottom panel of **a** shows the PYL1 proteins' input to the pull-down assays (each in the amount of 50 pmol). The top labels of **a** show the loops including the designated residues; Phe 189 is located on helix α3 of PYL1. WT, wild type. **c**, PPase activity of ABI1 and its mutants. The initial reaction velocities of ABI1(125–429) and its mutants were detected by using the system designed with RPA(pT)VA phosphopeptide as a substrate ($n = 3$). Error bars represent s.d. **d**, ABA-dependent inhibition of ABI1 PPase activity by PYL1. The relative PPase activities are measured and plotted in the presence of different concentrations of (+)-ABA ($n = 3$). The relative activities in the absence of (+)-ABA are defined as 100%. Error bars represent s.d.

surrounding loops of the ABA-binding site and the hydrophobic pocket of PYL1 accompanying the complete formation of the complex with ABI1. In the (+)-ABA-bound form of PYL1, the lid loops may take metastable conformations and then be refined to the stable conformations by contacting the PP2C domain of ABI1. This structural consideration explains that the interaction between ABI1 and RCAR1 (also known as PYL9), which is a member of the ABA receptors, increases the ABA-binding affinity of this ABA receptor protein[12]. Our model leads to the idea that the amount of metastable PYR/PYL/RCAR–(+)-ABA complex is a limiting factor for ABA signalling in plant cells.

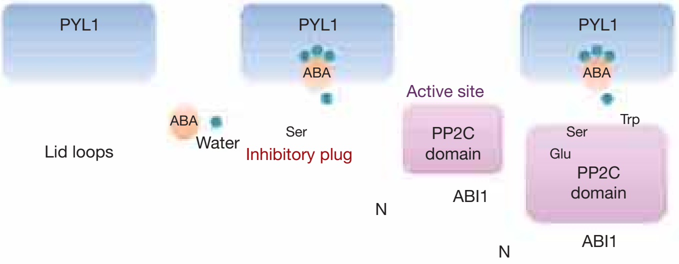


Figure 6 | A model of (+)-ABA-induced ABI1 regulation by PYL1. PYL1 is an intracellular (+)-ABA receptor in (+)-ABA-dependent signal transduction. When a (+)-ABA molecule enters into an ABA-binding site of PYL1, the conformations of lid loops mainly composed of the β3–β4 (red) and β5–β6 (green) loops change from the mobile, open state to the fixed, closed state. The fixed loops of PYL1 act as a scaffold to interact with the PP2C domain of ABI1 (for example, the hydrophobic pocket accepting the Trp 300 of ABI1). This ABA-dependent contact between PYL1 and ABI1 allows access of the β3–β4 loop to the active site of the PP2C domain. Thus, PYL1–(+)-ABA activates the SnRK2-dependent phosphorylation pathways by inhibiting the phosphatase activity of ABI1.

The (+)-ABA-mediated phosphorylation/dephosphorylation signalling pathway is a potent strategy for regulating the signalling that controls responses to environmental stresses and developmental processes. Our structural investigation has provided the detailed structural mechanism for the (+)-ABA-dependent inhibition of a negative regulator ABI1 in the pathway by a (+)-ABA receptor PYL1, which has deepened our understanding of the hormonal-signal-dependent phosphorylation regulation of plant cellular functions.

METHODS SUMMARY

The recombinant PYL1 was purified by using a histidine-tag affinity column and anion-exchange chromatography. The purified protein was concentrated to 10 mg ml^{-1} in a buffer containing 10 mM Tris-HCl, pH 7.0, 2 mM TCEP and 4 mM (+)-ABA/(+/−)-ABA. The crystals of the protein were obtained under reservoir solution conditions of 100 mM HEPES, pH 8.0, 24% PEG3350 and 0.2 M NaCl.

The recombinant ABI1 was purified by a histidine-tag affinity column and was mixed with PYL1–(+)-ABA to form the complex. The complex was further purified by anion-exchange chromatography and gel filtration chromatography. The purified protein was concentrated to 22 mg ml^{-1} in the following buffer: 10 mM Tris-HCl, pH 8.0, 100 mM NaCl and 1 mM TCEP. The crystals of the complex were obtained under reservoir solution conditions of 100 mM sodium citrate, pH 5.0 and 21% PEG3000.

The structure of PYL1–(+)-ABA was determined by the Se-MAD method. The final model of PYL1–(+)-ABA was refined to 2.05 Å resolution with R and R_{free} values of 19.2% and 22.9%, respectively. The structure of PYL1–(+)-ABA–ABI1 was determined by the molecular replacement method using the coordinates of PYL1 and human PP2C[16]. The final model of PYL1–(+)-ABA–ABI1 was refined to 2.10 Å resolution with R and R_{free} values of 20.1% and 24.9%, respectively.

The ABA-binding assays were performed by ITC, SPR and NMR (see Methods). The GST pull-down assays were performed in the presence of 10 μM (+)-ABA. The PP2C inhibitory assays were performed using a serine/threonine phosphatase assay system.

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

Received 23 September; accepted 13 October 2009.

1. Boudsocq, M., Barbier-Brygoo, H. & Laurière, C. Identification of nine sucrose nonfermenting 1-related protein kinases 2 activated by hyperosmotic and saline stresses in *Arabidopsis thaliana*. *J. Biol. Chem.* **279,** 41758–41766 (2004).
2. Kobayashi, Y., Yamamoto, S., Minami, H., Kagaya, Y. & Hattori, T. Differential activation of the rice sucrose nonfermenting 1-related protein kinase 2 family by hyperosmotic stress and abscisic acid. *Plant Cell* **16,** 1163–1177 (2004).
3. Mustilli, A., Merlot, S., Vavasseur, A., Fenzi, F. & Giraudat, J. *Arabidopsis* OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. *Plant Cell* **14,** 3089–3099 (2002).
4. Yoshida, R. *et al.* ABA-activated SnRK2 protein kinase is required for dehydration stress signaling in *Arabidopsis. Plant Cell Physiol.* **43,** 1473–1483 (2002).
5. Fujii, H. & Zhu, J. *Arabidopsis* mutant deficient in 3 abscisic acid-activated protein kinases reveals critical roles in growth, reproduction, and stress. *Proc. Natl Acad. Sci. USA* **106,** 8380–8385 (2009).
6. Nakashima, K. *et al.* Three *Arabidopsis* SnRK2 protein kinases, SRK2D/SnRK2.2, SRK2E/SnRK2.6/OST1 and SRK2I/SnRK2.3, involved in ABA signaling are essential for the control of seed development and dormancy. *Plant Cell Physiol.* **50,** 1345–1363 (2009).
7. Furihata, T. *et al.* Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. *Proc. Natl Acad. Sci. USA* **103,** 1988–1993 (2006).
8. Fujii, H., Verslues, P. & Zhu, J. Identification of two protein kinases required for abscisic acid regulation of seed germination, root growth, and gene expression in *Arabidopsis. Plant Cell* **19,** 485–494 (2007).
9. Hirayama, T. & Shinozaki, K. Perception and transduction of abscisic acid signals: keys to the function of the versatile plant hormone ABA. *Trends Plant Sci.* **12,** 343–351 (2007).
10. Park, S. *et al.* Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324,** 1068–1071 (2009).
11. Pennisi, E. Plant biology. Stressed out over a stress hormone. *Science* **324,** 1012–1013 (2009).
12. Ma, Y. *et al.* Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324,** 1064–1068 (2009).
13. Lee, B. & Richards, F. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55,** 379–400 (1971).
14. Tanokura, M. & Yamada, K. Heat capacity and entropy changes of calmodulin induced by calcium binding. *J. Biochem.* **95,** 643–649 (1984).
15. Tanokura, M. & Yamada, K. A calorimetric study of Ca^{2+} binding by wheat germ calmodulin. Regulatory steps driven by entropy. *J. Biol. Chem.* **268,** 7090–7092 (1993).
16. Das, A., Helps, N., Cohen, P. & Barford, D. Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 Å resolution. *EMBO J.* **15,** 6798–6809 (1996).
17. Meyer, K., Leube, M. & Grill, E. A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. *Science* **264,** 1452–1455 (1994).
18. Leung, J. *et al. Arabidopsis* ABA response gene ABI1: features of a calcium-modulated protein phosphatase. *Science* **264,** 1448–1452 (1994).
19. Koornneef, M., Reuling, G. & Karssen, C. The isolation and characterization of abscisic acid-insensitive mutants of *Arabidopsis thaliana. Physiol. Plant.* **61,** 377–383 (1984).

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

Acknowledgements The synchrotron-radiation experiments were performed at BL5A in the Photon Factory (2008S2-001). This work was supported by the Targeted Proteins Research Program (TPRP) of the Ministry of Education, Culture, Sports, Science, and Technology, Japan. We thank Y. Sakaguchi (DKSH Japan K.K.) and Y. Asami (TA Instruments Japan Inc.) for the ITC measurements and T. Mitani (GE Healthcare Japan Corp.) for the SPR experiments.

Author Contributions M. Tanokura conceived and designed the project. K.-i.M., T.M., Y.S. and K. Kubota performed the construct design, subcloning, protein expression, purification, crystallization, structure determination and all biochemical assays. H.-J.K., A.A., Y.M, M. Takahashi and Y.Z. also assisted in the subcloning, protein expression, purification and crystallization. Y.F., T.Y. and K. Kodaira performed the cloning. K.-i.M., T.M., Y.S, K. Kubota, K.Y.-S. and M. Tanokura wrote the manuscript and M. Tanokura edited the manuscript.

Author Information Structure coordinates and structural factors are deposited in the Protein Data Bank under accession numbers 3JRS (PYL1–(+)-ABA) and 3JRQ (PYL1–(+)-ABA–ABI1). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M. Tanokura (amtanok@mail.ecc.u-tokyo.ac.jp).

METHODS

Protein expression and purification. The PYL1 construct composed of 8–211 residues (PYL1(8–211)) was used for the crystallization of PYL1–(+)-ABA, and was also applied to all biochemical assays, such as the ITC, SPR, pull-down binding assay and ABI1 inhibitory assay. The shorter construct with 28–210 residues (PYL1(28–210)) was used for the structure analysis of the PYL1–(+)-ABA–ABI1 complex and the ABA-binding analysis by NMR. The cDNA sequences corresponding to each PYL1 construct were inserted into the NdeI/BamHI site of the pET28a vector (Novagen) and transformed into *Escherichia coli* Rosetta(DE3) for protein expressions. The recombinant proteins were purified by an Ni-Sepharose resin (GE Healthcare) and were then excised from the histidine tag by thrombin. Further purification was performed by ResourceQ (GE Healthcare). The Se-Met derivative of PYL1(8–211) was overexpressed using the methionine biosynthesis inhibition method and then purified by the same method as the native protein.

The cDNA fragments of ABI1 (residues 1–124, 125–429 and 1–434) were cloned into the BamHI/NotI site of the pGEX6P-3 vector. The fragment of 125–429 was also cloned into the NdeI/BamHI site of pET28a. Each vector was transformed into *E. coli* Rosetta(DE3) for protein expressions. GST-fused ABI1 constructs were purified using glutathione Sepharose resin (GE Healthcare) and were then purified by ResourceQ. For crystallization of the PYL1–(+)-ABA–ABI1 complex, histidine-tagged ABI1(125–429) was purified by an Ni-NTA resin (QIAGEN) and then PYL1(28–210) in 10 mM Tris-HCl, pH 8.0, 200 mM NaCl, 1 mM DTT and 5 mM (+)-ABA/(+/−)-ABA was eluted into the column to form the complex. The complex was eluted with 10 mM Tris-HCl, pH 8.0, 200 mM NaCl and 200 mM imidazole. Further purification was performed by ResourceQ and Superdex 75 10/30 (GE Healthcare).

Crystallization and data collection. Crystallizations of PYL1–(+)-ABA and PYL1–(+)-ABA–ABI1 were carried out at 20 °C by the sitting-drop vapour-diffusion method. The Se-Met derivative crystals of PYL1 were obtained under the same reservoir solution conditions as the native protein.

The crystals of PYL1–(+)-ABA were transferred into the reservoir solution containing 26% PEG3350 as a cryoprotectant and flash-cooled at 95 K. X-ray diffraction data were collected on the BL-5A beamline at the Photon Factory. All X-ray diffraction data were integrated and scaled using the program XDS[20]. The native crystal belongs to the space group $P4_12_12$ with unit cell parameters of $a = b = 97.73$ and $c = 135.27$ Å and a Matthews coefficient of 2.25 Å^3 Da^{-1} (ref. 21), and the parameters of the Se-Met derivative crystal were almost completely consistent with those of the native one.

The crystals of PYL1–(+)-ABA–ABI1 were transferred into the reservoir solution containing 20% ethylene glycol as a cryoprotectant and flash-cooled at 95 K. X-ray diffraction data were collected on the NW-12A beamline at the Photon Factory. X-ray diffraction data were integrated and scaled using the program XDS[20]. The crystal of the PYL1–(+)-ABA–ABI1 complex belongs to the space group $P2_1$ with unit cell parameters of $a = 49.01$, $b = 60.62$, $c = 84.96$ Å and $\beta = 104.56^\circ$ with a Matthews coefficient of 2.24 Å^3 Da^{-1}.

Structure determination and refinement. The structure of PYL1–(+)-ABA was determined by the multiwavelength anomalous dispersion method using the X-ray diffraction data sets of the Se-Met derivative crystal. We found six of nine selenium sites in the asymmetric unit using the program SHELXD[22]. Refinement of the selenium sites and phasing were performed using the program SHARP[23]. The initial model of the Se-Met derivative of PYL1 was built automatically using the program Resolve[24]. The model was refined and rebuilt using the X-ray diffraction data of the native crystal by the programs Refmac5[25] and XtalView[26]. The structure of the PYL1–(+)-ABA–ABI1 complex was determined by the molecular replacement method using the coordinates of PYL1 and human PP2C protein[16] and was refined and rebuilt by the same procedure as used for the structure determination of PYL1–(+)-ABA. The data collection and refinement statistics are summarized in Supplementary Table 2. In the Ramachandran plot using Rampage[27], no residues were in the outer region.

ABA-binding assays. ABA-binding assays were performed by ITC and SPR. The ITC experiments were studied at 20 °C using an iTC$_{200}$ calorimeter (Microcal) and a nano ITC^{2G} (TA Instruments). The solvent conditions were 20 mM phosphate buffer, pH 7.0, 100 mM NaCl, 1 mM TCEP and 0.5% DMSO. All ITC experiments were designed as a series of 20 injections of 6 mM ABA into 250 μM PYL1. The obtained data were corrected by subtracting the mixing enthalpies of the (+)-ABA solution into protein-free solution and fitted by Origin for ITC (Microcal) with a 1:1 binding model. The ITC experiments were also performed using a nano ITC^{2G} (TA Instruments) according to the manufacturer's protocol.

The SPR measurements were performed by a Biacore T100 system (GE Healthcare) and sensor chip SA (GE Healthcare). The protein samples were biotinylated by being mixed with a 1:1 molar ratio of EZ-Link sulpho-NHS-LC-LC-biotin (Pierce), and were then separated from non-reacted biotin by Superdex75. The labelled proteins were captured on the sensor chips with 2,000–5,000 resonance units in the running buffer (10 mM HEPES, pH 7.5, 150 mM NaCl and 0.05% (v/v) surfactant P20). Binding analyses were performed using a range of (+)-ABA concentrations (31–500 μM) dissolved in the running buffer containing 0.025% DMSO at 30 μl min^{-1}. The data were analysed by Biacore T100 Evaluation software 2.0.1 using a 1:1 binding model.

NMR measurements. For NMR measurements, ^{15}N-labelled PYL1(28–210) was produced in M9 minimal medium containing 0.1% $^{15}NH_4Cl$. The stable-isotope-labelled protein was purified by the same method as the non-labelled protein, and was then dissolved at 0.2 mM in 92% H_2O, 8% D_2O, 0.2% DMSO and 10 mM phosphate buffer at pH 7.0. The TROSY-based ^{1}H-^{15}N HSQC spectra of PYL1 in the absence and presence of 0.5 mM (+)-ABA or (−)-ABA were measured at 20 °C on a Varian Unity INOVA-600 spectrometer equipped with a cryogenic probe. All NMR spectra were acquired by a BioPack module on the operation system VNMR, processed by the program NMRPipe[28] and visualized using the program NMRViewJ[29].

Pull-down binding assays. The wild-type and mutants of PYL1(8–211) (300 pmol) were incubated with the glutathione resins that bound the ABI1 proteins (300 pmol) in a pull-down buffer containing 20 mM HEPES, pH 8.0, 150 mM NaCl, 1 mM TCEP and 5 mM $MgCl_2$, 10 μM (+)-ABA. The proteins remaining on the resins were then eluted with 50 mM Tris-HCl, pH 8.0 and 10 mM reduced glutathione. The eluted proteins were separated by SDS–PAGE and visualized by Coomassie staining. In the mutation analyses, the concentrations of protein bands were quantified using ImageJ software (NIH) and the concentrations of the PYL1 proteins were normalized by those of the ABI1 proteins in the same lane of SDS–PAGE.

PP2C-activity assays. The PP2C activity was assayed using a serine/threonine phosphatase assay system (Promega) according to the manufacturer's protocols. 0.1 μM GST-fused ABI1(125–429) and its mutants were reacted with 0.1 mM RRA(pT)VA phosphopeptide substrate, 1.5 μM PYL1(8–211) and several concentrations (0–30 μM) of (+)-ABA in a reaction buffer (50 mM HEPES, pH 8.0, 5 mM $MgCl_2$, 40 μM TCEP and 0.1 mg ml^{-1} BSA).

20. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
21. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).
22. Schneider, T. & Sheldrick, G. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
23. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr. D* **59**, 2023–2030 (2003).
24. Terwilliger, T. Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr. D* **59**, 38–44 (2003).
25. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
26. McRee, D. E. XtalView/Xfit–A versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
27. Lovell, S. C. *et al.* Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins* **50**, 437–450 (2003).
28. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
29. Johnson, B. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313–352 (2004).

Vol 462 | 3 December 2009 | doi:10.1038/nature08595 **nature**

Detection of sequential polyubiquitylation on a millisecond timescale

Nathan W. Pierce[1], Gary Kleiger[1], Shu-ou Shan[2]* & Raymond J. Deshaies[1]*

The pathway by which ubiquitin chains are generated on substrate through a cascade of enzymes consisting of an E1, E2 and E3 remains unclear. Multiple distinct models involving chain assembly on E2 or substrate have been proposed. However, the speed and complexity of the reaction have precluded direct experimental tests to distinguish between potential pathways. Here we introduce new theoretical and experimental methodologies to address both limitations. A quantitative framework based on product distribution predicts that the really interesting new gene (RING) E3 enzymes SCF$^{\mathrm{Cdc4}}$ and SCF$^{\beta\text{-TrCP}}$ work with the E2 Cdc34 to build polyubiquitin chains on substrates by sequential transfers of single ubiquitins. Measurements with millisecond time resolution directly demonstrate that substrate polyubiquitylation proceeds sequentially. Our results present an unprecedented glimpse into the mechanism of RING ubiquitin ligases and illuminate the quantitative parameters that underlie the rate and pattern of ubiquitin chain assembly.

Attachment of a polyubiquitin chain with at least four ubiquitins linked together through their lysine 48 residue (Lys 48) targets proteins to the proteasome for degradation[1]. A cascade of three enzymes performs the synthesis of polyubiquitin chains: a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2) and a ubiquitin ligase (E3)[2]. RING E3 enzymes catalyse the direct transfer of ubiquitin from an E2 to a lysine on a target protein[3]. SCF$^{\mathrm{Cdc4}}$ is the founding member of the largest family of E3 enzymes—the cullin–RING ubiquitin ligases that may comprise most of all human ubiquitin ligases[3]. Thus, unravelling the mechanism of SCF will have broad functional ramifications for the preponderance of human E3 enzymes.

Different pathways for ubiquitin chain assembly by RING E3 enzymes have been envisioned based on indirect evidence. On the one hand, Cdc34–SCF ubiquitylates substrates bearing a single ubiquitin significantly faster than non-ubiquitylated substrates[4,5], suggesting that it processively builds polyubiquitin chains on substrates with an initial slow transfer of ubiquitin followed by rapid elongation into a Lys-48-linked polyubiquitin chain. On the other hand, the E2 Ube2g2, a close relative of Cdc34, collaborates with the E3 gp78 to build a polyubiquitin chain on its active site cysteine that can be transferred en bloc to substrate[6,7]. Various permutations of the en bloc mechanism have been entertained, in which the chain is built either from proximal to distal end or vice versa[8–10]. Owing to the rapid speed of ubiquitin chain synthesis, intermediates that would reveal the underlying pathway cannot be kinetically resolved. Thus, it has not been possible to establish definitively the pathway of chain assembly for any RING E3. Here we introduce new theoretical and experimental methodologies to address both limitations.

Quantitative analysis of product distribution

Processivity emerges from the relationships between reaction and dissociation rates for different product intermediates[11]. To quantify the processivity of SCF, we established an assay capable of simultaneously monitoring the concentrations of substrate and its different ubiquitylated product intermediates. Our assay consisted of an engineered phosphopeptide substrate (CYCE) derived from human cyclin E1 (also known as CCNE1) and purified *Saccharomyces cerevisiae* Cdc34–SCF$^{\mathrm{Cdc4}}$ (refs 5, 12). CYCE was selected because it is a defined, chemically homogeneous substrate that binds with high affinity to the substrate-binding pocket of SCF$^{\mathrm{Cdc4}}$ (refs 12, 13). Moreover, intact cyclin E is a substrate of SCF$^{\mathrm{Cdc4}}$ *in vivo*[14] and the degron from CYCE can support turnover *in vivo* of an engineered substrate, Sic1, from which the endogenous degrons have been eliminated[12]. To examine the simplest system that recapitulated the processive behaviour of Cdc34–SCF$^{\mathrm{Cdc4}}$, we focused on single turnover reaction conditions containing an excess of SCF$^{\mathrm{Cdc4}}$ over radiolabelled CYCE. We initiated reactions by combining two pre-incubated mixtures: the 'charged E2' mixture containing ubiquitin, E1, ATP and Cdc34 was pre-incubated for 2 minutes to ensure the formation of saturating concentrations of Cdc34–ubiquitin thioesters (E2–Ub); and the 'substrate–ligase' mixture containing SCF and radiolabelled substrate was pre-incubated to ensure the formation of an enzyme–substrate complex. To maximize resolution of ubiquitin conjugates, the reaction products were fractionated on long SDS-polyacrylamide gels. Consistent with previous assays performed with Sic1 (ref. 5), conjugation of the Nedd8 homologue Rub1 to the Cdc53 subunit of budding yeast SCF$^{\mathrm{Cdc4}}$ did not alter the ubiquitylation kinetics of CYCE (Supplementary Fig. 1); thus all subsequent SCF$^{\mathrm{Cdc4}}$ assays were performed with unmodified E3.

Under these reaction conditions, CYCE was extensively polyubiquitylated by Cdc34–SCF$^{\mathrm{Cdc4}}$ within 30 s (reaction 1, Fig. 1a), and products containing at least six ubiquitins were visible within 10 s (Supplementary Fig. 1). Thus, with the time resolution offered by manual mixing, it was not apparent whether ubiquitin conjugates were formed by multiple sequential transfers of monoubiquitin or by en bloc transfer of pre-formed chains. However, we reasoned that quantitative analysis of the length distribution of polyubiquitin chains attached to CYCE during a single encounter with SCF might provide clues to the pathway of chain assembly. To determine the length distribution, we performed reactions with 1,000-fold excess chase of the unlabelled substrate peptide added to the 'charged E2' mixture (reaction 2, Fig. 1a). Under these conditions, radiolabelled substrate pre-bound to SCF was rapidly ubiquitylated, but upon dissociation

---

[1]Howard Hughes Medical Institute, Division of Biology, MC 156-29, [2]Division of Chemistry and Chemical Engineering, MC 147-75, California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA.
*These authors contributed equally to this work.

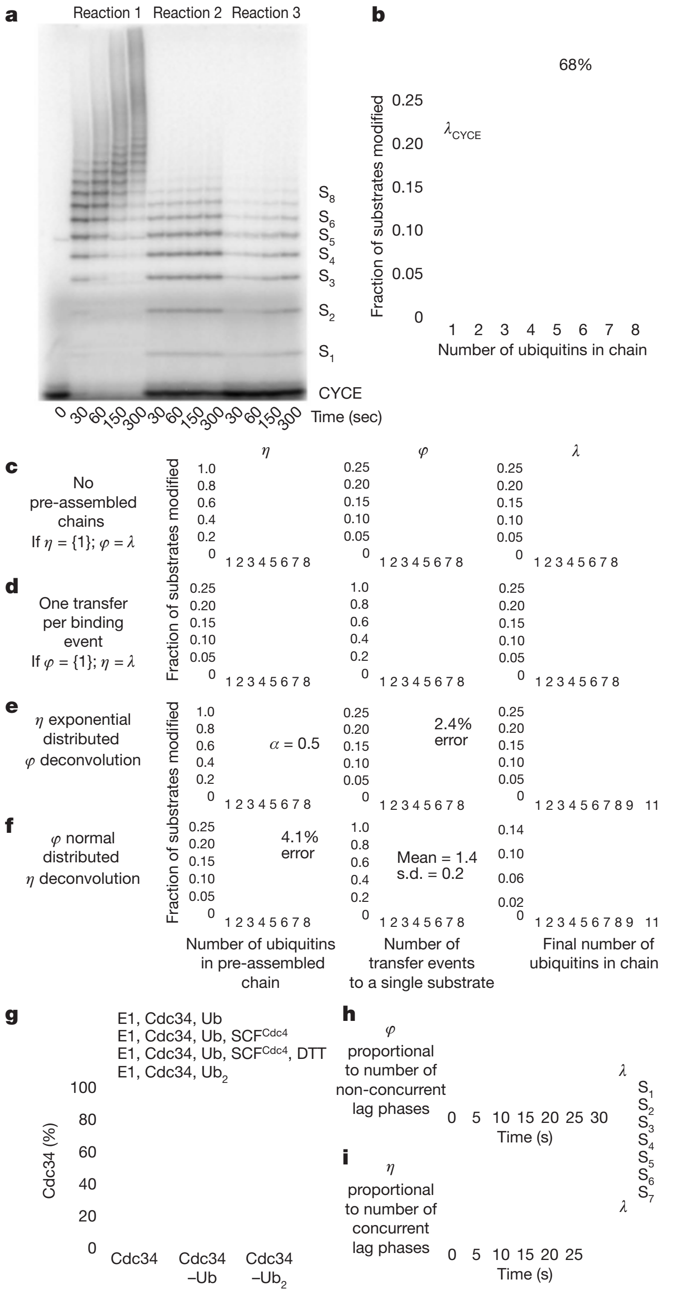


Figure 1 | Final product distribution for SCF^{Cdc4} and CYCE. **a**, In reaction 1, pre-incubated ^{32}P-labelled CYCE and SCF^{Cdc4} were added to the charged E2 mix. In reactions 2 and 3, excess unlabelled CYCE was pre-incubated with charged E2 mix and labelled CYCE, respectively. **b**, The single-encounter polyubiquitin chain-length distribution, λ_{CYCE}. Error bars, ± s.d. ($n = 3$). **c**, If $\eta(1) = 100\%$, then $\varphi = \lambda$. **d**, If $\varphi(1) = 100\%$, then $\eta = \lambda$. **e**, Deconvolution of λ_{CYCE} and exponentially distributed η. **f**, Deconvolution of λ_{CYCE} and normal distributed φ. **g**, Mass spectrometry of Cdc34 thioesterified for 2 min with indicated components. **h**, Simulated kinetics $\eta(1) = 100\%$. **i**, Simulated kinetics $\varphi(1) = 100\%$.

further ubiquitylation occurred at a significantly reduced rate due to competition from the chase peptide. To evaluate the effectiveness of the chase, we performed a parallel reaction in which the chase peptide was added to the 'substrate–ligase' mixture before initiation (reaction 3, Fig. 1a). The distribution of products in reaction 3 was subtracted from the distribution of products in reaction 2 at each time point (Supplementary Fig. 2) to yield the average distribution for substrate, λ (Fig. 1b). Three main points were highlighted by these experiments. First, it is evident from reaction 2 that the single-encounter reaction was complete within 30 s. Second, 72% of CYCE encounters with SCF^{Cdc4} resulted in no ubiquitin modification (Fig. 1a and Supplementary Fig. 2). Third, of those substrates that were modified, 68% of CYCE acquired a polyubiquitin chain with four or more ubiquitins (Fig. 1b).

We next sought to develop a quantitative framework to address whether the experimentally determined product distribution λ_{CYCE} (Fig. 1b) places constraints on the potential pathways of ubiquitin chain assembly. We considered three hypothetical situations. First, we imagined that only monoubiquitin was attached in each transfer event (Fig. 1c, 'sequential'). Binning all of the transfer events per substrate gave the transfer distribution φ, which in this case would equal λ. Second, we imagined the other extreme in which only one transfer event occurs per substrate (Fig. 1d, 'en bloc'). In this case, λ would be equal to the distribution of pre-assembled polyubiquitin chains thioesterified to E2, which we named η. Third, we considered permutations that combined sequential and en bloc transfers. For example, there are eight possible ways of making substrate modified with four ubiquitins (S_n, where $n = 4$), including two transfers of diubiquitin or transfer of monoubiquitin followed by transfer of triubiquitin, etc. From this analysis, a key point emerged: regardless of the type of distribution we started with, the family of η and φ distributions compatible with λ_{CYCE} (see Supplementary Methods) was restricted to extreme cases where either η or φ was nearly equal to λ_{CYCE} (Fig. 1e, f and Supplementary Figs 3–7). Therefore, most substrates either underwent one transfer per binding event or received a single ubiquitin per transfer event. Thus, accurately measuring product distribution constrained the number of possible pathways that could give rise to the reaction products we observed.

As a first test of whether ubiquitins were transferred all at once or sequentially, we measured the distribution of polyubiquitin chain lengths present on the active site of Cdc34 in the presence or absence of SCF^{Cdc4} by intact mass spectrometry. Cdc34 subjected to our standard 'charged E2' pre-incubation was completely converted to thioesters carrying a single ubiquitin (Cdc34–Ub; Fig. 1g and Supplementary Fig. 10). In the presence of SCF^{Cdc4}, 89% of Cdc34 was detected as Cdc34–Ub and 11% was unmodified; no Cdc34 species with more than one ubiquitin attached was detected. A control experiment run with diubiquitin confirmed that our assay was able to detect diubiquitin chains thioesterified to the active site of Cdc34 (Cdc34–Ub$_2$; Supplementary Fig. 11), but charging of Cdc34 with diubiquitin occurs with poor efficiency (~20%). Thus, our analysis of the product distribution λ coupled with measurement of the ubiquitin population thioesterified to Cdc34 under our reaction conditions (an estimate of η) strongly predicts that Cdc34–SCF^{Cdc4} assembled ubiquitin chains on substrate primarily by sequential transfers of single ubiquitin molecules.

Millisecond kinetics of SCF

As a second, more definitive test of the hypothesis stated above, we sought to measure directly how the product distribution (Fig. 1b) developed as a function of time. During a single encounter between a RING ubiquitin ligase and substrate, each intermediate should either undergo a transfer event or dissociate. If monoubiquitin composed 100% of η as in Fig. 1c, the products of the reaction should appear sequentially in time starting with S$_1$ and followed by S$_2$, then S$_3$, etc. Thus, the appearance of each sequential product should be delayed by a 'lag' phase (Fig. 1h). In contrast, if a single transfer composed 100% of φ as in Fig. 1d, then the pattern of ubiquitin chains attached to substrate at the earliest time-points should reveal the distribution of pre-assembled chains thioesterified to Cdc34 (Fig. 1i). Thus, products of increasing mass should accumulate sequentially if chain synthesis is sequential, but should accumulate contemporaneously if chains are transferred en bloc. Therefore, with sufficient time resolution, a single-encounter experiment would provide definitive data to distinguish between the alternative models. To achieve the necessary time resolution, we performed our single-encounter reactions on a quench flow apparatus that allowed us to take measurements on a timescale ranging from 10 ms to 30 s (Fig. 2a). To facilitate quantification of S$_2$

and S_5 in the CYCE reaction, the same reaction from Fig. 2a was fractionated on a gel with different resolving capabilities (Supplementary Fig. 12). Three major conclusions arose from these experiments. First, the product CYCE–Ub (S_1) was formed starting at the earliest time points (10–20 ms) without a lag phase, indicating that E2–Ub binding to SCF was rapid. This is consistent with stopped-flow measurements performed with $SCF^{\beta\text{-TrCP}}$ and hCdc34 (ref. 15). Second, each new ubiquitylated product appeared sequentially with non-concurrent lag phases (Fig. 2a, b and Supplementary Fig. 12). Third, the early reaction products S_1–S_3 'overshot' their final levels, indicating that these reaction intermediates serve as templates for the formation of subsequent products, supporting the model that polyubiquitin chains are built from multiple transfer events (Supplementary Fig. 16). Combined with the constraints on η and φ calculated above as well as our direct evaluation of the Cdc34–Ub pool (Fig. 1g), these data demonstrate that the underlying kinetic mechanism of our system was principally derived from sequential transfers of single ubiquitins.

To ensure that our conclusions were not an artefact of the reaction design, we changed the order of addition in our reactions. SCF^{Cdc4} was pre-incubated with the 'charged E2' mixture for 2 min (in which case 89% of Cdc34 is present in thioesterified form; Fig. 1g) and reactions were initiated by combining with radiolabelled CYCE. Products appeared after non-concurrent lag phases of increasing duration (Fig. 2c), analogous to that observed when the reaction was initiated by addition of Cdc34–Ub to CYCE prebound to SCF^{Cdc4} (Fig. 2a). Thus, regardless of whether CYCE first encountered Cdc34–Ub–SCF or Cdc34–Ub encountered CYCE–SCF, single ubiquitins were transferred to substrate in a sequential manner. Interestingly, reactions initiated by addition of CYCE were delayed compared with those initiated by addition of Cdc34–Ub, indicating that Cdc34–Ub productively associates with SCF^{Cdc4} faster than does CYCE.

$SCF^{\beta\text{-TrCP}}$ is sequentially processive

We next sought to test whether the sequential processive chain assembly we observed for SCF^{Cdc4} is unique or illuminates a general principle of SCF ubiquitin ligase mechanism. To address this issue, we evaluated ubiquitylation of a phosphopeptide derived from β-catenin (β-Cat) by its cognate E2–E3 complex, hCdc34 and human $SCF^{\beta\text{-TrCP}}$. Nedd8-conjugated E3 (N8-$SCF^{\beta\text{-TrCP}}$) was used for these experiments, because previous work demonstrated a potent stimulation of β-Cat ubiquitylation upon Nedd8 conjugation[4]. As was seen with CYCE–SCF^{Cdc4}, β-Cat was rapidly modified by N8-$SCF^{\beta\text{-TrCP}}$ and it was not possible to resolve intermediates in chain assembly by manual mixing[4] (Fig. 3a). Quantification of product distribution $\lambda_{\beta\text{-Cat}}$ revealed that 6% of β-Cat molecules were modified in a single encounter with N8-$SCF^{\beta\text{-TrCP}}$, of which 85% received at least four ubiquitins (Fig. 3b and Supplementary Fig. 2). Distribution analysis of $\lambda_{\beta\text{-Cat}}$ (Fig. 3c) and kinetic resolution of β-Cat ubiquitylation by quench-flow (Fig. 3d, e) revealed sequential appearance of intermediates analogous to those observed with CYCE ubiquitylation by SCF^{Cdc4}.

Although the general behaviour of SCF^{Cdc4} and N8-$SCF^{\beta\text{-TrCP}}$ were similar, the enzymes differed in the extent to which they

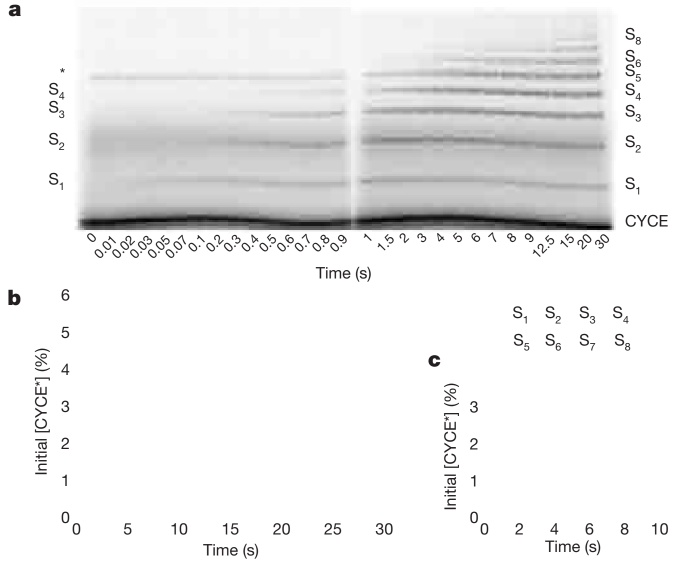


Figure 2 | Millisecond kinetics of a single-encounter reaction reveal sequential processivity. **a**, To achieve millisecond temporal resolution, CYCE reactions were performed on a quench flow apparatus and products were evaluated by SDS–polyacrylamide gel electrophoresis and phosphorimaging. The reaction scheme matched reaction 2 of Fig. 1a. The asterisk marks a contaminant. S_n refers to CYCE modified with n ubiquitins. **b**, Quantification shows successively longer lag phases for each additional ubiquitin added in the chain. The data were fitted using closed-form solutions refined by global regression analysis to a model with $\eta = 1$. The error bars represent the range of values ($n = 2$). **c**, SCF^{Cdc4} was pre-incubated with charged E2 mix and combined with ^{32}P-labelled CYCE.

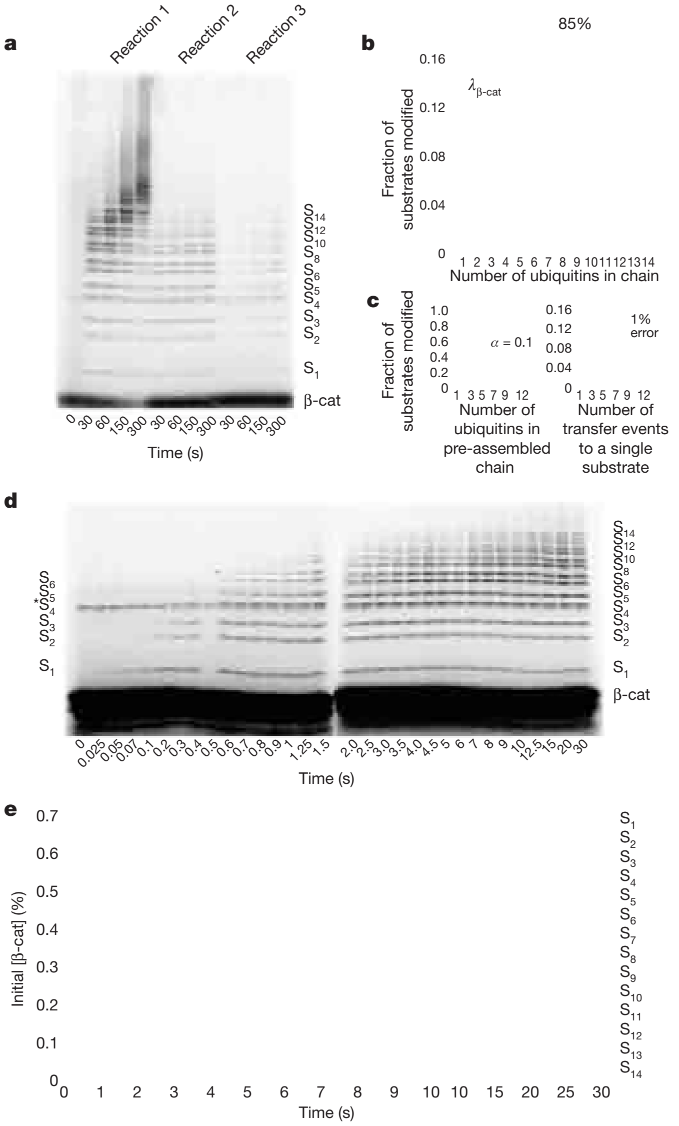


Figure 3 | Human Cdc34–$SCF^{\beta\text{-TrCP}}$ is sequentially processive. **a**, The same as Fig. 1a, except that human Cdc34 and Nedd8-conjugated $SCF^{\beta\text{-TrCP}}$ were assayed with ^{32}P-labelled β-Cat substrate. **b**, Product distribution ($\lambda_{\beta\text{-Cat}}$) was quantified as in Fig. 1b. Error bars, ± s.d. ($n = 3$). **c**, The Poisson distribution of φ using $\lambda_{\beta\text{-Cat}}$ that deviated the most from $\varphi(1) = 100\%$ within our set error bounds with $\alpha = 0.2$. **d**, β-Cat reactions with the scheme of reaction 2 (Fig. 1a) performed on a quench flow apparatus. **e**, Quantification shows successively lengthening lag phases for each additional ubiquitin added in the chain. The data were fitted as in Fig. 2b.

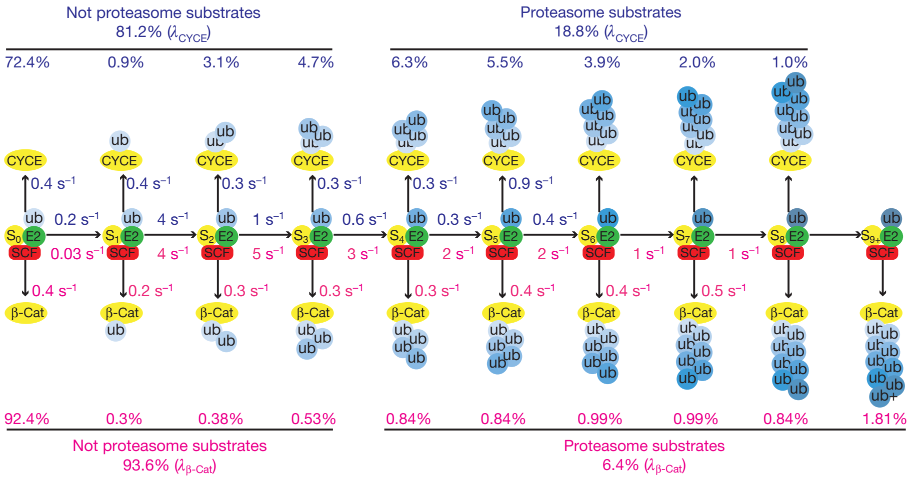


Figure 4 | Kinetic basis for Cdc34–SCF processivity. The millisecond kinetics of a single-encounter reaction were fitted to a sequential model revealing estimates for individual transfer and dissociation rates for each intermediate in the generation of polyubiquitylated CYCE (blue numbers) and β-Cat (red numbers) products. The percentages listed above or below each product were those from the final product distributions (λ) shown in Figs 1b and 3b.

converted bound substrate to product and elongated ubiquitin chains. Using a kinetic model in which monoubiquitin composed 100% of η, we were able to use methods borrowed from the study of nucleic acid polymerases[16] to extrapolate estimates for the individual reaction and dissociation rate constants from our single-encounter quench-flow experiments (Fig. 4 and Supplementary Figs 13–15).

Functional implications of our model

The model shown in Fig. 4 reveals the kinetic basis of processive polyubiquitin chain synthesis by budding yeast Cdc34–$\mathrm{SCF}^{\mathrm{Cdc4}}$ and human Cdc34–$\mathrm{SCF}^{\beta\text{-TrCP}}$, and accounts for the differences in their behaviour. Most encounters of substrate and SCF are unproductive because k_{off} is faster than k_{Ub1}. This is particularly exaggerated for β-Cat owing to its low value for k_{Ub1}. Once a single ubiquitin is attached, most substrates are committed to polyubiquitylation because of the drastic increase in k_{Ub2} relative to a nearly constant k_{off}. This gives rise to the high percentage of modified substrates with four or more ubiquitins in their chain (68% for CYCE and 85% for β-Cat). The overall chain length is limited by the progressive decrease in transfer rates (k_{Ubn}) as the chain becomes longer matched against the relatively constant rate at which product intermediates dissociate. This reduction in transfer rate most likely arises because the distal end of the flexible chain samples a progressively larger volume as it increases in length[17]. The longer chains on β-Cat are a result of a less dramatic decline in k_{Ubn} after the second ubiquitin is attached. We do not understand the basis for this difference. Meanwhile, the constant rate of dissociation for both CYCE and β-Cat implies that ubiquitin chains of increasing length do not change the intrinsic affinity of these substrates for SCF.

Casual inspection of our model suggests that modest changes in the ratio $k_{\mathrm{Ub1}}/k_{\mathrm{off}}$ for the first step would substantially alter the fraction of substrate that acquires a chain of at least four ubiquitins in a single encounter with SCF. This in turn provides a simple basis for SCF to modulate a substrate's degradation half-life (that is, the larger k_{off} is or smaller k_{Ub1} is, the lower the probability that a substrate is modified in a single encounter with SCF, which would translate to a longer half-life). Comparison of CYCE and β-Cat, which have distinct $k_{\mathrm{Ub1}}/k_{\mathrm{off}}$ ratios, underscores how the efficiency and pattern of substrate ubiquitylation can be tuned by these parameters. Despite these differences, it is remarkable how similar the reaction parameters are for two different enzymes from organisms separated by over 1 billion years of evolution. In both cases, k_{off} was about $0.4\,\mathrm{s}^{-1}$ and the fastest rate of ubiquitin chain elongation was $4\text{–}5\,\mathrm{s}^{-1}$. This suggests that true substrates are tuned to dissociate within a few seconds and that a transfer rate of $5\,\mathrm{s}^{-1}$ may be imposed by a conserved rate-limiting step. It will be of great interest to determine what molecular event enforces this speed limit.

We conclude that polyubiquitin chains are built on SCF substrates by sequential transfers of single ubiquitins. We establish a mechanistic framework that can be applied to other cullin–RING ubiquitin ligases and RING ubiquitin ligases to obtain individual rate constants for substrate dissociation and ubiquitin transfer at each step in the process of chain assembly. Our model indicates that the processivity, efficiency and pattern of ubiquitylation is governed by the sharp discontinuity in rates between the first transfer and subsequent transfers, contrasted with the shared dissociation rate among substrate and product intermediates.

METHODS SUMMARY

Proteins. CYCE and β-Cat phosphopeptide were purchased from New England Peptide. Ubiquitin and K48 diubiquitin were purchased from Boston Biochem. Uba1 and $\mathrm{SCF}^{\mathrm{Cdc4}}$ were prepared and purified as described[5]. Full-length yeast Cdc34 was purified as described[18]. His7–Rub1 was purified from *Escherichia coli* inclusion bodies[4] and human E1, UbcH3B (hCdc34) and Nedd8-$\mathrm{SCF}^{\beta\text{-TrCP}}$ were prepared and purified as described[4]. Yeast Ubc12 and Ula1–Uba3 were purified as described[19]. Rub1, Ubc12, Ula1–Uba3 and ATP were incubated with immobilized $\mathrm{SCF}^{\mathrm{Cdc4}}$ to make Rub1-conjugated $\mathrm{SCF}^{\mathrm{Cdc4}}$. Protein kinase A was purchased from New England Biolabs.

Ubiquitylation assay. CYCE (200 nM) or β-Cat (2 μM) was incubated with [γ-^{32}P]ATP (132 nM) and protein kinase A for 45 min at 30 °C to make radiolabelled CYCE or β-Cat. Yeast ubiquitylation reactions contained ATP (2 mM), ubiquitin (60 μM), Uba1 (0.8 μM), Cdc34 (10 μM), $\mathrm{SCF}^{\mathrm{Cdc4}}$ (150 nM) and radiolabelled CYCE (10 nM).

Human ubiquitylation reactions contained ATP (2 mM), ubiquitin (60 μM), E1 (1 μM), Cdc34 (10 μM), $\mathrm{SCF}^{\beta\text{-TrCP}}$ (500 nM) and radiolabelled β-Cat (100 nM). As indicated, single-encounter reactions contained an unlabelled CYCE chase (10 μM) or β-Cat chase (100 μM). Millisecond reactions were performed on a quench flow apparatus (Kintek RQF-3 Rapid Quench Flow). Reactions contained a buffer previously described[20] at 23 °C. Reactions were quenched with SDS–polyacrylamide gel electrophoresis buffer with β-mercaptoethanol and run on 20-cm 5–20% tricine gels (CYCE) or glycine gels (β-Cat) that were quantified with a phosphor screen (Molecular Devices). Thioester formation assays contained Cdc34 (10 μM), Uba1 (1 μM), ATP (2 mM), ubiquitin or K48 diubiquitin (15 μM) and $\mathrm{SCF}^{\mathrm{Cdc4}}$ (100 nM) as indicated. After 2 min, reactions were stopped with excess 5% acetic acid and analysed on an Agilent LC-MSD.

Analysis. Deconvolutions and regression were performed in Matlab. Global fitting was performed with KinTek Global Kinetic Explorer. Mass spectrometry data were processed using the Chemstation software package.

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

Received 25 September; accepted 19 October 2009.

1. Thrower, J. S. *et al.* Recognition of the polyubiquitin proteolytic signal. *EMBO J.* 19, 94–102 (2000).

2. Dye, B. T. & Schulman, B. Structural mechanisms underlying posttranslational modification by ubiquitin-like proteins. *Annu. Rev. Biophys. Biomol. Struct.* **36,** 131–150 (2007).
3. Petroski, M. & Deshaies, R. Function and regulation of cullin–RING ubiquitin ligases. *Nature Rev. Mol. Cell Biol.* **6,** 9–20 (2005).
4. Saha, A. & Deshaies, R. Multimodal activation of the ubiquitin ligase SCF by Nedd8 conjugation. *Mol. Cell* **32,** 21–31 (2008).
5. Petroski, M. & Deshaies, R. Mechanism of lysine 48-linked ubiquitin-chain synthesis by the cullin-RING ubiquitin-ligase complex SCF-Cdc34. *Cell* **123,** 1107–1120 (2005).
6. Ravid, T. & Hochstrasser, M. Autoregulation of an E2 enzyme by ubiquitin-chain assembly on its catalytic residue. *Nature Cell Biol.* **9,** 422–427 (2007).
7. Li, W. *et al.* A ubiquitin ligase transfers preformed polyubiquitin chains from a conjugating enzyme to a substrate. *Nature* **446,** 333–337 (2007).
8. Hochstrasser, M. Lingering mysteries of ubiquitin-chain assembly. *Cell* **124,** 27–34 (2006).
9. Li, W. *et al.* Mechanistic insights into active site-associated polyubiquitination by the ubiquitin-conjugating enzyme Ube2g2. *Proc. Natl Acad. Sci. USA* **106,** 3722–3727 (2009).
10. Deshaies, R. J. & Joazeiro, C. A. RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.* **78,** 399–434 (2009).
11. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* Ch. 14 (Freeman, 1999).
12. Nash, P. *et al.* Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **414,** 514–521 (2001).
13. Orlicky, S. *et al.* Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase. *Cell* **112,** 243–256 (2003).
14. Strohmaier, H. *et al.* Human F-box protein hCdc4 targets cyclin E for proteolysis and is mutated in a breast cancer cell line. *Nature* **413,** 316–322 (2001).
15. Kleiger, G., Saha, A., Lewis, S., Kuhlman, B. & Deshaies, R. Rapid E2–E3 assembly and disassembly enable processive ubiquitylation of cullin-RING ubiquitin ligase substrates. *Cell*. (in the press).
16. Kati, W. M. *et al.* Mechanism and fidelity of HIV reverse transcriptase. *J. Biol. Chem.* **267,** 25988–25997 (1992).
17. Petroski, M. *et al.* Evaluation of a diffusion-driven mechanism for substrate ubiquitination by the SCF-Cdc34 ubiquitin ligase complex. *Mol. Cell* **24,** 523–534 (2006).
18. Feldman, R. M. *et al.* A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* **91,** 221–230 (1997).
19. Kamura, T. *et al.* The Rbx1 subunit of SCF and VHL E3 ubiquitin ligase activates Rub1 modification of cullins Cdc53 and Cul2. *Genes Dev.* **13,** 2928–2933 (1999).
20. Petroski, M. & Deshaies, R. *In vitro* reconstitution of SCF substrate ubiquitination with purified proteins. *Methods Enzymol.* **398,** 143–158 (2005).

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

Acknowledgements We thank J. Vielmetter for providing SCF^{Cdc4}, β-TrCP-Skp1 and human E1; S. Hess, R. L. J. Graham and the Proteome Exploration Laboratory for providing assistance with mass spectrometry of CYCE and Cdc34 thioester. We thank S. Schwarz, B. Schulman and G. Wu for gifts of reagents. We thank D. Sprinzak and all the members of the Deshaies and Shan laboratories for support and discussions. N.W.P. was supported by the Gordon Ross Fellowship and a National Institutes of Health Training Grant. R.J.D. is an Investigator of the Howard Hughes Medical Institute. This work was supported in part by National Institutes of Health GM065997.

Author Contributions N.W.P. performed all computational modelling and experiments, except G.K. performed the mass spectrometry experiments in Fig. 1g. N.W.P., R.J.D. and S.-o.S. conceived the experiments. N.W.P. and R.J.D. wrote the manuscript with editorial input from the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.J.D. (deshaies@caltech.edu).

METHODS

RDB 2289 with pGEX2-T His7–Rub1 was made by cloning a His7 tag in place of the GST tag in RDB 1436. To make Rub1-conjugated SCF^{Cdc4}, the procedure for isolating SCF^{Cdc4} was modified[20]. Instead of eluting SCF^{Cdc4} from Py-conjugated protein A beads after washing, 50 μM Rub1, 10 μM Ubc12, 1 μM Ula1–Uba3 and 2 mM ATP were incubated together overnight. The beads were washed before elution.

For all reactions, gels were dried and exposed to phosphor screens (Molecular Dynamics). Images were scanned and then quantified in ImageQuant using a rolling ball background subtraction. For each lane, every band was quantified as a percentage of the total signal in all bands.

The relation between η, φ and λ was mathematically analogous to the probability of the sum of multiple dice throws. However, the probability of throwing each number on the dice was a weighted normalized distribution (analogous to η) and the number of throws was also a weighted normalized distribution (analogous to φ). A distribution that is normalized sums to 1. Thus, λ equalled the weighted sum of multiple discrete convolutions of η with itself as governed by φ, as shown by example in Supplementary Fig. 3. Knowledge of λ and η allowed us to calculate φ by multiple weighted deconvolutions, as shown by example in Supplementary Fig. 4. This was true for calculating η from λ and φ, as shown by example in Supplementary Fig. 5a. If we assigned a distribution to η, we determined φ by deconvolutions with λ, and vice versa. Considering normalized distributions of η that only contained $\eta(1)$ and $\eta(2)$, exponential distributions, Poisson distributions and normal distributions, we varied parameters over a wide range and performed deconvolutions, as shown by example in Supplementary Fig. 5b. An exponential distribution is described by a single parameter, here called α. A Poisson distribution is also described by a single parameter, here called α. The normal distribution is described by two parameters, the mean and the standard deviation (s.d.). Parameters were varied starting at 0 and increasing by step sizes of 0.1 until they equalled 10. For the normal distribution, each value of the mean was held constant while the s.d. was varied. We sought the distribution that deviated most from $\eta(1) = 100\%$ whose φ did not contain values >1 or <0 and that when convoluted with φ, the sum of λ fell within 0.95 and 1.05, or an error rate of ± 5% was found. This was repeated for φ. These distributions are shown in Supplementary Figs 6 and 7. Random distributions were also considered (data not shown).

For mass spectrometry analysis, Uba1 (1 μM), Cdc34-Δ270 (10 μM) and ubiquitin or K48-linked di-ubiquitin (15 μM, Boston Biochem) were incubated for 2 min in reaction buffer (30 mM Tris, pH 7.5, 100 mM NaCl, 5 mM $MgCl_2$, 2 mM dithiothreitol and 2 mM ATP) in a volume of 10 μl both in the presence and absence of SCF (100 nM). Reactions were quenched by the addition of 90 μl 5% acetic acid. Quenching was verified by an order of addition reaction where E1 was left out of the initial incubation and was added after quenching. This resulted in 100% quenching of the thioester charging reaction. Separation of E2 thioesters in the presence of SCF was accomplished by the addition of 100 mM dithiothreitol after the 2 min incubation period. The dithiothreitol was incubated with the reaction mixture for 5 min followed by the addition of 90 μl of 5% acetic acid. Detection of proteins was done on an Agilent LC-MSD. Mass spectra were acquired in positive-ion mode, scanning from 500 to 1700 m/z. The electrospray voltage was set to 4 kV and the gas temperature in the spray chamber was maintained at 350 °C. A stationary phase, Zorbax 300SB C3 150 mm × 2.1 mm column was used for separation (Agilent). Mobile phase A was 0.2% formic acid; mobile phase B was 0.2% formic acid, 10% methanol and 90% acetonitrile. The flow rate was 0.200 ml min^{-1}. After a 25-min delay, the effluent was directed into the mass spectrometer. Linear gradients started with 5% mobile phase B and finished at 95% from 25–50 min. Data were processed using the ChemStation software package. The sequence of yeast Cdc34-Δ270 contains the amino acids from positions 1 to 270 of the yeast Cdc34 sequence followed by the sequence ARPLHHHHHH, yielding a theoretical molecular mass of 32,245 daltons. The theoretical mass of Cdc34-Δ270 thioesterified with ubiquitin (40,792) was calculated by summing the masses of Cdc34-Δ270 (32,245) and ubiquitin (8,565) and subtracting the mass of a water molecule, which is lost during formation of the thioester bond.

For CYCE global fitting with KinTek Global Kinetic Explorer, the average of two independent experiments was fitted to a model with $\eta = 1$, and the fit for k_1–k_4 used the normalized option whereas the rest of the rate constants did not. For β-Cat global fitting, rate constants were fitted without normalization. To improve fitting, neighbouring rate constants were constrained by the end point. The dissociation constants (K_d) for CYCE/Cdc4 (ref. 21) and β-Cat/β-TrCP[22] were used to derive starting concentrations of bound substrate.

21. Hao, B. *et al.* Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol. Cell* **26**, 131–143 (2007).

22. Wu, G. *et al.* Structure of a β-TrCP1-Skp1-β-catenin complex. *Mol. Cell* **11**, 1445–1456 (2003).

nature
Vol 462|3 December 2009|doi:10.1038/nature08578

LETTERS

Extreme particle acceleration in the microquasar Cygnus X-3

M. Tavani[1,2,3,4], A. Bulgarelli[5], G. Piano[1,4], S. Sabatini[2,4], E. Striani[2,4], Y. Evangelista[1], A. Trois[1], G. Pooley[6], S. Trushkin[7], N. A. Nizhelskij[7], M. McCollough[8], K. I. I. Koljonen[9], G. Pucella[10], A. Giuliani[11], A. W. Chen[3,11], E. Costa[1], V. Vittorini[1,3], M. Trifoglio[5], F. Gianotti[5], A. Argan[1], G. Barbiellini[3,12,13], P. Caraveo[11], P. W. Cattaneo[14], V. Cocco[2], T. Contessi[11], F. D'Ammando[1,2], E. Del Monte[1], G. De Paris[1], G. Di Cocco[5], G. Di Persio[1], I. Donnarumma[1], M. Feroci[1], A. Ferrari[3,15], F. Fuschino[5], M. Galli[16], C. Labanti[5], I. Lapshov[17], F. Lazzarotto[1], P. Lipari[18,19], F. Longo[12,13], E. Mattaini[11], M. Marisaldi[5], M. Mastropietro[20], A. Mauri[5], S. Mereghetti[11], E. Morelli[5], A. Morselli[4], L. Pacciani[1], A. Pellizzoni[21], F. Perotti[11], P. Picozza[2,4], M. Pilia[21,22], M. Prest[22], M. Rapisarda[10], A. Rappoldi[14], E. Rossi[5], A. Rubini[1], E. Scalise[1], P. Soffitta[1], E. Vallazza[13], S. Vercellone[23], A. Zambra[3,11], D. Zanello[18,19], C. Pittori[24], F. Verrecchia[24], P. Giommi[24], S. Colafrancesco[24], P. Santolamazza[24], A. Antonelli[25] & L. Salotti[26]

Super-massive black holes in active galaxies can accelerate particles to relativistic energies[1], producing jets with associated γ-ray emission. Galactic 'microquasars', which are binary systems consisting of a neutron star or stellar-mass black hole accreting gas from a companion star, also produce relativistic jets, generally together with radio flares[2]. Apart from an isolated event detected[3] in Cygnus X-1, there has hitherto been no systematic evidence for the acceleration of particles to gigaelectronvolt or higher energies in a microquasar, with the consequence that we are as yet unsure about the mechanism of jet energization. Here we report four γ-ray flares with energies above 100 MeV from the microquasar Cygnus X-3 (an exceptional X-ray binary[4–6] that sporadically produces radio jets[7–9]). There is a clear pattern of temporal correlations between the γ-ray flares and transitional spectral states of the radio-frequency and X-ray emission. Particle acceleration occurred a few days before radio-jet ejections for two of the four flares, meaning that the process of jet formation implies the production of very energetic particles. In Cygnus X-3, particle energies during the flares can be thousands of times higher than during quiescent states.

Cygnus X-3 (Cyg X-3) is a powerful X-ray binary that has a period[5] of $P_{\rm orb} = 4.8$ h and a typical luminosity near the maximum accretion power of a solar-mass compact star[10], $L_{\rm X} \approx 10^{38}$ erg s^{-1}, for a 10-kpc distance[5]. The compact star powering the system is either a neutron star in an unusual state of accretion or a black hole of 10–20 solar masses orbiting around a Wolf–Rayet companion star[11]. So far, emission up to ~300 keV has been detected[12,13] from Cyg X-3 and usually shows a complex X-ray spectrum with two main states ('soft' and 'hard').

The Italian Space Agency's Astro-rivelatore Gamma ad Immagini Leggero (AGILE) satellite[14] dedicated a special programme of extensive pointings to the Cygnus region between mid 2007 and mid 2009. The AGILE γ-ray instrument[14] is very compact and is capable of monitoring cosmic sources simultaneously in the γ-ray (0.1–10-GeV) and the hard-X-ray (18–60-keV) energy bands with good sensitivity and optimal angular resolution. AGILE operates in a fixed-pointing mode, implying that it can accumulate data on a source within its large field of view (2.5 sr for the γ-ray imager) fourteen times a day, taking into account Earth occultations during each spacecraft orbit. A typical one-day exposure above 100 MeV can reach 10^7 cm^2 s and is ideal for detecting microquasar variability on short timescales (as short as a few hours).

For Cyg X-3, AGILE accumulated a total exposure above 100 MeV of almost 10^9 cm^2 s (equivalent to an effective duration of about five months). The γ-ray intensity map of the Cygnus region is shown in Supplementary Fig. 1. This region is complex, hosting star formation sites, OB associations and several prominent X-ray sources. Galactic diffuse γ-ray radiation is taken into account by modelling[15] the interaction of cosmic rays with interstellar matter and radiation fields along the line of sight. AGILE's angular resolution satisfactorily resolves the field surrounding Cyg X-3 at γ-ray energies. A dominant source 0.4° from the Cyg X-3 position is the steady γ-ray pulsar[16] 1AGL J2032+4102/0FGL J2032.2+4122. By integrating all AGILE data, we find a weak (4.6σ) γ-ray source consistent with the position of Cyg X-3 and a flux of $F = (10 \pm 2) \times 10^{-8}$ photons cm^{-2} s^{-1} above 100 MeV.

We also detect transient γ-ray emission from a flaring source consistent with the Cyg X-3 position, in four distinct episodes (Table 1; see also Supplementary Fig. 2). These flares were found in an independent multi-source maximum-likelihood search for transients in all available AGILE data (with off-axis angles less than 45°, and which covered an area of $5^\circ \times 5^\circ$ centred at the Cyg X-3 position). The statistical significance of each flare was assessed using both the maximum-likelihood analysis and the false-discovery-rate[17,18] method. They all passed stringent post-trial

---

[1]INAF-IASF Roma, Via del Fosso del Cavaliere 100, I-00133 Roma, Italy. [2]Dipartimento di Fisica, Università Tor Vergata, Via della Ricerca Scientifica 1, I-00133 Roma, Italy. [3]Consorzio Interuniversitario Fisica Spaziale, Villa Gualino, Viale Settimio Severo 63, I-10133 Torino, Italy. [4]INFN Roma Tor Vergata, Via della Ricerca Scientifica 1, I-00133 Roma, Italy. [5]INAF-IASF Bologna, Via Gobetti 101, I-40129 Bologna, Italy. [6]Astrophysics Group, Cavendish Laboratory, 19 J. J. Thomson Avenue, Cambridge CB3 0HE, UK. [7]Special Astrophysical Observatory RAS, Karachaevo-Cherkassian Republic, Nizhnij Arkhyz 36916, Russia. [8]Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [9]TKK/Metsähovi Radio Observatory, Metsähovintie 114, Kylmälä 02540, Finland. [10]ENEA Frascati, Via Enrico Fermi 45, I-00044 Frascati (RM), Italy. [11]INAF-IASF Milano, Via E. Bassini 15, I-20133 Milano, Italy. [12]Dipartimento di Fisica, Università di Trieste, Via A. Valerio 2, I-34127 Trieste, Italy. [13]INFN Trieste, Padriciano 99, I-34012 Trieste, Italy. [14]INFN Pavia, Via Bassi 6, I-27100 Pavia, Italy. [15]Dipartimento di Fisica, Università di Torino, Via P. Giuria 1, I-10125 Torino, Italy. [16]ENEA Bologna, Via don Fiammelli 2, I-40128 Bologna, Italy. [17]IKI, Profsoyuznaya Street 84, Moscow 117997, Russia. [18]INFN Roma 1, Piazza Aldo Moro 2, I-00185 Roma, Italy. [19]Dipartimento di Fisica, Università La Sapienza, Piazza Aldo Moro 2, I-00185 Roma, Italy. [20]CNR, IMIP, I-00016 Montelibretti (Rome), Italy. [21]INAF Osservatorio Astronomico di Cagliari, Poggio dei Pini, I-09012 Capoterra (Cagliari), Italy. [22]Dipartimento di Fisica, Università Insubria, Via Valleggio 11, I-22100 Como, Italy. [23]INAF-IASF Palermo, Via La Malfa 153, I-90146 Palermo, Italy. [24]ASI Science Data Center, ESRIN, I-00044 Frascati, Italy. [25]Osservatorio Astronomico di Roma, Via di Frascati 33, I-00040 Monte Porzio Catone, Italy. [26]Agenzia Spaziale Italiana, Viale Liegi 26, I-00198 Roma, Italy.

Table 1 | Major γ-ray flares of Cyg X-3

Gamma-ray flaring date	X-ray state	Radio state	ΔT_1 (d)	Radio-flare flux (detection frequency)	ΔT_2 (d)	Gamma-ray flux (10^{-8} photons cm^{-2} s^{-1})
16–17 April 2008 (MJD = 54,572–54,573)	Soft	Pre-flare	—	~16 Jy (11 GHz)	~0–1	260 ± 80
2–3 November 2008 (MJD = 54,772–54,773)	Soft	Pre-quenched	3–4	~1 Jy (15 GHz)	~8–9	258 ± 83
11–12 December 2008 (MJD = 54,811–54,812)	Soft	Optically thick–thin change	—	~3 Jy (11 GHz)	~9–10	210 ± 73
20–21 June 2009 (MJD = 55,002–55,003)	Soft	Pre-quenched	~4–5	—	—	212 ± 75

Column one shows the dates of the γ-ray flares from Cyg X-3. In all cases, Cyg X-3 showed a low-intensity or non-detectable hard-X-ray flux above 20 keV. Column two shows the X-ray state as determined by the public data in the band (2–12 keV) of the All-Sky Monitor on board NASA's Rossi X-ray Timing Explorer. Column three shows the radio-flux state at the time of the γ-ray flare. Column four shows the time delay, ΔT_1, between the γ-ray flare and the radio-state minimum, if applicable. Column five shows the magnitude of the major radio flare following the γ-ray flare (detection radio frequencies given in parentheses). Column six shows the time delay, ΔT_2, between the γ-ray flare and the major radio flare. Column seven shows the photon flux of the γ-ray flare above 100 MeV. The errors show the statistical uncertainties and take into account the effects of the Galactic diffuse emission and the presence of nearby sources. MJD, modified Julian date.

significance requirements, implying an occurrence rate for equivalent statistical fluctuations greater than one over several hundred one-day map replications (Supplementary Information). Our values for the flux, significance and position of each of these events are reported in Supplementary Table 1. By integrating all flaring data, we find a transient γ-ray source detected at the 5.5σ level at the average Galactic coordinates (longitude, latitude) $(79.6^\circ, 0.5^\circ) \pm 0.5^\circ_{\text{stat}} \pm 0.1^\circ_{\text{syst}}$ (both types of error apply to both coordinates), with an average flaring flux of $F = (190 \pm 40) \times 10^{-8}$ photons cm^{-2} s^{-1} above 100 MeV. After careful investigation, we attribute this flaring source to Cyg X-3: no other prominent source is known within the AGILE γ-ray error box. The average spectrum between 100 MeV and 3 GeV is well described by a power law with photon index 1.8 ± 0.2.

These flares are all associated with special Cyg X-3 radio and X-ray/hard X-ray states. Figure 1 shows the Cyg X-3 daily flux light curve at hard-X-ray energies (15–50 keV) as monitored using the Burst Alert Telescope on board NASA's Swift spacecraft between 1 January 2008 and 30 June 2009. Notably, all γ-ray flares in Table 1 (Fig. 1, red arrows) occur either in coincidence with low hard-X-ray fluxes or during transitions from low hard-X-ray fluxes to high. The Cyg X-3 hard-X-ray fluxes (20 keV and above) and regular-X-ray fluxes (1–10 keV) are anticorrelated[6,12]. Therefore, γ-ray flares occur only during soft-X-ray states or during transitions between such states and quenched hard-X-ray states.

A further crucial piece of information is provided by the Cyg X-3 radio states as monitored by our group using the Arcminute Microkelvin Imager Large Array, UK, and the RATAN-600 radio telescope, Russia. We find that two of four γ-ray flares are distinctly produced before major radio flares. Figure 2 shows the radio, γ-ray, soft-X-ray and hard-X-ray energy data of the strongest radio flare of

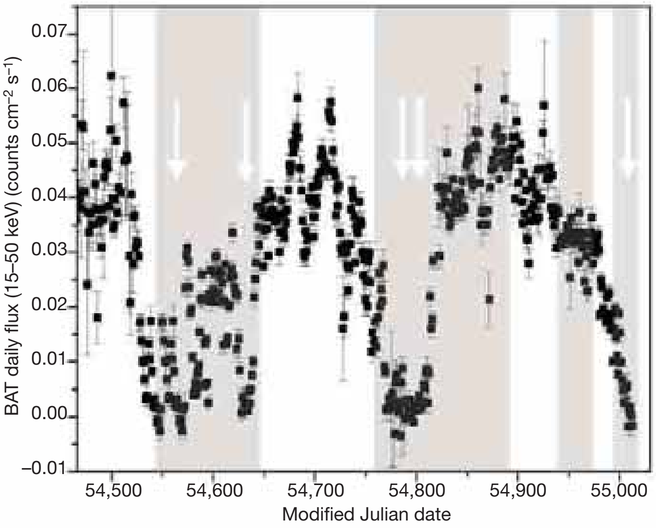


Figure 1 | Hard-X-ray flux from Cyg X-3 and γ-ray flares. Daily averaged data (in counts per second in the energy range 15–50 keV) as a function of time, as monitored by the Burst Alert Telescope (BAT) between 1 January 2008 and 25 June 2009 (http://heasarc.gsfc.nasa.gov/docs/swift/results/transients). Data errors are 1 s.d. The red arrows mark the dates of major γ-ray flares of Cyg X-3 as detected by the AGILE instrument above 100 MeV and reported in Table 1 (16–17 April 2008, 3–4 November 2008, 20–21 December 2008, 22 June 2009). The orange arrow marks a low-intensity γ-ray flare (20–21 June 2008). Grey areas show the intervals of good AGILE γ-ray exposure of Cyg X-3 with off-axis angles less than 45°.

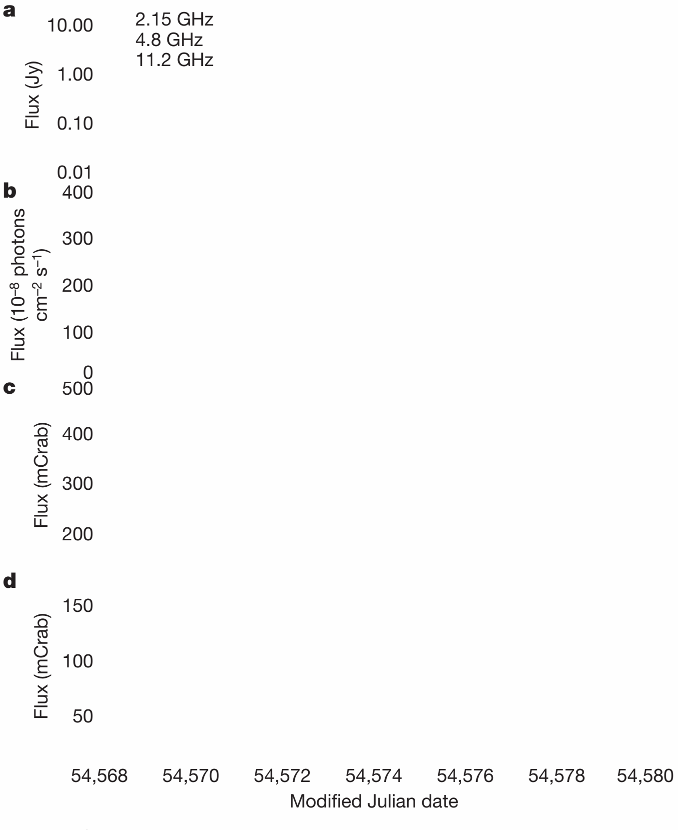


Figure 2 | Multi-frequency data for Cyg X-3 during the period 13–27 April 2008. **a**, The major radio flare of Cyg X-3 detected by the RATAN-600 radio telescope. Peak radio emission above 10 Jy at 11 GHz was detected[19] on 18 April 2008 (MJD = 54,574). **b**, Simultaneous AGILE γ-ray monitoring. A major γ-ray flare from a source consistent with Cyg X-3 above 100 MeV is detected $\Delta T_2 \approx 1$ d before the measured radio peak time. The γ-ray upper limits are at the 2σ level. **c**, X-ray flux from Cyg X-3 as monitored by the All-Sky Monitor in the energy band 1.3–12.1 keV (6-h averaged values). The hardness ratio $F(5\text{–}12\text{ keV})/F(3\text{–}5\text{ keV})$ increases by a factor of three on 17 April 2008. **d**, Hard-X-ray flux from Cyg X-3 as monitored quasi-continuously using Super-AGILE (one of the detectors on board AGILE) in the energy range 20–60 keV for an average of 14 observations per day. Errors, 1 s.d.; 1 mCrab corresponds to 9.3×10^{-12} erg cm^{-2} s^{-1} in the 20–60-keV energy range.

our sample (18 April 2008, when Cyg X-3 reached a flux level of 16 Jy at 11 GHz; refs 19, 20). Supplementary Fig. 4 shows the multi-frequency data of the 12 December 2008 γ-ray flare, which was also followed by a very strong radio outburst. On 2–3 November 2008 (and probably also on 21–22 June 2009; see Supplementary Figs 3 and 5), the γ-ray activity was associated with the source entering a 'quenched radio state'[21], that is, a rare state (occurring 2% of the time) of radio-flux minimum that usually anticipates a major radio flare[8,12].

Our results show that the flaring γ-ray emission occurs only in special transitional states associated with bright soft-X-ray states and/or the low radio emission that precedes major radio flares. We can clarify this point by using our data to update the schematic view of the radio/X-ray transitions as done in ref. 12. Figure 3 shows the peculiar 'hysteresis curve' that Cyg X-3 follows in a radio and soft-X-ray flux diagram representing all spectral states of the source. Remarkably, the γ-ray flares detected by AGILE tend to occur in the 'gully' of the hysteresis curve, which corresponds to states of very low radio flux and strong soft-X-ray flux that are preludes to major radio flares. Cyg X-3 spends only a few per cent of its time in these rare states[8,12].

Our detection of transient γ-ray emission above 100 MeV from Cyg X-3 provides direct evidence that extreme particle acceleration and non-thermalized emission can occur in microquasars with a repetitive pattern. So far, the complex interplay between inner-disk emission (and presumably soft-X-ray emission) and coronal or jet emission (most probably related to hard-X-ray emission) in Cyg X-3 has been addressed by Comptonization models that reproduce[12,13,22] the spectral states up to ~300 keV. Despite having a number of important differences with other Galactic systems[22], the Cyg X-3 states resemble those of accreting black holes with a mixed population of thermal and non-thermal relativistic electrons. The Cyg X-3 X-ray spectral states are fitted by optically thick Comptonized models that include a reflection component[23] and an additional high-energy power-law component[12,13]. Typical temperatures of the hot coronal plasma reach an equivalent energy of ~10 keV, and the high-energy tail has been modelled to extend to photon energies of one megaelectronvolt or slightly higher[12,13].

However, Cyg X-3 is capable of accelerating particles by a very efficient mechanism leading to photon emission at energies thousands of times larger than the maximum energy so far detected ($E \approx 300$ keV). Furthermore, photons have to escape from regions that might be opaque to γ-rays because of pair production, unless special conditions are satisfied. The peak γ-ray isotropic luminosity is $L_\gamma \approx 3 \times 10^{36}$ erg s^{-1} above 100 MeV. It is unknown whether this emission is leptonic or hadronic. The site and ultimate origin of this extreme particle acceleration depend on the disk–corona dynamics that lead to jet formation and relativistic propagation. A number of complex phenomena such as coronal magnetic field reconnection or shock acceleration along the proto-jets can take place and influence the formation of particle energy distribution functions at kinetic energies of up to 1 GeV and greater. For leptonic emission (see, for example, refs 24, 25 for microquasar studies), synchrotron radiation in the gigaelectronvolt domain requires large magnetic fields of ~10^4 G and Lorentz factors of $\gamma \approx 10^5$ or greater. Alternatively, inverse Compton emission, at gigaelectronvolt energies, of the soft X-rays from the disk and/or optical or infrared radiation from the companion star also requires Lorentz factors of $\gamma \approx 10^5$ before being suppressed in the Klein–Nishina regime at teraelectronvolt energies. Depending on the geometry (the Cyg X-3 jet inclination angle is constrained[7] to be ~10°–20°), γ-ray emission can be boosted or partly suppressed. Hadronic emission involves proton–proton interactions (in the shock-accelerated front interacting, for example, with the Wolf–Rayet mass outflow) in relatively dense environments (see, for example, ref. 26). In this case, protons have to be accelerated to Lorentz factors of $\gamma \approx 10^2$ or greater, and the interaction may require a critical target density to occur.

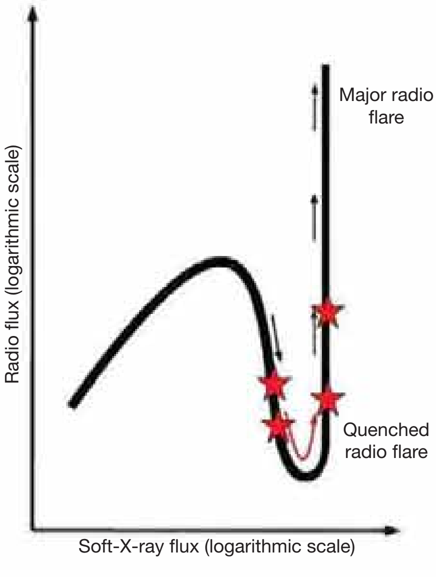


Figure 3 | Schematic representation of the radio and X-ray spectral states of Cyg X-3. The trajectory shown represents the correlated radio/soft X-ray spectral states of Cyg X-3 (adapted from ref. 12). The red stars mark the approximate positions of the major γ-ray flares detected by AGILE (Table 1), which tend to occur in low-flux/pre-flare radio states. Cyg X-3 is rarely in these 'gully' states (only a few per cent of the total time[8,12]; Supplementary Information). For all γ-ray flaring episodes, the radio and hard-X-ray fluxes are low or very low, whereas the X-ray flux (1–10 keV) is large.

Received 7 August; accepted 8 October 2009.

1. Krolik, J. H. *Active Galactic Nuclei* (Princeton Univ. Press, 1998).
2. Mirabel, I. F. & Rodriguez, L. F. A superluminal source in the galaxy. *Nature* **371,** 46–48 (1994).
3. Albert, J. *et al.* Very high energy gamma-ray radiation from the stellar mass black hole binary Cygnus X-1. *Astrophys. J.* **665,** 51–54 (2007).
4. Giacconi, R., Gorenstein, P., Gurski, H. & Waters, J. R. An X-ray survey of the Cygnus region. *Astrophys. J.* **148,** L119–L127 (1967).
5. Bonnet-Bidaud, J.-M. & Chardin, G. Cygnus X-3, a critical review. *Phys. Rep.* **170,** 325–404 (1988).
6. McCollough, M. L. *et al.* Discovery of correlated behavior between the hard X-ray and the radio bands in Cygnus X-3. *Astrophys. J.* **517,** 951–955 (1999).
7. Mioduszewski, A. J., Rupen, M. P., Hjellming, R. M., Pooley, G. G. & Waltman, E. B. A one-sided highly relativistic jet from Cygnus X-3. *Astrophys. J.* **553,** 766–775 (2001).
8. Waltman, E. B., Friedler, R. L., Johnston, K. J. & Ghigo, F. D. The quiescent level of Cygnus X-3 at 2.25 and 8.3 GHz: 1988–1992. *Astron. J.* **108,** 179–187 (1994).
9. Martì, J., Paredes, J. M. & Peracaula, M. The Cygnus X-3 radio jets at arcsecond scales. *Astrophys. J.* **545,** 939–944 (2000).
10. Vilhu, O., Hakala, P., Hannikainen, D. C., McCollough, M. & Koljonen, K. Orbital modulation of X-ray emission lines in Cygnus X-3. *Astron. Astrophys.* **501,** 679–686 (2009).
11. van Kerkwijk, M. H. *et al.* Infrared helium emission lines from Cygnus X-3 suggesting a Wolf–Rayet star companion. *Nature* **355,** 703–705 (1992).
12. Szostek, A., Zdziarski, A. A. & McCollough, M. L. A classification of the X-ray and radio states of Cyg X-3 and their long-term correlations. *Mon. Not. R. Astron. Soc.* **388,** 1001–1010 (2008).
13. Hjalmarsdotter, L., Zdziarski, A. A., Szostek, A. & Hannikainen, D. C. Spectral variability in Cygnus X-3. *Mon. Not. R. Astron. Soc.* **392,** 251–263 (2009).
14. Tavani, M. *et al.* The AGILE mission. *Astron. Astrophys.* **502,** 995–1013 (2009).
15. Giuliani, A. *et al.* Gamma-ray emission from the galaxy: a new model for AGILE. *Mem. Soc. Astron. Ital.* **5,** 135–140 (2004).
16. Abdo, A.A., *et al.* Detection of 16 Gamma-Ray Pulsars through blind frequency searches using the Fermi LAT. *Science* **325,** 840–844 (2009).
17. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate – a practical and powerful approach to multiple testing. *J. R. Stat. Soc. A* **57,** 289–300 (1995).
18. Miller, C. J. *et al.* Controlling the false-discovery rate in astrophysical data analysis. *Astron. J.* **122,** 3492–3505 (2001).
19. Trushkin, S. *et al.* A giant radio flare of Cygnus X-3. *Astron. Telegr.* **1483** (2008).
20. Pal, S., Trushkin, S. & Chandra, I. GMRT low frequency radio observation of the giant flare from Cygnus X-3. *Astron. Telegr.* **1486** (2008).
21. Pooley, G. Cygnus X-3 radio observations. *Astron. Telegr.* **1828** (2008).

22. Zdziarski, A. A. & Gierlinksi, M. Radiative processes, spectral states and variability of black-hole binaries. *Prog. Theor. Phys.* **155** (suppl.), 99–119 (2004).
23. Coppi, P. S. in *High Energy Processes in Accreting Black Holes* (eds Poutanen, J. & Svensson, R.) 375–403 (ASP Conf. Ser. 161, Astronomy Society of the Pacific, 1999).
24. Atoyan, A. M. & Aharonian, F. A. Modelling of the non-thermal flares in the Galactic microquasar GRS 1915+105. *Mon. Not. R. Astron. Soc.* **302,** 253–276 (1999).
25. Markoff, S., Falcke, H. & Fender, R. P. A jet model for the broadband spectrum of XTE J1118+480. Synchrotron emission from radio to X-rays in the low/hard spectral state. *Astron. Astrophys.* **372,** L25–L28 (2001).
26. Romero, G. E., Torres, D. F., Kaufman Bernadò, M. M. & Mirabel, I. F. Hadronic gamma-ray emission from windy microquasars. *Astron. Astrophys.* **410,** L1–L4 (2003).

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

Acknowledgements The AGILE mission is funded by the Italian Space Agency with scientific and programmatic participation by the Italian Institute of Astrophysics and the Italian Institute of Nuclear Physics.

Author Contributions M. Tavani, A.B., G. Piano, S.S., E. Striani, G. Pucella and A.G. carried out the γ-ray analysis; G. Pooley, S.T. and N.A.N. obtained and analysed the radio data; and M. McCollough., K.I.I.K., Y.E., E.C. and M.F. analysed and discussed the X-ray data. 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 M. Tavani (pi.agile@iasf-roma.inaf.it).

nature Vol 462|3 December 2009|doi:10.1038/nature08579

LETTERS

Supernova 2007bi as a pair-instability explosion

A. Gal-Yam[1], P. Mazzali[2,3], E. O. Ofek[4], P. E. Nugent[5], S. R. Kulkarni[4], M. M. Kasliwal[4], R. M. Quimby[4], A. V. Filippenko[6], S. B. Cenko[6], R. Chornock[6], R. Waldman[7], D. Kasen[8], M. Sullivan[9], E. C. Beshore[10], A. J. Drake[4], R. C. Thomas[5], J. S. Bloom[6], D. Poznanski[6], A. A. Miller[6], R. J. Foley[11], J. M. Silverman[6], I. Arcavi[1], R. S. Ellis[4] & J. Deng[12]

Stars with initial masses such that $10M_\odot \le M_{\mathrm{initial}} \le 100M_\odot$, where $M_\odot$ is the solar mass, fuse progressively heavier elements in their centres, until the core is inert iron. The core then gravitationally collapses to a neutron star or a black hole, leading to an explosion—an iron-core-collapse supernova[1,2]. By contrast, extremely massive stars with $M_{\mathrm{initial}} \ge 140M_\odot$ (if such exist) develop oxygen cores with masses, M_{core}, that exceed $50M_\odot$, where high temperatures are reached at relatively low densities. Conversion of energetic, pressure-supporting photons into electron–positron pairs occurs before oxygen ignition and leads to a violent contraction which triggers a nuclear explosion[3–5] that unbinds the star in a pair-instability supernova. Transitional objects with $100M_\odot < M_{\mathrm{initial}} < 140M_\odot$ may end up as iron-core-collapse supernovae following violent mass ejections, perhaps as a result of brief episodes of pair instability, and may already have been identified[6–8]. Here we report observations of supernova SN 2007bi, a luminous, slowly evolving object located within a dwarf galaxy. We estimate the exploding core mass to be $M_{\mathrm{core}} \approx 100M_\odot$, in which case theory unambiguously predicts a pair-instability supernova. We show that $>3M_\odot$ of radioactive ^{56}Ni was synthesized during the explosion and that our observations are well fitted by models of pair-instability supernovae[9,10]. This indicates that nearby dwarf galaxies probably host extremely massive stars, above the apparent Galactic stellar mass limit[11], which perhaps result from processes similar to those that created the first stars in the Universe.

We identified a new optical transient (SNF20070406-008) on 2007 April 6.5 UT (universal time is used for all dates in this paper) at right ascension 13 h 19 min 20.2 s and declination 08° 55′ 44.0″ (J2000). Follow-up spectroscopic observations showed that this is a supernova (SN 2007bi), with no trace of either hydrogen or helium, leading to its classification as a type Ic supernova[12], albeit of a peculiar nature and with only one previously known counterpart (SN 1999as[13]; Fig. 1). No signs of interaction with circumstellar material (a major source of uncertainty in the analysis of previous luminous supernovae[7,14]) are observed throughout the evolution of this event (Fig. 1). A search for pre-discovery data recovered observations by the University of Arizona's Catalina Sky Survey[15] that allowed an accurate determination of the date and magnitude of the supernova brightness peak. Photometric observations continued for 555 days during our intensive follow-up campaign. A red (R-band) light curve is plotted in Fig. 2a.

The measured light curve is unique, showing a very long rise time to peak (~70 d; Fig. 2; Supplementary Information, section 2), an extreme luminosity reaching a peak R-band absolute magnitude of $M_{\mathrm{R}} = -21.3$ mag and a slow decline (0.01 mag d^{-1} over >500 d),

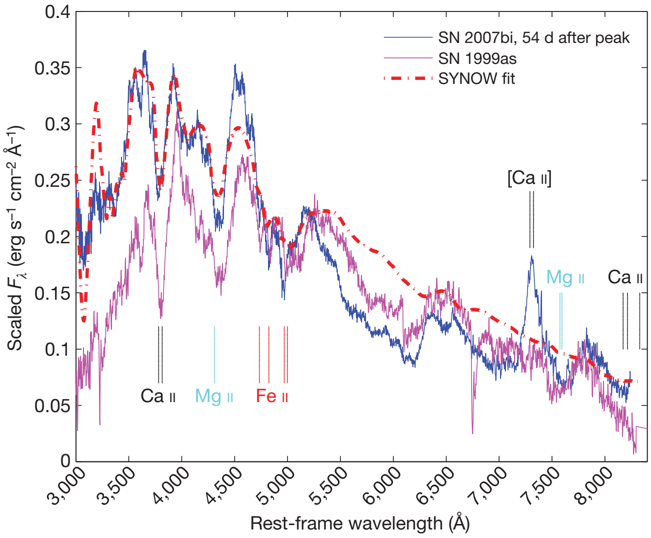


Figure 1 | Spectra of the unusual type Ic supernova SN 2007bi. We observed SN 2007bi on 2007 April 15.6 and 16.4, using the Low Resolution Imaging Spectrometer[26] (LRIS) mounted on the Keck I 10-m telescope (W. M. Keck Observatory, Hawaii). Here we plot the specific flux, F_λ, with arbitrary scaling. Narrow emission lines (see Fig. 3 for details) indicate a redshift of $z = 0.1279$. A survey of our databases of supernova spectra shows that this event is similar to only a single previous example, SN 1999as[13], which until now was the most luminous known type Ic supernova by a wide margin. SN 2007bi has a comparable luminosity (Fig. 2). We identify the most prominent features (marked) as arising from calcium, magnesium and iron, and derive a photospheric velocity of 12,000 km s^{-1}. A model fit (Supplementary Information, section 5) confirms these line identifications and shows that the absorption in the blue–ultraviolet part of the spectrum is dominated by blends of iron, cobalt and nickel lines. The shallow, poorly defined trough seen around 5,500–6,000 Å could arise from blends of silicon and sulphur lines, and the lines of neutral oxygen and sodium that are usually prominent in type Ic spectra are remarkably weak. No hydrogen or helium lines are seen, confirming the type Ic classification and strongly disfavouring the possibility that interaction with circumstellar material contributes significantly to the luminosity[7,14,27]. No narrow sodium absorption is seen, indicating little absorption by dust in the host galaxy[19]. The strong emission line around 7,300 Å seems to arise from [Ca II] emission at zero expansion velocity, and is usually observed in supernova only in late-time nebular spectra. A complete analysis of our full set of photospheric spectra will be presented in a forthcoming publication; see also additional details in ref. 19.

---

[1]Benoziyo Center for Astrophysics, Faculty of Physics, The Weizmann Institute of Science, Rehovot 76100, Israel. [2]Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, 85748 Garching, Germany. [3]Scuola Normale Superiore, Piazza Cavalieri 7, 56127 Pisa, Italy. [4]Department of Astronomy, 105-24, California Institute of Technology, Pasadena, California 91125, USA. [5]Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720-3411, USA. [6]Department of Astronomy, University of California, Berkeley, California 94720-3411, USA. [7]The Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel. [8]Department of Astronomy and Astrophysics, University of California, Santa Cruz, California 95064, USA. [9]Department of Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH, UK. [10]Department of Planetary Sciences, Lunar and Planetary Laboratory, 1629 East University Boulevard, Tucson, Arizona 85721, USA. [11]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. [12]National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China.

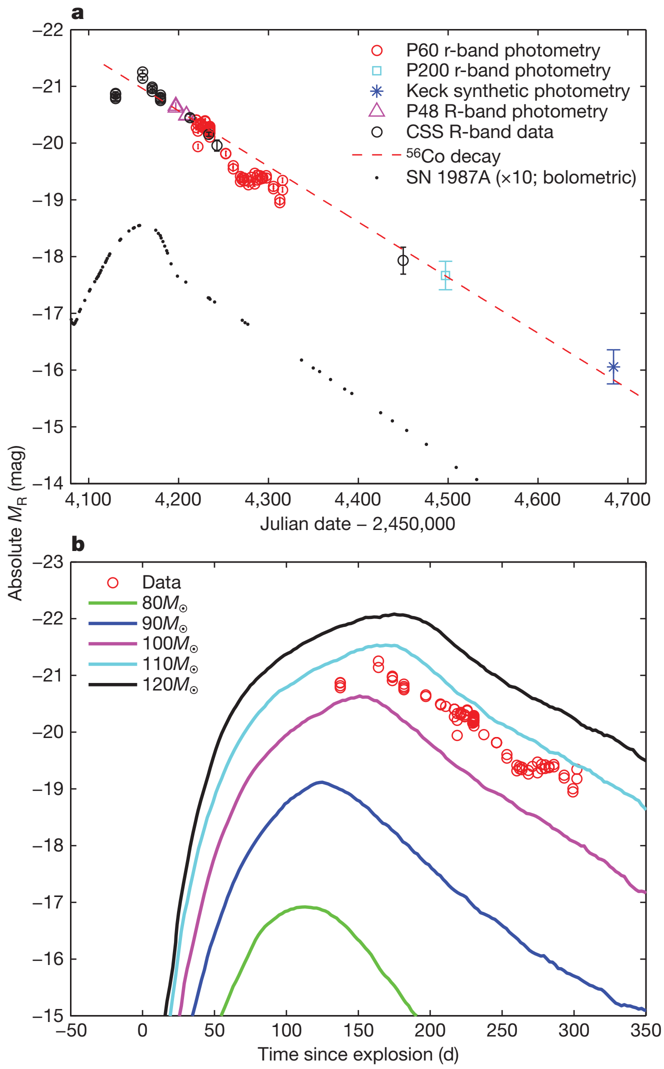


Figure 2 | Radioactive ^{56}Ni and total ejected mass from the light-curve evolution of SN 2007bi are well fitted using PISN models. a, The R-band light curve of SN 2007bi. We have compiled observations obtained using the 48-inch (1.2-m) Samuel Oschin Telescope (P48), the 60-inch (1.5-m) robotic telescope (P60) and the 200-inch (5-m) Hale Telescope (P200) at Palomar Observatory, California, as well as photometry from the Catalina Sky Survey[15] (CSS) and synthetic photometry integrated from our late-time Keck spectrum (Fig. 3; see Supplementary Information, sections 1 and 2, for additional details). We find a peak magnitude of $M_R = -21.3 \pm 0.1$ mag on 2007 February 21 (Supplementary Information, section 2). The error is dominated by the absolute zero-point calibration uncertainty. The outstanding peak luminosity of this event, if radioactively driven, suggests that a remarkable amount of ^{56}Ni was produced ($>3M_\odot$; ref. 28; Supplementary Information, section 3). The slow rise time derived from our fit (77 d; Supplementary Information, section 2), combined with the measured photospheric velocity (12,000 km s^{-1}; Fig. 1), requires very massive ejecta ($M_{ej} \approx 100M_\odot$) and a huge release of kinetic energy ($E_k \approx 10^{53}$ erg; Supplementary Information, section 3), where we apply the commonly used scaling relations[28,29]. An independent direct estimate for the ^{56}Ni yield is obtained from the luminosity during the late-time radioactive-decay phase, compared with the observed decay of SN 1987A[30] (Supplementary Information, section 3). Given the uncertainty in the explosion date of SN 2007bi and a range of bolometric correction values (Supplementary Information, section 2), the ^{56}Ni mass produced by SN 2007bi satisfied $4M_\odot < M_{^{56}\mathrm{Ni}} < 7M_\odot$. The total radiated energy we measure by direct integration of the light curve is $E_{rad} \approx (1\text{–}2) \times 10^{51}$ erg (Supplementary Information, section 3), which is comparable to that of the most luminous supernovae known[7]. Errors, 1σ. **b**, Comparison of the observations of SN 2007bi with models calculated before the supernova's discovery[5,9]. The curves presented are for various helium cores (masses as indicated) exploding as PISNs, and cover the photospheric phase. The data are well fitted by $100M_\odot$–$110M_\odot$ models. At later times, the emission is nebular and bolometric corrections used to calculate the model R-band light curve cease to apply (Supplementary Information, section 4). In comparing with these rest-frame models, cosmological time dilation for $z = 0.1279$ has been taken into account.

consistent with the decay rate of radioactive ^{56}Co. These properties suggest that the very massive ejecta were energized by a large amount of radioactive nickel ($>3M_\odot$; Figs 2 and 3; Supplementary Information, section 3), as expected from models of pair-instability supernovae[4,5,10] (PISNs). Our spectra, lacking any signs of hydrogen or helium, indicate that this mass is dominated by carbon, oxygen and heavier elements. The large amount of kinetic energy released, $E_k \approx 10^{53}$ erg (Fig. 2; Supplementary Information, section 3), is comparable to those derived for the most energetic γ-ray bursts[16], placing this event among the most extreme explosions known. In Fig. 2b, we show theoretical light curves calculated from PISN models[5,9] before our discovery. The data fit the models very well, suggesting that we observed the explosion of a star with a helium core mass of around $100M_\odot$.

PISN models imply that such an explosion would synthesize $3M_\odot$–$10M_\odot$ of radioactive ^{56}Ni (Table 1). Such a large amount of newly synthesized radioactive material would energize the supernova debris for an extended period of time, ionizing the expanding gas cloud. Collisional excitation would lead to strong nebular emission lines, whose strength should be roughly proportional to the amount of radioactive source material, providing another testable prediction. Figure 3a shows a comparison of the nebular spectrum of SN 2007bi with that of the well-studied, ^{56}Ni-rich SN 1998bw, which produced $\sim 0.5M_\odot$ of radioactive nickel[17], suggesting that SN 2007bi produced $\gtrsim 7M_\odot$ of nickel (Supplementary Information, section 3), again supporting its interpretation as a PISN.

By modelling the nebular spectrum, we are able to resolve the elemental composition of the fraction of the ejected mass that is illuminated by radioactive nickel. We can directly measure the abundances of carbon, oxygen, sodium, magnesium, calcium and iron, and derive the mass of radioactive ^{56}Ni. Our elemental abundance ratios are in good agreement with model predictions[5] for heavier elements, but lighter elements (carbon, oxygen and magnesium) seem to be under-observed. By adopting the calculated model output[5] for elements that do not have strong nebular emission in the optical (mostly silicon and sulphur, and some neon and argon), we arrive at a total illuminated mass of $>50M_\odot$, with a composition as described in Table 1. We note that this falls well below the total mass derived from photometry, indicating that even the unprecedented amount of radioactive nickel produced by SN 2007bi was not sufficient to energize the entire mass ejected by this extreme explosion (Supplementary Information, section 6). The unilluminated mass probably contains more light elements that originated in the outer envelopes of the exploding star, and in which our nebular observations are deficient (see Supplementary Information, sections 3 and 6, for additional details).

Our data thus provide strong evidence that we have observed the explosion of a helium core with $M \approx 100M_\odot$, which, according to theory, can only result in a PISN[3–6,10,18]. The measured light curve, yield of radioactive nickel and elemental composition of the ejecta are consistent with models of PISNs that were calculated before our discovery. On the basis of fewer observations of SN 2007bi, combined with their analysis of the host-galaxy properties, the authors of ref. 19 consider both a PISN model and an interpretation in terms of a massive iron-core-collapse supernova[20], slightly favouring the latter. However, our quantitative estimate of the helium core mass from our peak light-curve shape and analysis of the nebular spectra is inconsistent with iron-core-collapse models[20] and theoretically requires a PISN[5,10]. We thus conclude that we have most likely discovered the first clear example of a PISN.

There are several implications of this discovery. Theory allows stars as massive as $1,000M_\odot$ to have formed in the very early Universe[21].

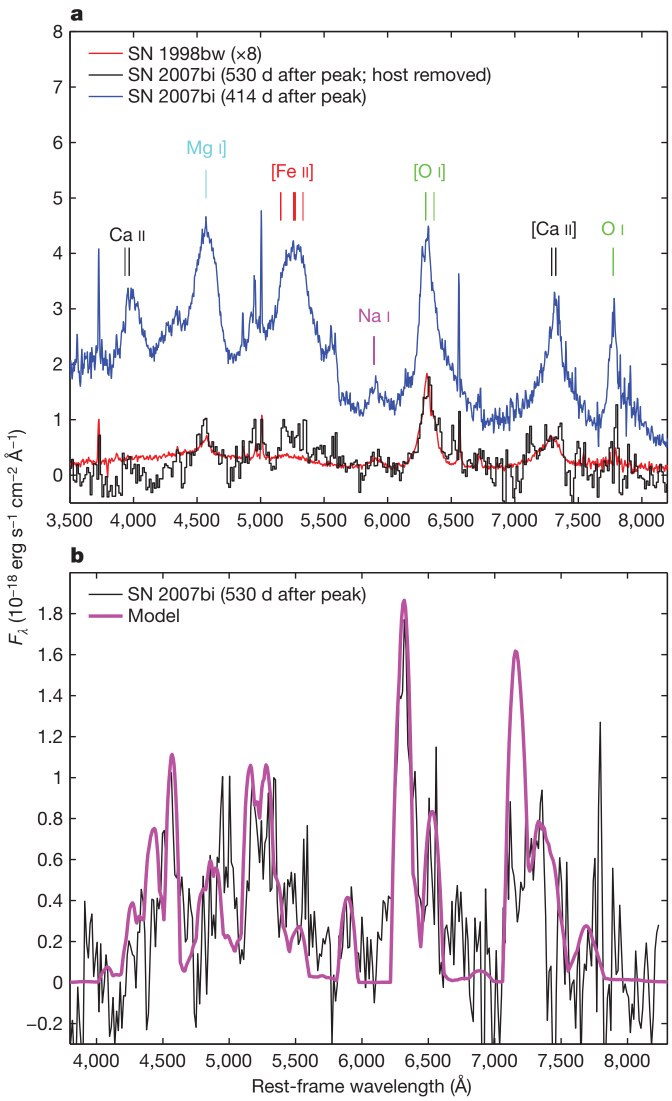


Figure 3 | Ejecta composition from nebular spectra of SN 2007bi. **a**, Two late-time spectra of SN 2007bi. A spectrum obtained with the FORS2 spectrograph mounted on the European Space Observatory's 8.1-m Very Large Telescope at the Paranal facilities of the La Silla Paranal Observatory, Chile, on 2008 April 10 (414 d after peak; 367 d rest frame) is not completely nebular yet. Prominent broad supernova emission peaks from neutral and singly ionized elements are marked. Multiple narrow host-galaxy emission lines ([O II], wavelength $\lambda = 3{,}727$ Å; Hβ; [O III], $\lambda = 4{,}959$ Å, 5,007 Å; Hα; [S II], $\lambda = 6{,}716$ Å, 6,731 Å) are seen at $z = 0.1279$. A later spectrum (530 d after peak; 470 d rest frame) was obtained using the LRIS mounted on the Keck I 10-m telescope on 2008 August 4. This spectrum is fully nebular, although of lower signal-to-noise ratio. Comparison with a late-time spectrum of the ^{56}Ni-rich SN 1998bw, which produced $0.5M_\odot$ of ^{56}Ni (ref. 17), adjusted for the larger distance and later rest-frame spectroscopic observations of SN 2007bi, and multiplied by eight, provides a good fit for intermediate-mass elements (oxygen, sodium, magnesium and calcium) but under-predicts the strength of the iron lines. Assuming that emission-line luminosity scales with the mass of energizing ^{56}Ni, we derive from the scaling factor a ^{56}Ni mass such that $7.7M_\odot < M_{^{56}\mathrm{Ni}} < 11.3M_\odot$ (Supplementary Information, section 3), in reasonable agreement with estimates from early (peak) and late-time (radioactive-tail) photometric estimates. We note that the lack of Hα emission at these late epochs is an especially strong argument against interaction with circumstellar material. A lower limit of $\sim 5 \times 10^{16}$ cm on the distance to any hydrogen-rich material (in particular recent mass loss) is derived from this non-detection assuming an expansion velocity of 12,000 km s^{-1} (Fig. 1). **b**, Modelling the nebular spectrum. By using our nebular spectroscopy code[17], we are able to constrain the composition of the ejecta. We find that the emission-line luminosity requires an initial ^{56}Ni mass of $3.7M_\odot$–$7.4M_\odot$ and the composition given in Table 1. As can be seen there, our measurements are well fitted by the theoretical predictions of PISN models[5] calculated before we discovered SN 2007bi.

However, the most massive stars known in the local Universe (for example luminous blue variables) have estimated masses of $\sim 150M_\odot$ (ref. 11). In the single example known so far, such a hypergiant star exploded in a normal core-collapse event (SN 2005gl)[2]. Our detection of a PISN forming from a $\sim 100M_\odot$ core suggests a progenitor with an estimated initial mass of $\sim 200M_\odot$ (ref. 5), assuming very low mass loss rates appropriate for zero-metallicity stars. We note that this estimate is highly sensitive to assumed mass-loss processes, which are poorly understood, and that high-metallicity mass-loss prescriptions would require an even higher initial mass. In a sense, our discovery of such a core is in conflict with commonly used mass-loss calculations, which do not allow such a high-mass core to form at the measured metallicities[19]. Our finding probably requires the modification of mass-loss models, perhaps through increased clumping in massive-star winds[7,14,18], at least during the hydrogen-rich mass-loss phase.

Regardless of the exact mass loss assumed, our data indicate that extremely massive stars above the Galactic limit ($M > 150M_\odot$) are formed in the local Universe. Perhaps the dwarf, metal-poor host galaxy of SN 2007bi ($M_\mathrm{B} = -16.3$ mag at redshift $z = 0.1279$, indicating an approximate metallicity of $12 + \log[\mathrm{O/H}] = 8.25$ (ref. 22; see ref. 19 for additional details), where O and H respectively denote the number densities of atomic oxygen and hydrogen) retained conditions that were similar to those prevalent in the early Universe. Luminous events such as SN 2007bi can therefore serve as beacons, focusing our attention on otherwise unremarkable local dwarf galaxies that can be used as fossil laboratories to study the early Universe.

Our observational confirmation of PISN models supports their use in predicting the detectability and observed properties of PISNs formed from the first stars, by future missions such as NASA's James Webb Space Telescope; in estimating their contribution to

Table 1 | Predicted PISN ejecta elemental composition compared with our measurements

	C ($M_\odot$)	O ($M_\odot$)	Ne ($M_\odot$)	Na ($M_\odot$)	Mg ($M_\odot$)	Si ($M_\odot$)	S ($M_\odot$)	Ar ($M_\odot$)	Ca ($M_\odot$)	^{56}Ni ($M_\odot$)	Total ($M_\odot$)
Measured (1)	1.0*	10.1*	4.0	0.0012*	0.068*	22.0	10.0	1.3	0.75*	4.5*	53.8
Measured (2)	1.0*	12.0*	4.0	0.0013*	0.095*	22.0	10.0	1.3	0.90*	5.7*	57.1
Measured (3)	1.2*	14.6*	4.0	0.0018*	0.13*	22.0	10.0	1.3	1.00*	7.4*	61.7
Measured (4)	1.0*	7.5*	4.0	0.0015*	0.065*	22.0	10.0	1.3	0.95*	3.65*	50.6
Measured (5)	1.0*	9.1*	4.0	0.0018*	0.085*	22.0	10.0	1.3	1.10*	4.6*	53.3
Measured (6)	1.0*	11.3*	4.0	0.0023*	0.12*	22.0	10.0	1.3	1.00*	6.0*	56.8
$95M_\odot$ model[5]	4.1	45.2	4.0	0.0028	4.38	21.3	8.8	1.3	0.99	2.98	95
$100M_\odot$ model[5]	4.0	43.9	4.1	0.0028	4.41	23.1	10.0	1.5	1.22	5.82	100
$105M_\odot$ model[5]	3.9	42.7	3.9	0.0028	4.40	24.5	10.8	1.7	1.40	9.55	105
$95M_\odot$ model[10]	0.28	38.1	1.60	0.0011	2.35	20.9	13.5	2.32	1.98	5.18	95
$100M_\odot$ model[10]	0.25	36.4	1.51	0.0010	2.45	22.6	15.0	2.54	2.23	8.00	100
$105M_\odot$ model[10]	0.23	35.2	1.47	0.0009	2.47	23.9	15.9	2.73	2.46	11.89	105

We report six different estimates based on the two available late-time spectra (Fig. 3): Very Large Telescope day 414 (rows 1–3) and Keck day 530 (rows 4–6). Each spectrum is modelled assuming three different explosion dates: 45 d (rows 1, 4), 77 d (rows 2, 5) and 113 d (rows 3, 6) before peak (observed). The results are qualitatively similar, with ^{56}Ni masses of $3.7M_\odot$–$7.4M_\odot$ and total masses of $51M_\odot$–$62M_\odot$. Elements indicated with an asterisk are directly constrained from optical nebular emission lines. Abundances of other elements are probed through their cooling effects using lines outside the optical range (which are modelled), but we consider these constraints to be weaker.

the chemical evolution of the Universe[5]; and in calculating their effect on the reionization of the Universe. With the advent of new wide-field surveys, such as the Palomar Transient Factory[23,24] and the Catalina Real-Time Transient Survey[15], that monitor millions of nearby low-luminosity, anonymous galaxies[25], many additional such events should soon be discovered and will further illuminate these important questions.

Received 6 August; accepted 8 October 2009.

1. Smartt, S. J. Progenitors of core-collapse supernovae. *Annu. Rev. Astron. Astrophys.* **47,** 63–106 (2009).
2. Gal-Yam, A. & Leonard, D. C. A massive hypergiant star as the progenitor of the supernova SN 2005gl. *Nature* **458,** 865–867 (2009).
3. Rakavy, G. & Shaviv, G. Instabilities in highly evolved stellar models. *Astrophys. J.* **148,** 803–816 (1967).
4. Barkat, Z., Rakavy, G. & Sack, N. Dynamics of supernova explosion resulting from pair formation. *Phys. Rev. Lett.* **18,** 379–381 (1967).
5. Heger, A. & Woosley, S. E. The nucleosynthetic signature of population III. *Astrophys. J.* **567,** 532–543 (2002).
6. Woosley, S. E., Blinnikov, S. & Heger, A. Pulsational pair instability as an explanation for the most luminous supernovae. *Nature* **450,** 390–392 (2007).
7. Smith, N., Chornock, R., Silverman, J. M., Filippenko, A. V. & Foley, R. J. Spectral evolution of the extraordinary type IIn supernova 2006gy. Preprint at ⟨http://arxiv.org/abs/0906.2200⟩ (2009).
8. Miller, A. A. *et al.* New observations of the very luminous supernova 2006gy: evidence for echoes. Preprint at ⟨http://arxiv.org/abs/0906.2201⟩ (2009).
9. Kasen, D., Heger, A. & Woosley, S. The first stellar explosions: theoretical light curves and spectra of pair-instability supernovae. *Proc. Am. Inst. Phys. Conf.* **990,** 263–267 (2008).
10. Waldman, R. The most massive core-collapse supernova progenitors. *Astrophys. J.* **685,** 1103–1108 (2008).
11. Figer, D. F. An upper limit to the masses of stars. *Nature* **434,** 192–194 (2005).
12. Filippenko, A. V. Optical spectra of supernovae. *Annu. Rev. Astron. Astrophys.* **35,** 309–355 (1997).
13. Kasen, D. N. *Aspherical Supernovae*. PhD thesis, Univ. California, Berkeley (2004).
14. Agnoletto, I. *et al.* SN 2006gy: was it really extraordinary? *Astrophys. J.* **691,** 1348–1359 (2009).
15. Drake, A. J. *et al.* First results from the Catalina Real-Time Transient Survey. *Astrophys. J.* **696,** 870–884 (2009).
16. Cenko, S. B. *et al.* The collimation and energetics of the brightest Swift gamma-ray bursts. Preprint at ⟨http://arxiv.org/abs/0905.0690⟩ (2009).
17. Mazzali, P. A., Nomoto, K., Patat, F. & Maeda, K. The nebular spectra of the hypernova SN 1998bw and evidence for asymmetry. *Astrophys. J.* **559,** 1047–1053 (2001).
18. Langer, N. *et al.* Pair creation supernovae at low and high redshift. *Astron. Astrophys.* **475,** L19–L23 (2007).
19. Young, D. R. *et al.* Two peculiar type Ic supernovae in low-metallicity, dwarf galaxies. *Astron. Astrophys.*. (submitted).
20. Umeda, H. & Nomoto, K. How much ^{56}Ni can be produced in core-collapse supernovae? Evolution and explosions of 30–100$M_\odot$ stars. *Astrophys. J.* **673,** 1014–1022 (2008).
21. Bromm, V. & Larson, R. B. The first stars. *Annu. Rev. Astron. Astrophys.* **42,** 79–118 (2004).
22. Tremonti, C. A. *et al.* The origin of the mass–metallicity relation: insights from 53,000 star-forming galaxies in the Sloan Digital Sky Survey. *Astrophys. J.* **613,** 898–913 (2004).
23. Law, N. M. *et al.* The Palomar Transient Factory: system overview, performance and first results. *Publ. Astron. Soc. Pacif.* (in the press); preprint at ⟨http://arxiv.org/abs/0906.5350⟩ (2009).
24. Rau, A. *et al.* Exploring the optical transient sky with the Palomar Transient Factory. *Publ. Astron. Soc. Pacif.* (in the press); preprint at ⟨http://arxiv.org/abs/0906.5355⟩ (2009).
25. Young, D. R. *et al.* Core-collapse supernovae in low-metallicity environments and future all-sky transient surveys. *Astron. Astrophys.* **489,** 359–375 (2008).
26. Oke, J. B. *et al.* The Keck Low-Resolution Imaging Spectrometer. *Publ. Astron. Soc. Pacif.* **107,** 375–385 (1995).
27. Miller, A. A. *et al.* The exceptionally luminous Type II-linear supernova 2008es. *Astrophys. J.* **690,** 1303–1312 (2009).
28. Perets, H. B. *et al.* A new type of stellar explosion. Preprint at ⟨http://arxiv.org/abs/0906.2003⟩ (2009).
29. Foley, R. J. *et al.* SN 2008ha: an extremely low luminosity and exceptionally low energy supernova. *Astron. J.* **138,** 376–391 (2009).
30. Pun, C. S. J. *et al.* Ultraviolet observations of SN 1987A with the IUE satellite. *Astrophys. J. Suppl. Ser.* **99,** 223–261 (1995).

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

Acknowledgements We gratefully acknowledge advice and help from E. Pian and discussions with Z. Barkat, E. Livne, E. Nakar, N. Langer and P. Podsiadlowski. This work benefited from useful interaction during the Fireworks meetings held at the Weizmann Institute (2008) and at the University of Bonn (2009). Work related to the CSS data reported here was supported by the US National Aeronautics and Space Administration (NASA) under a grant issued through the Science Mission Directorate Near-Earth Object Observations program. The joint work of A.G.-Y. and P.M. is supported by a Weizmann-Minerva grant. A.G.-Y. acknowledges support from the Israeli Science Foundation, a European Union Seventh Framework Programme Marie Curie IRG fellowship, the Benoziyo Center for Astrophysics, a research grant from the Peter and Patricia Gruber Awards, and the William Z. and Eda Bess Novick New Scientists Fund at the Weizmann Institute. P.E.N. is supported by the US Department of Energy's Scientific Discovery through Advanced Computing programme. The A.V.F. group at the University of California, Berkeley is grateful for financial support from the US National Science Foundation, the US Department of Energy, the TABASGO Foundation, Gary and Cynthia Bengier, and the Richard and Rhoda Goldman Fund. J.D. is supported by the National Natural Science Foundation of China and by the Chinese 973 Program. This work is based in part on data from the W. M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and NASA; it was made possible by the generous financial support of the W. M. Keck Foundation. This work made use of the NASA/IPAC Extragalactic Database, which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA. R.C.T. is a Luis W. Alvarez Fellow at the Lawrence Berkeley National Laboratory. R.J.F is a Clay Fellow at the Harvard-Smithsonian Center for Astrophysics.

Author Contributions A.G.-Y. initiated, coordinated and managed the project, carried out photometric and spectroscopic analysis, and wrote the manuscript. P.M. was responsible for obtaining the Very Large Telescope late-time observations, carried out spectroscopic modelling and led the theoretical interpretation. E.O.O. led the Palomar photometry, obtained P200 and Keck observations, and performed the photometric calibration analysis. P.E.N. discovered SN 2007bi, identified its peculiarity and similarity to SN 1999as, initiated some of the early spectroscopic analysis and led the recovery of pre-discovery data from DeepSky and the Catalina Real-Time Transient Survey. S.R.K., M.M.K. and R.M.Q. obtained key late-time Keck spectra and helped with the P60 observations. A.V.F., S.B.C. and R.C. analysed early Keck data and contributed to manuscript preparation and editing, including final proofreading (A.V.F.). R.W. and D.K. carried out custom PISN modelling for comparison with the observations. M.S. undertook custom reduction of the key late-time Keck spectrum. E.C.B. is the principal investigator for the CSS, and his team acquired the CSS data and provided preliminary calibration of the results. A.J.D. helped recover CSS data and advised about their calibration. R.C.T. analysed early spectra using his automated SYNOW code. J.S.B., D.P. and A.A.M. obtained early spectroscopic observations of SN 2007bi as well as infrared observations using the Peters Automated Infrared Imaging Telescope, and contributed to analysis and manuscript editing. R.J.F. and J.M.S. contributed to spectral observations and reductions, and advised during manuscript preparation. I.A. helped with P60 photometry and calibration, and with manuscript editing. R.S.E. obtained Keck observations of SN 2007bi. J.D. contributed to the Very Large Telescope programme that resulted in the observations of SN 2007bi, and proofread the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.G.-Y. (avishay.gal-yam@weizmann.ac.il).

nature
Vol 462|3 December 2009|doi:10.1038/nature08609

LETTERS

Synthetic magnetic fields for ultracold neutral atoms

Y.-J. Lin[1], R. L. Compton[1], K. Jiménez-García[1,2], J. V. Porto[1] & I. B. Spielman[1]

Neutral atomic Bose condensates and degenerate Fermi gases have been used to realize important many-body phenomena in their most simple and essential forms[1–3], without many of the complexities usually associated with material systems. However, the charge neutrality of these systems presents an apparent limitation—a wide range of intriguing phenomena arise from the Lorentz force for charged particles in a magnetic field, such as the fractional quantum Hall effect in two-dimensional electron systems[4,5]. The limitation can be circumvented by exploiting the equivalence of the Lorentz force and the Coriolis force to create synthetic magnetic fields in rotating neutral systems. This was demonstrated by the appearance of quantized vortices in pioneering experiments[6–9] on rotating quantum gases, a hallmark of superfluids or superconductors in a magnetic field. However, because of technical issues limiting the maximum rotation velocity, the metastable nature of the rotating state and the difficulty of applying stable rotating optical lattices, rotational approaches are not able to reach the large fields required for quantum Hall physics[10–12]. Here we experimentally realize an optically synthesized magnetic field for ultracold neutral atoms, which is evident from the appearance of vortices in our Bose–Einstein condensate. Our approach uses a spatially dependent optical coupling between internal states of the atoms, yielding a Berry's phase[13] sufficient to create large synthetic magnetic fields, and is not subject to the limitations of rotating systems. With a suitable lattice configuration, it should be possible to reach the quantum Hall regime, potentially enabling studies of topological quantum computation.

In classical electromagnetism, the Lorentz force for a particle of charge q moving with velocity $\mathbf{v}$ in a magnetic field $\mathbf{B}$ is $\mathbf{v} \times q\mathbf{B}$. In the Hamiltonian formulation of quantum mechanics, where potentials play a more central role than fields, the single-particle Hamiltonian is $\mathcal{H} = \hbar^2 (\mathbf{k} - q\mathbf{A}/\hbar)^2 / 2m$, where $\mathbf{A}$ is the vector potential giving rise to the field $\mathbf{B} = \nabla \times \mathbf{A}$, $\hbar\mathbf{k}$ is the canonical momentum and m is the mass. In both formalisms, only the products $q\mathbf{B}$ and $q\mathbf{A}$ are important. To generate a synthetic magnetic field $\mathbf{B}^*$ for neutral atoms, we engineered a Hamiltonian with a spatially dependent vector potential $\mathbf{A}^*$ producing $\mathbf{B}^* = \nabla \times \mathbf{A}^*$.

The quantum mechanical phase is the relevant and significant quantity for charged particles in magnetic fields. A particle of charge q travelling along a closed loop acquires a phase $\phi = 2\pi \Phi_B/\Phi_0$ due to the presence of magnetic field $\mathbf{B}$, where Φ_B is the enclosed magnetic flux and $\Phi_0 = h/q$ is the flux quantum. A similar path-dependent phase, the Berry's phase[13], is the geometric phase acquired by a slowly moving particle adiabatically traversing a closed path in a Hamiltonian with position-dependent parameters. The Berry's phase depends only on the geometry of the parameters along the path, and is distinct from the dynamic contribution to the phase, which depends upon the speed of the motion.

The close analogy with the Berry's phase implies that properly designed position-dependent Hamiltonians for neutral particles can simulate the effect of magnetic fields on charged particles. We created such a spatially varying Hamiltonian for ultracold atoms by dressing them in an optical field that couples different spin states. The appropriate spatial dependence can originate from the laser beams' profile[10,14,15] or, as here, from a spatially dependent laser–atom detuning[16]. An advantage of this optical approach over rotating gases is that the synthetic field exists at rest in the laboratory frame, allowing all trapping potentials to be time-independent.

The large synthetic magnetic fields accessible by this approach make possible the study of unexplored bosonic quantum-Hall states, labelled by the filling factor $\nu = \Phi_B/\Phi_0$, the ratio of atom number to the number of flux quanta. The outstanding open questions in quantum-Hall physics centre on systems whose elementary quasiparticle excitations are anyons: neither bosons nor fermions. In some cases these anyons may be non-abelian, meaning that moving them about each other can implement quantum gates, which makes non-abelian anyons of great interest for this 'topological' quantum computation[17]. In electronic systems, the observed $\nu = 5/2$ quantum-Hall state may be such a system, but its true nature is still uncertain[18]. In contrast, the $\nu = 1$ bosonic quantum-Hall state with contact interactions has the same non-abelian anyonic excitations as the $\nu = 5/2$ state in electronic systems is hoped to have[19].

To engineer a vector potential $\mathbf{A}^* = A_x^* \hat{x}$, we illuminated a ^{87}Rb Bose–Einstein condensate (BEC) with a pair of Raman laser beams with momentum difference along $\hat{x}$ (Fig. 1a). These coupled together the three spin states, $m_F = 0$ and ± 1, of the $5S_{1/2}$, $F = 1$ electronic ground state (Fig. 1b), producing three dressed states whose energy–momentum dispersion relations $E_j(k_x)$ are experimentally tunable. Example dispersions are illustrated in Fig. 1c. The lowest of these, with minimum at k_{min}, corresponds to a term in the Hamiltonian associated with the motion along $\hat{x}$, namely $\mathcal{H}_x^* \approx \hbar^2 (k_x - k_{\mathrm{min}})^2 / 2m^* = \hbar^2 (k_x - q^* A_x^*/\hbar)^2 / 2m^*$, where A_x^* is an engineered vector potential that depends on an externally controlled Zeeman shift for the atom with synthetic charge q^*, and m^* is the effective mass along $\hat{x}$. To produce the desired spatially dependent $A_x^*(y)$ (see Fig. 1d), generating $-B^* \hat{z} = \nabla \times \mathbf{A}^*$, we applied a Zeeman shift that varied linearly along $\hat{y}$. The resulting B^* was approximately uniform near $y = 0$, at which point $A_x^* = B^* y$. (Here, the microscopic origin of the synthetic Lorentz force[20] was optical along $\hat{x}$, depending upon the velocity along $\hat{y}$; the force along $\hat{y}$ was magnetic, depending upon the $\hat{x}$ velocity.) In this way, we engineered a Hamiltonian for ultracold atoms that explicitly contained a synthetic magnetic field, with vortices in the ground state of a BEC. This is distinctly different from all existing experiments, where vortices are generated by phase imprinting[21,22], rotation[7–9], or a combination thereof[23]. Each of these earlier works presents a different means of imparting angular momentum to the system yielding rotation. Figure 1e shows an experimental image of the atoms with $B^* = 0$. Figure 1f, with $B^* > 0$, shows vortices. This demonstrates an observation of an optically induced synthetic magnetic field.

We created a ^{87}Rb BEC in a 1,064-nm crossed dipole trap, loaded into the lowest-energy dressed state[24] with atom number N up to

---

[1]Joint Quantum Institute, National Institute of Standards and Technology, and University of Maryland, Gaithersburg, Maryland, 20899, USA. [2]Departamento de Física, Centro de Investigación y Estudios Avanzados del Instituto Politécnico Nacional, México DF, 07360, México.

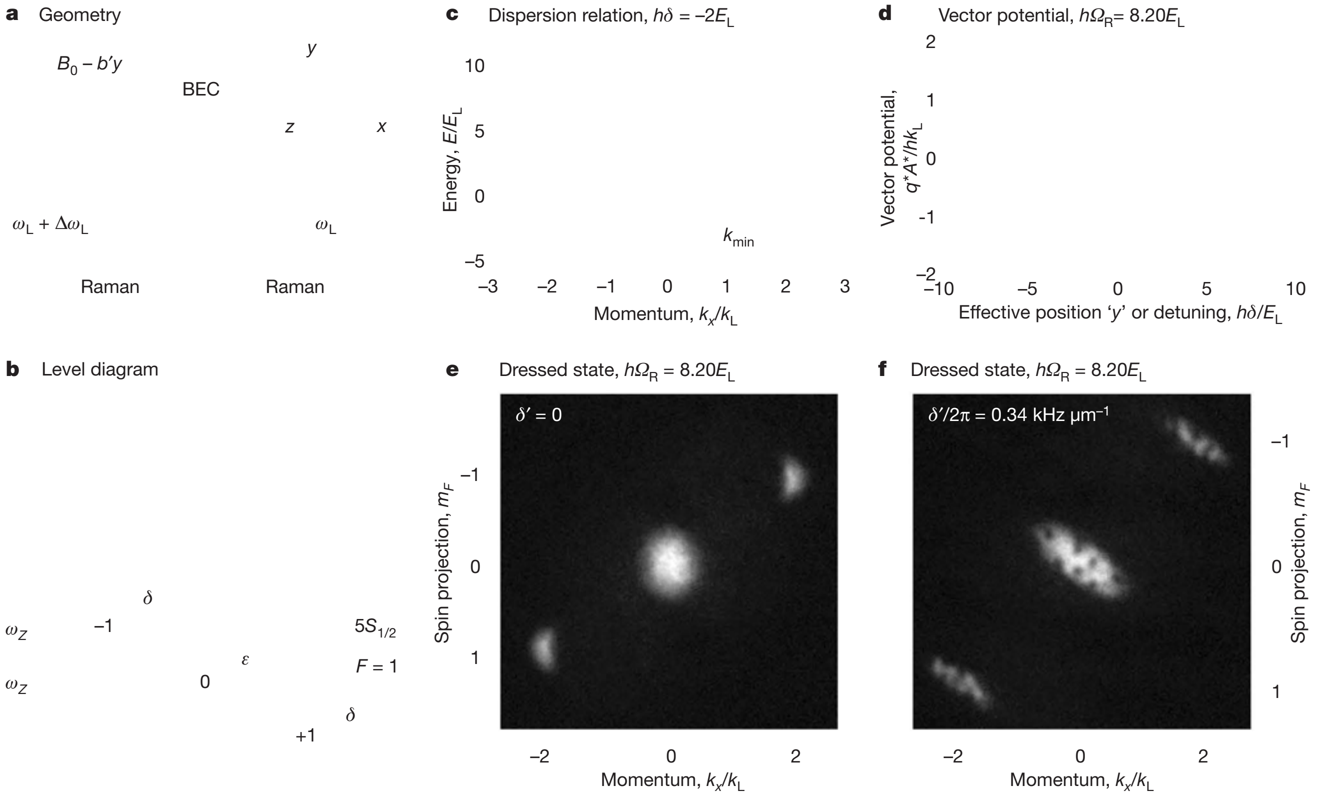


Figure 1 | Experiment summary for synthesizing magnetic fields. a, The BEC is in a crossed dipole trap in a magnetic field $B = (B_0 - b'y)\hat{y}$. Two Raman beams propagating along $\hat{y} \mp \hat{x}$ (linearly polarized along $\hat{y} \pm \hat{x}$) have frequencies ω_L and $\omega_L + \Delta\omega_L$. **b**, Raman coupling scheme within the $F = 1$ manifold: ω_Z and ε are the linear and quadratic Zeeman shifts, and δ is the Raman detuning. **c**, Energy–momentum dispersion relations. The grey curves represent the states without Raman coupling; the three coloured curves represent $E_j(k_x)$ of the dressed states. The arrow indicates the minimum at k_{min}. **d**, Vector potential $q^*A_x^* = \hbar k_{min}$ versus Raman detuning δ. The insets show the dispersion $E_1(k_x)$ for $\hbar\delta = 0$ (top inset) and $-2E_L$ (bottom inset). **e**, **f**, Dressed BEC imaged after a 25.1-ms TOF without (**e**) and with (**f**) a gradient. The spin components $m_F = 0$ and ± 1 separate along $\hat{y}$ owing to the Stern–Gerlach effect.

2.5×10^5, and a Zeeman shift $\omega_Z/2\pi = g\mu_B B/h \approx 2.71$ MHz, produced by a real magnetic bias field $B\hat{y}$. The $\lambda = 801.7$ nm Raman beams propagated along $\hat{y} \pm \hat{x}$ and differed in frequency by a constant $\Delta\omega_L \simeq \omega_Z$, where a small Raman detuning $\delta = \Delta\omega_L - \omega_Z$ largely determined the vector potential A_x^*. The scalar light shift from the Raman beams, combined with the dipole trap, gave an approximately symmetric three-dimensional potential, with frequencies $f_x, f_y, f_z \approx 70$ Hz. Here, $\hbar k_L = h/(\sqrt{2}\lambda)$ and $E_L = \hbar^2 k_L^2/2m$ are the appropriate units for the momentum and energy.

The spin and momentum states $|m_F, k_x\rangle$ coupled by the Raman beams can be grouped into families of states labelled by the momentum $\hbar k_x$. Each family $\Psi(k_x) = \{|-1, k_x+2k_L\rangle, |0, k_x\rangle, |+1, k_x - 2k_L\rangle\}$ is composed of states that differ in linear momentum along $\hat{x}$ by $\pm 2\hbar k_L$, and are Raman-coupled with strength $\hbar\Omega_R$. For each k_x, the three dressed states are the eigenstates in the presence of the Raman coupling, with energies[24] $E_j(k_x)$. The resulting vector potential is tunable within the range $-2k_L < q^*A_x^*/\hbar < 2k_L$. In addition, $E_j(k_x)$ includes a scalar potential[16] V'. A_x^*, V', and m^* are functions of Raman coupling Ω_R and detuning δ, and for our typical parameters $m^* \approx 2.5m$, reducing f_x from about 70 Hz to about 40 Hz. The BEC's chemical potential $\mu/h \approx 1$ kHz is much smaller than the $\sim h \times 10$ kHz energy separation between dressed states, so the BEC only occupies the lowest-energy dressed state. Further, it justifies the harmonic expansion around $q^*A_x^*/\hbar$, valid at low energy. Hence, the complete single-atom Hamiltonian is $\mathcal{H} = \mathcal{H}_x^* + \hbar^2\left(k_y^2 + k_z^2\right)/2m + V(\mathbf{r})$, where $V(\mathbf{r})$ is the external potential including $V'(\Omega_R, \delta)$.

The dressed BEC starts in a uniform bias field $B = B_0\hat{y}$, at Raman resonance ($\delta = 0$), corresponding to $A_x^* = 0$[24]. To create a synthetic field B^*, we applied a field gradient b' such that $B = (B_0 - b'y)\hat{y}$, ramping in 0.3 s from $b' = 0$ to a variable value up to 0.055 T m^{-1}, and then held it constant for t_h to allow the system to equilibrate. The detuning gradient $\delta' = g\mu_B b'/\hbar$ generates a spatial gradient in A_x^*. For the detuning range in our experiment, $\partial A_x^*/\partial\delta$ is approximately constant, leading to an approximately uniform synthetic field B^* given by $B^* = \partial A_x^*/\partial y = \delta' \partial A_x^*/\partial\delta$ (see Fig. 1d). To probe the dressed state, we switched off the dipole trap and the Raman beams in less than 1 μs, projecting each atom into spin and momentum components. We absorption-imaged the atoms after a time-of-flight (TOF) ranging from 10.1 ms to 30.1 ms (Fig. 1e, f).

For a dilute BEC in low synthetic fields, we expect to observe vortices. In this regime, the BEC is described by a macroscopic wavefunction $\psi(\mathbf{r}) = |\psi(\mathbf{r})|e^{i\phi(\mathbf{r})}$, which obeys the Gross–Pitaevskii equation (GPE). The phase ϕ winds by 2π around each vortex, with amplitude $|\psi| = 0$ at the vortex centre. The magnetic flux Φ_{B^*} results in N_v vortices and for an infinite, zero-temperature system, the vortices are arrayed in a lattice[25] with density q^*B^*/h. For finite systems vortices are energetically less favourable, and their areal density is below this asymptotic value, decreasing to zero at a critical field B_c^*. For a cylindrically symmetric BEC, B_c^* is given by $q^*B_c^*/h = 5/(2\pi R^2)\ln(0.67R/\xi)$ where R is the Thomas–Fermi radius and ξ is the healing length[26]. B_c^* is larger for smaller systems. For our non-cylindrically symmetric system, we numerically solve the GPE to determine B_c^* for our experimental parameters (see Methods).

For synthetic fields greater than the critical value, we observed vortices that enter the condensate and reach an equilibrium vortex number N_v after about 0.5 s. Owing to a shear force along $\hat{x}$ when the Raman beams are turned off, the nearly symmetric *in situ* atom cloud tilts during TOF. Although the vortices' positions may rearrange, any initial order is not lost. During the time of our experiment, the

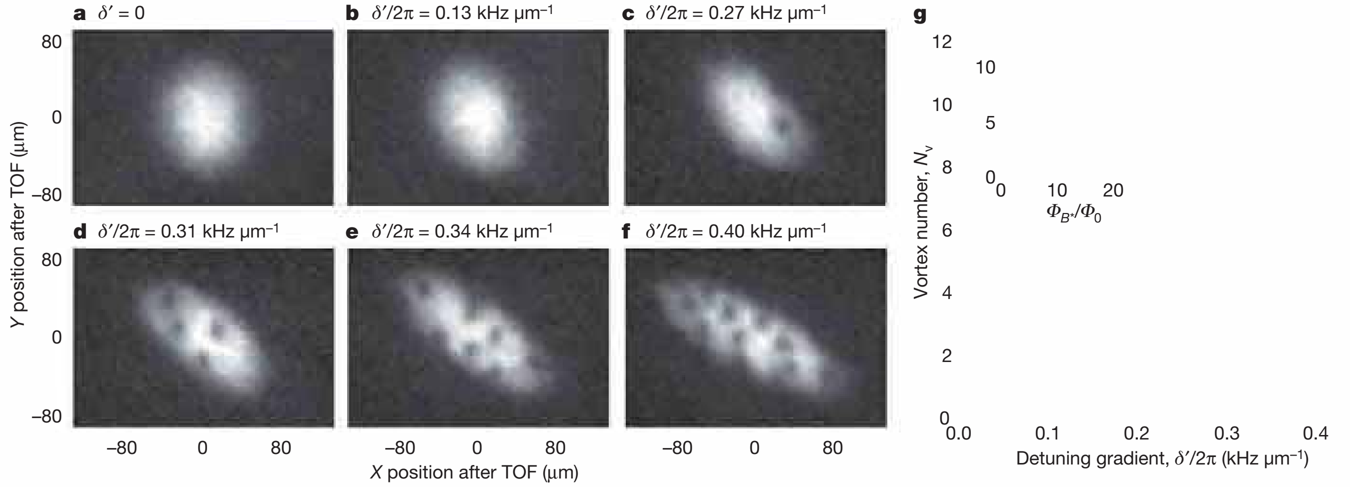


Figure 2 | Appearance of vortices at different detuning gradients. Data was taken for $N = 1.4 \times 10^5$ atoms at hold time $t_h = 0.57$ s. **a–f**, Images of the $|m_F = 0\rangle$ component of the dressed state after a 25.1-ms TOF with detuning gradient $\delta'/2\pi$ from 0 to 0.43 kHz μm^{-1} at Raman coupling $\hbar\Omega_R = 8.20E_L$. **g**, Vortex number N_v versus δ' at $\hbar\Omega_R = 5.85E_L$ (blue circles) and $8.20E_L$ (red circles). Each data point is averaged over at least 20 experimental realizations, and the uncertainties represent one standard deviation σ. The inset displays N_v versus the synthetic magnetic flux $\Phi_{B^*}/\Phi_0 = \mathcal{A}q^*\langle B^*\rangle/h$ in the BEC. The dashed lines indicate δ', below which vortices become energetically unfavourable according to our GPE computation, and the shaded regions show the 1σ uncertainty from experimental parameters.

vortices did not form a lattice and the positions of the vortices were irreproducible between different experimental realizations, consistent with our GPE simulations. We measured N_v as a function of detuning gradient δ' at two couplings, $\hbar\Omega_R = 5.85E_L$ and $8.20E_L$ (Fig. 2). For each Ω_R, vortices appeared above a minimum gradient when the corresponding field $\langle B^*\rangle = \delta'\langle\partial A_x^*/\partial\delta\rangle$ exceeded the critical field B_c^*. (For our coupling, B^* is only approximately uniform over the system and $\langle B^*\rangle$ is the field averaged over the area of the BEC.) The inset shows N_v for both values of Ω_R plotted versus $\Phi_{B^*}/\Phi_0 = \mathcal{A}q^*\langle B^*\rangle/h$, the vortex number for a system of area $\mathcal{A} = \pi R_x R_y$ with the asymptotic vortex density, where R_x (or R_y) is the Thomas–Fermi radius along $\hat{x}$ (or $\hat{y}$). The system size, and thus B_c^*, are approximately independent of Ω_R, so we expected this plot to be nearly independent of Raman coupling. Indeed, the data for $\hbar\Omega_R = 5.85E_L$ and $8.20E_L$ only deviated for $N_v < 5$, probably owing to the intricate dynamics of vortex nucleation[27].

Figure 3 illustrates a progression of images showing that vortices nucleate at the system's edge, fully enter to an equilibrium density and then decay along with the atom number. The timescale for vortex nucleation depends weakly on B^*, and is more rapid for larger B^* with more vortices. It is about 0.3 s for vortex number $N_v \geq 8$, and increases to about 0.5 s for $N_v = 3$. For $N_v = 1$ (B^* near B_c^*), the single vortex always remains near the edge of the BEC. In the dressed state, spontaneous emission from the Raman beams removes atoms from the trap, causing the population to decay with a 1.4(2)-s lifetime, and the equilibrium vortex number decreases along with the area of the BEC.

To verify that the dressed BEC has reached equilibrium, we prepared nominally identical systems in two different ways. First, we varied the initial atom number and measured N_v as a function of atom number N at a fixed hold time of $t_h = 0.57$ s. Second, starting with a large atom number, we measured both N_v and N, as they

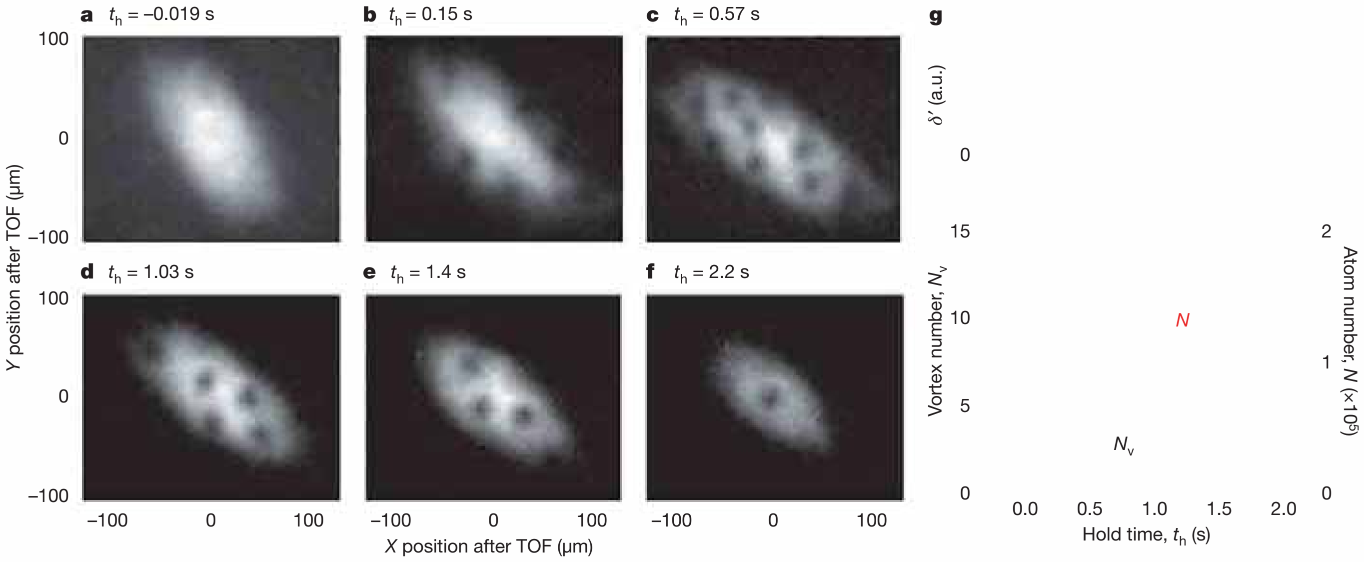


Figure 3 | Vortex formation. **a–f**, Images of the $|m_F = 0\rangle$ component of the dressed state after a 30.1-ms TOF for hold times t_h between -0.019 s and 2.2 s. The detuning gradient $\delta'/2\pi$ is ramped to 0.31 kHz μm^{-1} at the coupling $\hbar\Omega_R = 5.85E_L$. **g**, Top panel shows time sequence of δ'. (a.u., arbitrary units.) Bottom panel shows vortex number N_v (solid symbols) and atom number N (open symbols) versus t_h with a population lifetime of 1.4(2) s. The number in parentheses is the uncorrelated combination of statistical and systematic 1σ uncertainties.

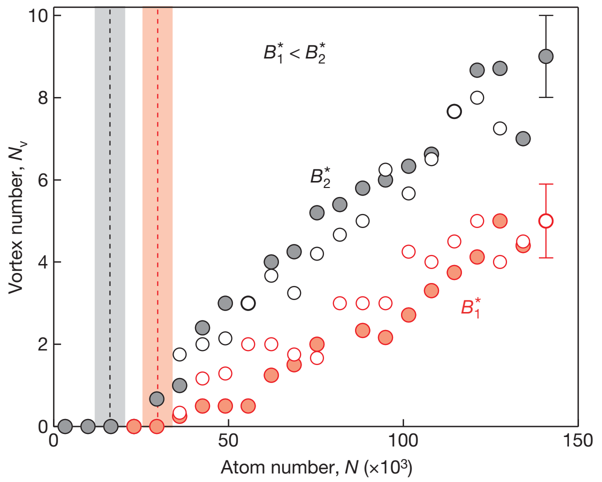


Figure 4 | Equilibrium vortex number. Vortex number N_v versus atom number N at detuning gradient $\delta_1'/2\pi = 0.26$ kHz μm^{-1} (red circles) and $\delta_2'/2\pi = 0.31$ kHz μm^{-1} (black circles), corresponding to synthetic fields $B_1^* < B_2^*$, at Raman coupling $\hbar\Omega_R = 5.85 E_L$. The two data points with the largest N show representative 1σ uncertainties, estimated from data in Fig. 2g. We vary N by its initial value with a fixed hold time $t_h = 0.57$ s (solid symbols), and by t_h with a fixed initial N (open symbols). The vertical dashed lines indicate N, below which vortices become energetically unfavourable computed using our GPE simulation. The shaded regions reflect the 1σ uncertainties from the experimental parameters.

decrease with t_h (Fig. 3). Figure 4 compares N_v versus N measured with both methods, each at two detuning gradients corresponding to fields $B_1^* < B_2^*$. The data show that N_v as a function of N is the same for these preparation methods, providing evidence that for $t_h \gtrsim 0.57$ s, N_v has reached equilibrium. As the atom number N falls, the last vortex departs the system when the critical field—increasing with decreasing N—surpasses the actual field.

In conclusion, we have demonstrated optically synthesized magnetic fields for neutral atoms resulting from the Berry's phase, a fundamental concept in physics. This novel approach differs from experiments with rotating gases, in which it is difficult to add optical lattices and rotation is limited by heating, metastability, and the difficulty of adding large angular momentum, preventing access to the quantum-Hall regime. A standout feature in our approach is the ease of adding optical lattices. For example, the addition of a two-dimensional (2D) lattice makes it immediately feasible to study the fractal energy levels of the Hofstadter butterfly[28]. Further, a one-dimensional lattice can divide the BEC into an array of 2D systems normal to the field. A suitable lattice configuration allows access to the $\nu \approx 1$ quantum-Hall regime, with an ensemble of 2D systems each with approximately 200 atoms, and with a realistic interaction energy of about $k_B \times 20$ nK.

METHODS SUMMARY

Dressed state preparation. We created a ^{87}Rb BEC in a crossed dipole trap[29], with $N \approx 4.7 \times 10^5$ atoms in $|F = 1, m_F = -1\rangle$. The quadratic Zeeman shift was $\hbar\varepsilon = 0.61 E_L$ for $\omega_Z/2\pi = g\mu_B B/h \approx 2.71$ MHz, where g is the Landé g-factor. To maintain $\delta = 0$ at the BEC's centre as we ramped the field gradient b', we changed $g\mu_B B_0$ by as much as $7E_L$. Simultaneously, we decreased the dipole beam power by 20%, producing our approximately 40-Hz trap frequency along $\hat{x}$. Additionally, the detuning gradient $\delta'\hat{y}$ made the scalar potential V' anti-trapping along $\hat{y}$, reducing f_y from 70 Hz to 50 Hz for our largest δ'. Spontaneous emission from the Raman beams decreased the atom number to $N \approx 2.5 \times 10^5$ for $t_h = 0$, with a condensate fraction of 0.85.

Numerical method. We compared our data to a finite temperature 2D stochastic GPE[30] simulation including the dressed state dispersion $E(k_x, y)$ that depends on y through the detuning gradient δ'. We evolved the time-dependent projected GPE:

$$i\hbar\frac{\partial\psi(\mathbf{x},t)}{\partial t} = \mathcal{P}\left\{\left[E\left(-i\hbar\frac{\partial}{\partial x}, y\right) - \frac{\hbar^2}{2m}\frac{\partial^2}{\partial y^2} + g_{2D}|\psi(\mathbf{x},t)|^2\right]\psi(\mathbf{x},t)\right\}$$

$\mathcal{P}$ projects onto a set of significantly occupied modes, and g_{2D} parameterizes the 2D interaction strength. The stochastic GPE models interactions between the highly occupied modes described by ψ and sparsely occupied thermal modes with dissipation and an associated noise term. We approximately accounted for the finite extent along $\hat{z}$ by making g_{2D} depend on the local 2D density. For low temperatures this 2D model correctly recovers the three-dimensional Thomas–Fermi radii, and gives the expected 2D density profile. These quantitative details are required to compute correctly the critical field or number for the first vortex to enter the system, which are directly tied to the 2D condensate area.

Received 7 August; accepted 20 October 2009.

1. Greiner, M., Mandel, O., Esslinger, T., Hänsch, T. W. & Bloch, I. Quantum phase transition from a superfluid to a Mott insulator in a gas of ultracold atoms. *Nature* **415,** 39–44 (2003).
2. Regal, C. A., Greiner, M. & Jin, D. S. Observation of resonance condensation of fermionic atom pairs. *Phys. Rev. Lett.* **92,** 040403 (2004).
3. Zwierlein, M. W. *et al.* Condensation of pairs of fermionic atoms near a Feshbach resonance. *Phys. Rev. Lett.* **92,** 120403 (2004).
4. Tsui, D. C., Stormer, H. L. & Gossard, A. C. Two-dimensional magnetotransport in the extreme quantum limit. *Phys. Rev. Lett.* **48,** 1559–1562 (1982).
5. Laughlin, R. B. Anomalous quantum Hall effect: an incompressible quantum fluid with fractionally charged excitations. *Phys. Rev. Lett.* **50,** 1395–1398 (1983).
6. Zwierlein, M. W., Abo-Shaeer, J. R., Schirotzek, A., Schunck, C. H. & Ketterle, W. Vortices and superfluidity in a strongly interacting Fermi gas. *Nature* **435,** 1047–1051 (2005).
7. Schweikhard, V., Coddington, I., Engels, P., Mogendorff, V. P. & Cornell, E. A. Rapidly rotating Bose-Einstein condensates in and near the lowest Landau level. *Phys. Rev. Lett.* **92,** 040404 (2004).
8. Madison, K. W., Chevy, F., Wohlleben, W. & Dalibard, J. Vortex formation in a stirred Bose-Einstein condensate. *Phys. Rev. Lett.* **84,** 806–809 (2000).
9. Abo-Shaeer, J. R., Raman, C., Vogels, J. M. & Ketterle, W. Observation of vortex lattices in Bose-Einstein condensates. *Science* **292,** 476–479 (2001).
10. Juzeliūnas, G. & Öhberg, P. Slow light in degenerate Fermi gases. *Phys. Rev. Lett.* **93,** 033602 (2004).
11. Jaksch, D. & Zoller, P. Creation of effective magnetic fields in optical lattices: the Hofstadter butterfly for cold neutral atoms. *N. J. Phys.* **5,** 56.1–56.11 (2003).
12. Sorensen, A. S., Demler, E. & Lukin, M. D. Fractional quantum Hall states of atoms in optical lattices. *Phys. Rev. Lett.* **94,** 086803 (2005).
13. Berry, M. V. Quantal phase factors accompanying adiabatic changes. *Proc. R. Soc. Lond. A* **392,** 45–57 (1984).
14. Juzeliūnas, G., Ruseckas, J., Öhberg, P. & Fleischhauer, M. Light-induced effective magnetic fields for ultracold atoms in planar geometries. *Phys. Rev. A* **73,** 025602 (2006).
15. Gunter, K. J., Cheneau, M., Yefsah, T., Rath, S. P. & Dalibard, J. Practical scheme for a light-induced gauge field in an atomic Bose gas. *Phys. Rev. A* **79,** 011604 (2009).
16. Spielman, I. B. Raman processes and effective gauge potentials. *Phys. Rev. A* **79,** 063613 (2009).
17. Nayak, C., Simon, S. H., Stern, A., Freedman, M. & Sarma, S. D. Non-abelian anyons and topological quantum computation. *Rev. Mod. Phys.* **80,** 1083–1159 (2008).
18. Radu, I. P. *et al.* Quasi-particle properties from tunneling in the formula fractional quantum Hall state. *Science* **320,** 899–902 (2008).
19. Cooper, N. R. Rapidly rotating atomic gases. *Adv. Phys.* **57,** 539–616 (2008).
20. Cheneau, M. *et al.* Geometric potentials in quantum optics: A semi-classical interpretation. *Europhys Lett.* **83,** 60001 (2008).
21. Leanhardt, A. E. *et al.* Imprinting vortices in a Bose-Einstein condensate using topological phases. *Phys. Rev. Lett.* **89,** 190403 (2002).
22. Andersen, M. F. *et al.* Quantized rotation of atoms from photons with orbital angular momentum. *Phys. Rev. Lett.* **97,** 170406 (2006).
23. Matthews, M. R. *et al.* Vortices in a Bose-Einstein condensate. *Phys. Rev. Lett.* **83,** 2498–2501 (1999).
24. Lin, Y.-J. *et al.* Bose-Einstein condensate in a uniform light-induced vector potential. *Phys. Rev. Lett.* **102,** 130401 (2009).
25. Yarmchuk, E. J., Gordon, M. J. V. & Packard, R. E. Observation of stationary vortex arrays in rotating superfluid helium. *Phys. Rev. Lett.* **43,** 214–217 (1979).
26. Lundh, E., Pethick, C. J. & Smith, H. Zero-temperature properties of a trapped Bose-condensed gas: beyond the Thomas-Fermi approximation. *Phys. Rev. A* **55,** 2126–2131 (1997).
27. Murray, D. R., Öhberg, P., Gomila, D. & Barnett, S. M. Vortex nucleation in Bose-Einstein condensates due to effective magnetic fields. *Phys. Rev. A* **79,** 063618 (2009).
28. Hofstadter, D. R. Energy levels and wave functions of Bloch electrons in rational and irrational magnetic fields. *Phys. Rev. B* **14,** 2239–2249 (1976).
29. Lin, Y.-J., Perry, A. R., Compton, R. L., Spielman, I. B. & Porto, J. V. Rapid production of ^{87}Rb Bose-Einstein condensates in a combined magnetic and optical potential. *Phys. Rev. A* **79,** 063631 (2009).
30. Blakie, P. B., Bradley, A. S., Davis, M. J., Ballagh, R. J. & Gardiner, C. W. Dynamics and statistical mechanics of ultra-cold Bose gases using c-field techniques. *Adv. Phys.* **57,** 363–455 (2008).

Acknowledgements We thank W. D. Phillips for discussions. This work was partially supported by ONR, ARO with funds from the DARPA OLE programme, and the NSF through the JQI Physics Frontier Center. R.L.C. acknowledges the NIST/NRC postdoctoral programme and K.J.-G. thanks CONACYT.

Author Contributions All authors contributed to writing of the manuscript. Y.-J.L. led the data collection effort (with assistance from R.L.C. and K.J.-G.). I.B.S. and J.V.P. designed the original apparatus, which was largely constructed by I.B.S., and Y.-J.L. implemented the specific changes required for the present experiment. I.B.S. conceived the experiment and performed numerical and analytic calculations. This work was supervised by I.B.S. with consultations from J.V.P.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to I.B.S. (ian.spielman@nist.gov).

Vol 462 | 3 December 2009 | **doi:10.1038/nature08584**

Controlling photonic structures using optical forces

Gustavo S. Wiederhecker[1], Long Chen[1], Alexander Gondarenko[1] & Michal Lipson[1]

The use of optical forces to manipulate small objects is well known. Applications include the manipulation of living cells by optical tweezers[1] and optical cooling in atomic physics[2]. The miniaturization of optical systems (to the micro and nanoscale) has resulted in very compliant systems with masses of the order of nanograms, rendering them susceptible to optical forces[3–6]. Optical forces have been exploited to demonstrate chaotic quivering of microcavities[7], optical cooling of mechanical modes[8–11], actuation of a tapered-fibre waveguide and excitation of the mechanical modes of silicon nano-beams[12,13]. Despite recent progress in this field[14–17], it is challenging to manipulate the optical response of photonic structures using optical forces; this is because of the large forces that are required to induce appreciable changes in the geometry of the structure. Here we implement a resonant structure whose optical response can be efficiently statically controlled using relatively weak attractive and repulsive optical forces. We demonstrate a static mechanical deformation of up to 20 nanometres in a silicon nitride structure, using three milliwatts of continuous optical power. Because of the sensitivity of the optical response to this deformation, such optically induced static displacement introduces resonance shifts spanning 80 times the intrinsic resonance linewidth.

The photonic structure is composed of two coupled ring resonators with an optical response sensitive to small changes in the distance between the rings. As shown in Fig. 1a, the structure is composed of a pair of vertically stacked silicon nitride (Si_3N_4) rings held by very thin spokes and a pedestal. The spokes are designed to reduce the mechanical stiffness of the cavity, increasing the sensitivity to the optical forces between the rings, and at the same time induce only negligible scattering for light circulating near the outside edge of the rings[18] (see Methods for fabrication details). The relatively small refractive index of Si_3N_4 ($n \approx 2.0$) makes strong optical coupling possible for relatively large gaps ($g \approx 640$ nm) between the top and the bottom resonator. As a result of this coupling, the transverse mode profile splits into symmetric (S, Fig. 1b) and antisymmetric (AS, Fig. 1c) combinations, leading to two distinct resonant frequencies. Figure 1d shows an example of the resonance wavelength splitting (centred at 1,493 nm) for the transverse electric (TE) mode of a 30-μm-diameter ring, obtained from numerical simulations (see Supplementary Information). The splitting clearly depends exponentially on the separation gap. Such steep dependence of the resonant frequencies on the gap between the rings translates into strong gradient optical forces (per photon) between the rings[5,6,10,17] (see Supplementary Information for details). In contrast to very recent demonstrations of tunable bipolar optical forces in waveguides[15,16], where modulated light was used to enhance the mechanical vibrational amplitude, optical forces in our device can be tuned to attractive or repulsive by tuning a continuous-wave laser frequency to either the S or AS resonance. Here we show that continuous-wave laser light at milliwatt power levels can induce significant static attraction or repulsion between two coupled resonant cavities through gradient optical forces.

Both symmetric and antisymmetric optical modes can be observed in the transmission spectrum of the fabricated structure, indicating the coupling between the two cavities. To ensure efficient coupling to both S and AS cavity modes, we use a monolithic integrated bus waveguide (see Fig. 1a). In Fig. 2a we show the optical transmission for input light with TE polarization for the 30-μm-diameter ring shown in Fig. 1a. The strong coupling in our cavity makes the resonance frequency splitting so large that it spans more than one cavity's free spectral range. As a result, the two neighbouring S and AS optical modes in Fig. 2a correspond to two azimuthal orders. For this device, the coupling gap between the bus waveguide and the rings is 650 nm. Each mode can be identified by comparing its measured free spectral ranges with numerical simulations (see Supplementary Information for details).

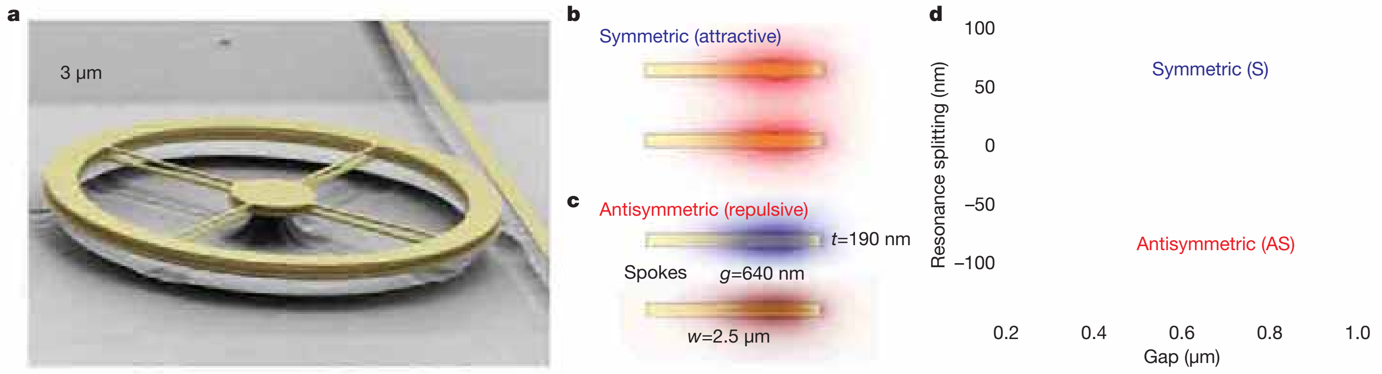


Figure 1 | Design of a resonant cavity sensitive to optical gradient forces. **a**, Scanning electron micrograph (SEM) of two vertically stacked ring cavities. **b**, **c**, Cross-section of the ring region, showing the calculated electric field profile for the symmetric (**b**) and antisymmetric (**c**) optical modes supported by the cavity. The ring thickness is $t = 190$ nm, width $w = 2.5$ μm and inter-ring gap $g \approx 640$ nm. The symmetric mode induces attractive optical forces, and the antisymmetric mode induces repulsive forces. **d**, Typical simulated resonant wavelength splitting of the S and AS optical modes as the gap between the rings is changed.

[1] School of Electrical and Computer Engineering, Cornell University, Ithaca, New York 14853, USA.

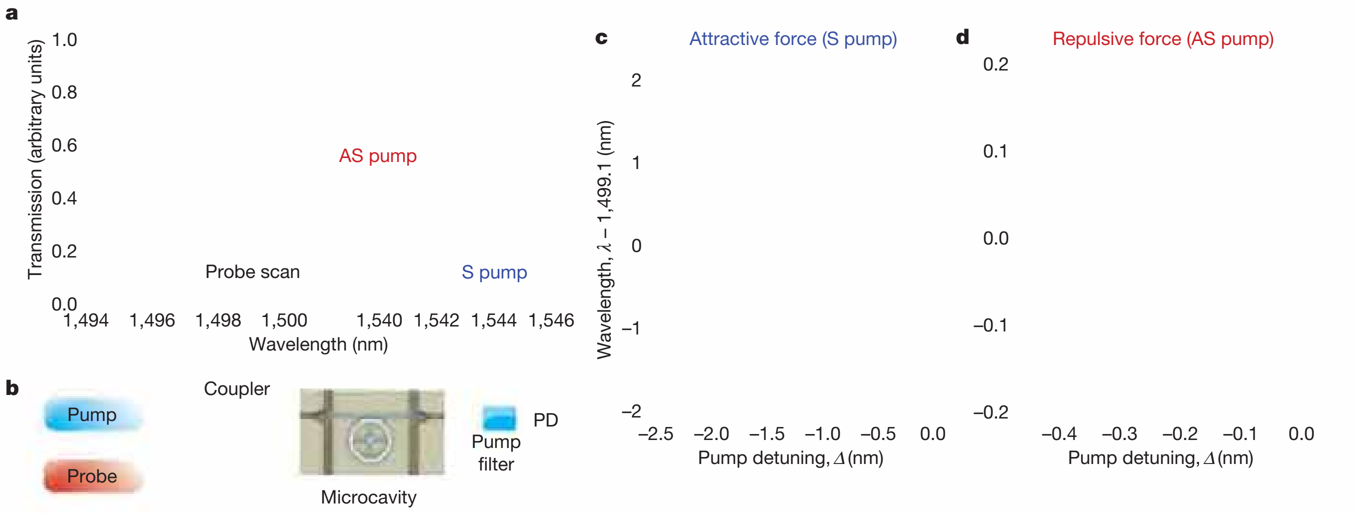


Figure 2 | Demonstration of gradient force control of cavity resonances. **a**, Optical transmission of the fabricated device shown in Fig. 1a. The different colours in the transmission spectrum indicate different optical modes. Blue, the fundamental symmetric TE mode ($TE_1^{\ +}$); red, the fundamental antisymmetric (AS) mode ($TE_1^{\ -}$); green, the second-order symmetric mode ($TE_2^{\ +}$). **b**, Schematic of the experimental set-up. The pump and probe laser light are coupled and launched into the device, and the residual pump is filtered out and detected by the photodiode (PD). **c**, Measured resonance splitting (centred at 1,499.1 nm) when the pump laser is tuned towards a symmetric resonance (blue arrow in panel **a**), inducing attractive forces. The blue circles represent the probed S mode and the red circles represent the probed AS mode. **d**, As in **c** but with the pump laser tuned to an antisymmetric resonance (red arrow in panel **a**), inducing repulsive forces. The cartoon insets to **c** and **d** illustrate the optical-force-induced deformation (out of scale) of the structure.

The typical loaded optical quality factors for the fundamental $TE_1^{\ +}$ and TE^- resonances are $Q = 6.8 \times 10^4$ (finesse $F = 720$) and $Q = 2.1 \times 10^4$ ($F = 200$), respectively. We attribute this difference mostly to the asymmetry between the top and the bottom ring geometry and material properties (see Methods and Supplementary Information for details).

We demonstrate experimentally the control of the photonic structure's response using both attractive and repulsive forces, that is, the tuning of the resonance splitting with optical forces. In contrast to other tuning methods based on thermal effect or free-carrier injection, this approach allows the cavity resonances to be moved to either shorter or longer wavelengths. As illustrated in the schematic of Fig. 2b, we use a pump laser to induce the optical force and a weak probe laser to read out the cavity response (see Methods for details). Attractive and repulsive forces are obtained by tuning the control laser towards a symmetric (blue arrow in Fig. 2a) or antisymmetric optical resonance (red arrow in Fig. 2a). As the detuning between the laser frequency and the cavity resonance is reduced, more energy is stored in the cavity, increasing the optical force between the rings. For each detuning between the pump laser and the respective cavity resonance, the probe laser wavelength is swept across a pair of S and AS resonances (shaded region in Fig. 2a) to read out the resonance shift. We note that the detuning range shown in Fig. 2c and d is significantly larger than the cavity linewidths shown in Fig. 2a. This is because the S/AS pump resonances also shift as the pump laser approaches the respective cavity resonant wavelength (see Supplementary Information for details). The probed wavelength shifts (relative to the central wavelength, $\lambda = 1{,}499.1$ nm) for both S excitation (attractive forces) and AS excitation (repulsive forces) are plotted in Fig. 2c and d. In Fig. 2c, while the AS resonance (red circles) is blue-shifted, the S resonance (blue circles) is red-shifted. This opposing behaviour can only occur if the gap between the two rings is reducing as the pump laser approaches the symmetric cavity resonance.

In Fig. 2d we show the wavelength splitting measured when the AS mode is excited. In contrast to the S excitation shown in Fig. 2c, the AS resonance is red-shifted while the S resonance is blue-shifted, indicating that the gap between the two rings is now increasing as the pump laser approaches the antisymmetric cavity resonance. The different slopes observed for attractive and repulsive pump schemes are due to the different quality factors and extinction ratios for the resonances. The observed opposite resonance shift uniquely demonstrates that optical gradient force is the dominant actuation effect in our cavity. Moreover, the demonstration that both attractive and repulsive optical forces can be induced in a microcavity structure paves the way towards devices with tailored optical potentials. Such tailorings could be used to explore novel effects and functionalities in optomechanical devices, such as self-alignment and or cavity trapping with optical forces[6].

Optical absorption means that thermal effects are also expected to induce resonance shifting. As the cavity temperature increases, the refractive index of Si_3N_4 will increase, owing to the thermo-optic effect. Also, because of the thermal expansion, the cavity structure may be deformed, inducing a thermo-mechanical effect. The thermo-optic effect causes the resonances to red-shift, creating thus an asymmetry between S and AS splitting (see Supplementary Information for details). Such thermo-optic asymmetry can be seen in Fig. 2c, where the S-mode red-shift is larger than the AS-mode blue-shift. In Fig. 2d, owing to the repulsive optical force, the asymmetry is reversed and the S-mode blue-shift is smaller than the AS-mode red-shift. The thermo-mechanical effect, however, can actually deform the mechanical structure. This deformation can either increase or decrease the gap between the rings. A rise in temperature, however, will always induce the same thermo-mechanical deformation, regardless of the optical mode that is providing the heat source. This is in clear contrast with the measurements shown in Fig. 2c and d, indicating that the thermo-mechanical effect is much weaker than the optomechanical effect. To infer its magnitude in our structure, we performed full numerical simulations of coupled thermo-mechanical problem using the finite-element method (see Supplementary Information). The results show that even a very large temperature increase of $\Delta T = 70$ K would change the gap between the rings by only 100 pm. Its contribution is so small because the thermal expansion is mainly in the radial direction, leaving the gap between the rings almost unchanged.

From the measured net resonance shift for both S and AS modes, the thermo-optic ($\Delta\lambda_{\mathrm{to}}$) and the optomechanical ($\Delta\lambda_{\mathrm{om}}$) contributions to the net resonance shift can be isolated. This is achieved by considering the measured group index of the optical modes, the thermo-optic coefficient of Si_3N_4 and the optomechanical tuning efficiency ($k_{\mathrm{om}} = \partial\lambda/\partial g$, where λ is the resonance wavelength and g is the gap between the rings) (see Methods and Supplementary Information). The optomechanical tuning efficiencies for S and AS

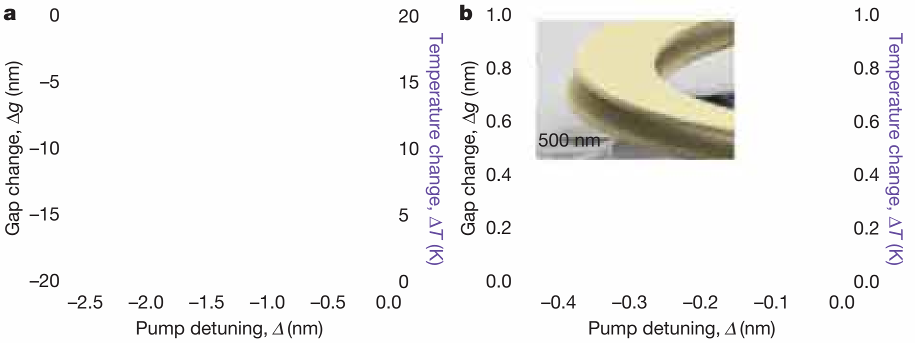


Figure 3 | Inter-cavity gap and temperature change. Change in the gap between the resonators (black circles, left scale) induced by the optical force, and the temperature increase due to optical absorption (purple squares, right scale). **a**, Symmetric pump (attractive forces). **b**, Antisymmetric pump (repulsive forces). The inset to **b** shows a zoomed-in SEM image of the cavity. The scale bar is 500 nm.

resonances are obtained from the numerical simulation of the cavity's optical response. For a 640-nm inter-ring gap the numerical values are $k_{\mathrm{om}}^{+} = -0.07$ and $k_{\mathrm{om}}^{-} = 0.1$. The maximum optomechanical shift obtained for attractive forces is calculated to be $\Delta\lambda_{\mathrm{om}}^{+} \approx 1.5$ nm (about 80 linewidths) for the S mode; $\Delta\lambda_{\mathrm{om}}^{-} \approx -2$ nm (about 30 linewidths) for the AS mode. The thermo-optic contribution is similar for both modes and is calculated to be $\Delta\lambda_{\mathrm{to}}^{\pm} \approx 0.6$ nm.

The change in the gap between the two rings induced by the optical force can also be estimated from the splitting curves shown in Fig. 2c and d and the optomechanical tuning efficiencies. The maximum gap change is found to be $\Delta g^{+} = -20$ nm for attractive forces and $\Delta g^{-} = 1$ nm for repulsive forces. This difference is mostly due to the different quality factors and extinction ratios of the modes. The tuning of this gap and temperature increase with respect to the pump laser detuning is shown in Fig. 3a and b. Using a spring constant of $k = 1.2\ \mathrm{N\,m^{-1}}$, calculated from numerical simulations of the cavity's mechanical response, the optical force on each ring can be estimated from the usual linear relation $F^{+} = k\Delta g^{+}/2 \approx 7.5$ nN. This value is of the same order of magnitude as the value expected from the cavity energy U and the optomechanical tuning efficiency[5,6] $F^{+} = (U^{+}/\lambda)k_{\mathrm{om}}^{+} \approx 12$ nN. For repulsive forces this yields $F^{-} = k\Delta g^{-}/2 \approx 600$ pN, whereas the predicted force is $F^{-} = (U^{-}/\lambda)k_{\mathrm{om}}^{-} \approx 350$ pN (see Supplementary Information for details).

We verify that the impact of undesirable optically induced mechanical oscillations in our structure is not detrimental by measuring the dynamic optical response (see Methods and Supplementary Information). Owing to its higher optical Q such oscillations would be more detrimental under symmetric pump excitation. However, because the optical quality factor ($Q = 6 \times 10^{4}$) of this S mode is still not very high and the mechanical quality factor ($Q_{\mathrm{m}} \approx 2$) is extremely low, the optical power threshold for regenerative mechanical oscillations[8] of the fundamental mechanical mode is estimated to be $P_{\mathrm{th}} \approx 65$ mW and was not achieved in our experiment. The thermal brownian motion of the cavity mechanical modes, however, can cause fluctuations in the gap between the rings. At thermal equilibrium, this amplitude can be calculated using the equipartition theorem to be $\sqrt{\langle \delta x^{2} \rangle} \approx 60$ pm. This corresponds to the relative wavelength shift $\langle \Delta\lambda \rangle/\delta\lambda = k_{\mathrm{om}}\sqrt{\langle \delta x^{2} \rangle}/\delta\lambda \approx 20\%$, where δλ is the resonance linewidth. In our cavity, however, a giant optical spring effect[19] greatly reduces such fluctuations. The strong optical spring effect is due to the small effective mass ($m_{\mathrm{eff}} \approx 85$ pg) and relatively large optomechanical tuning efficiency of the cavity. The frequency of the fundamental mechanical mode is observed to increase by 14-fold in our structure compared to its natural oscillation frequency, which corresponds to a 200-fold increase in the effective spring constant (see Supplementary Information). Because the thermal fluctuation scales with the inverse of the spring constant, the optical spring should reduce the thermal fluctuation to $\sqrt{\langle \delta x^{2} \rangle} \approx 4$ pm, corresponding to a relative wavelength fluctuation $\langle \Delta\lambda \rangle/\delta\lambda = k_{\mathrm{om}}\sqrt{\langle \delta x^{2} \rangle}/\delta\lambda \approx 1\%$. The relative power fluctuation of the transmitted pump laser was measured to be $\langle \Delta P \rangle / P \approx 0.05$ at small pump detunings. This corresponds to a relative wavelength shift of $\langle \Delta\lambda \rangle/\delta\lambda \approx 7\%$, which is larger than the predicted value. This larger predicted fluctuation is attributed to the reduction in the effective damping coefficient caused by the spring effect. Although these fluctuations are quite small, they could be further reduced by targeting an even lower mechanical quality factor cavity using, for instance, a pressurized environment or specific spoke design[18].

We show that significant optical force actuation using continuous-wave laser light is achievable by exploring interacting optical microcavities. The particular cavity design explored in this work provides high mechanical sensitivity to such forces. For instance, within the optical and mechanical quality factors explored here, the static optical force is dominant over both thermal effects and optomechanical oscillations. Further reduction in the mechanical quality factors, through squeeze-film damping[20], should render our static actuation scalable to even larger values. The demonstration of both attractive and repulsive forces in a resonant microstructure is an important step towards enabling recently proposed functionalities for optomechanically based devices, such as self-aligning and optical corralling behaviour in cavities[6]. Our monolithic integration with the bus waveguides also provides a more compact device that is more likely to be integrated into elaborate optical interconnect designs[21]. These advances on the static actuation of photonic microcavity structures using optical forces should enable future micro-optomechanical systems with novel and distinct functionalities.

METHODS SUMMARY

We fabricated our devices using electron beam lithography of silicon nitride and silicon oxide layers deposited over an oxidized silicon wafer. The patterned structure was then released using a wet hydrofluoric acid etch.

The experimental set-up used to demonstrate optical force tuning consisted of two external-cavity tunable lasers simultaneously launched into the bus waveguide of our device using a lensed optical fibre. The pump laser power coupled to the cavity was varied by changing the detuning between its frequency and the cavity resonant frequency. To avoid the instabilities associated with the red-side of the cavity resonance, the detuning was always to the blue-side (negative detuning).

The change in the separation gap between the two rings was obtained from the experimental data by considering both thermal and optomechanical effects. The optomechanical and thermal coupling parameters were obtained from experimental evidence and numerical simulations.

The dynamic response of the cavity was obtained by detecting the transmitted light with a 125-MHz bandwidth photodetector followed by a real-time oscilloscope and a radio frequency (r.f.) spectrum analyser. The oscilloscope was used to determine the relative oscillation amplitude whereas the r.f. analyser was used to identify the mechanical modes and their properties (see Supplementary Information for details).

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

Received 24 June; accepted 7 October 2009.

1. Ashkin, A. & Dziedzic, J. Optical trapping and manipulation of viruses and bacteria. *Science* 235, 1517–1520 (1987).

2. Hansch, T. & Schawlow, A. Cooling of gases by laser radiation. *Opt. Commun.* **13,** 68–69 (1975).
3. Antonoyiannakis, M. & Pendry, J. Electromagnetic forces in photonic crystals. *Phys. Rev. B* **60,** 2363–2374 (1999).
4. Ng, J., Chan, C., Sheng, P. & Lin, Z. Strong optical force induced by morphology-dependent resonances. *Opt. Lett.* **30,** 1956–1958 (2005).
5. Povinelli, M. *et al.* Evanescent-wave bonding between optical waveguides. *Opt. Lett.* **30,** 3042–3044 (2005).
6. Rakich, P., Solja, M. & Ippen, E. Trapping, corralling and spectral bonding of optical resonances through optically induced potentials. *Nature Photon.* **1,** 658–665 (2007).
7. Carmon, T., Cross, M. C. & Vahala, K. J. Chaotic quivering of micron-scaled on-chip resonators excited by centrifugal optical pressure. *Phys. Rev. Lett.* **98,** 167203 (2007).
8. Kippenberg, T. & Vahala, K. Cavity opto-mechanics. *Opt. Express* **15,** 17172–17205 (2007).
9. Kippenberg, T. & Vahala, K. Cavity optomechanics: back-action at the mesoscale. *Science* **321,** 1172–1176 (2008).
10. Lin, Q., Rosenberg, J., Jiang, X., Vahala, K. J. & Painter, O. Mechanical oscillation and cooling actuated by the optical gradient force *Phys. Rev. Lett.* **103,** 103601 (2009).
11. Schliesser, A., Rivière, R., Anetsberger, G., Arcizet, O. & Kippenberg, T. Resolved-sideband cooling of a micromechanical oscillator. *Nature Phys.* **4,** 415–419 (2008).
12. Li, M., Pernice, W. H. P. & Tang, H. X. Broadband all-photonic transduction of nanocantilevers. *Nature Nanotechnol.* **4,** 377–382 (2009).
13. Li, M. *et al.* Harnessing optical forces in integrated photonic circuits. *Nature* **456,** 480–484 (2008).
14. Eichenfield, M. & Painter, O. J. Optomechanics of strongly coupled stacked monolithic microdisks. *Conference on Lasers and Electro-Optics/Quantum Electronics and Laser Science Conference and Photonic Applications Systems Technologies* JMD1 ⟨http://www.opticsinfobase.org/abstract.cfm?URI=CLEO-2008-JMD1⟩ (2008).
15. Li, M., Pernice, W. H. P. & Tang, H. X. Tunable bipolar optical interactions between guided lightwaves. *Nature Photon.* **3,** 464–468 (2009).
16. Roels, J. *et al.* Tunable optical forces between nanophotonic waveguides. *Nature Nanotechnol.* **4,** 510–513 (2009).
17. Rosenberg, J., Lin, Q. & Painter, O. Static and dynamic wavelength routing via the gradient optical force. *Nature Photon.* **3,** 478–483 (2009).
18. Anetsberger, G., Rivi, R., Schliesser, A., Arcizet, O. & Kippenberg, T. Ultralow-dissipation optomechanical resonators on a chip. *Nature Photon.* **2,** 627–633 (2008).
19. Eichenfield, M., Camacho, R., Chan, J., Vahala, K. J. & Painter, O. A picogram- and nanometre-scale photonic-crystal optomechanical cavity. *Nature* **459,** 550–555 (2009).
20. Bao, M. & Yang, H. Squeeze film air damping in MEMS. *Sens. Actuat. A* **136,** 3–27 (2007).
21. Sherwood-Droz, N. *et al.* Optical 4x4 hitless silicon router for optical networks-on-chip (NoC). *Opt. Express* **16,** 15915–15922 (2008).

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

Acknowledgements This work was supported in part by the National Science Foundation under grant 00446571. We also acknowledge partial support by Cornell University's Center for Nanoscale Systems. This work was performed in part at the Cornell Nano-Scale Science and Technology Facility (a member of the National Nanofabrication Users Network) which is supported by the National Science Foundation, its users, Cornell University and Industrial users. G.S.W. thanks S. Lee for help in preparing some of the Supplementary Information.

Author Contributions G.S.W. designed, fabricated and tested the devices. L.C. helped in the design, fabrication and testing. A.G. helped with the fabrication. G.S.W., L.C, A.G. and M.L. discussed the results and their implications and contributed to writing this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.L. (ml292@cornell.edu).

METHODS

Device fabrication. The two layers (each 190-nm-thick) of stoichiometric Si_3N_4 are deposited using low-pressure chemical vapour deposition (CVD) while the 300-nm-thick SiO_2 layer was deposited by plasma-enhanced CVD. The underlying substrate is a 4-μm SiO_2 formed by thermal oxidation of a silicon wafer. Because the top Si_3N_4 layer is deposited over a plasma-enhanced CVD SiO_2, the surface roughness of the top layer is expected to be larger than the bottom layer, decreasing its optical quality. To have a single coupling waveguide, which increases the coupling and thus extinction ratio of the AS modes, we perform two lithography steps. First, circular pads are defined in the position where the cavities will be patterned. The upper Si_3N_4 layer is then etched everywhere else except the circular pads region. In the second lithography step, both the wheel pattern and the waveguides are defined, but the wheels are defined on top of the previously defined circular pads. Only then the intermediate SiO_2 and the bottom Si_3N_4 layer are etched. As a result the waveguides will have a single Si_3N_4 layer, whereas the wheels will have two layers of Si_3N_4. After depositing a 1-μm-thick protective SiO_2 cladding using plasma-enhanced CVD, we use optical lithography to pattern the spun photoresist with a rectangular window around the resonators. To release the structure, the device is immersed in buffered hydrofluoric acid for an isotropic etch of the SiO_2 in the window region. To avoid static friction of the stacked disks due to surface tension in the drying process, the device is subsequently dried using a critical point dryer. The internal tensile stress ($\sigma_0 = 1$ GPa) of the Si_3N_4 films causes a noticeable bending of the rings in the vertical direction (see the SEM in Fig. 1a). This bending means that the actual gap between the resonators is larger than the sacrificial SiO_2 layer thickness; it is estimated from the SEM image to be $g = 640 \pm 50$ nm.

Experimental set-up. The set-up used to actuate and probe the cavity is illustrated in Fig. 2b. It consisted of two tunable external cavity diode lasers combined using a 3-dB directional coupler. The laser light was coupled in and out of the chip using a pair of lensed optical fibres. The high insertion loss of our devices (12.0 ± 0.5 dB) means that the pump laser was pre-amplified using an erbium-doped fibre amplifier (not shown in the schematic of Fig. 2b). To minimize coupling fluctuations when the pump power is varied, we chose to keep the pump laser power constant and sweep the pump laser wavelength λ_p across the desired resonances (see Supplementary Information for details). To induce attractive forces, the pump was tuned to the symmetric optical resonance at $\lambda = 1{,}545$ nm. For repulsive forces the laser was tuned to the antisymmetric resonance at $\lambda = 1{,}555$ nm. In both cases, the laser approached the resonance from shorter wavelengths (blue-detuning). To monitor the pump-induced changes in the optical resonances, the probe laser wavelength was swept through a wavelength range spanning a pair of symmetric and antisymmetric resonances (see Fig. 2a). The light collected from the bus waveguide was split into two parts. One output was filtered with an interferometric filter (centre wavelength 1,550 nm, bandwidth 10 nm) to reject the residual pump light. The detected probe light provided the raw optical transmission spectra used to obtain the curves shown in Fig. 2c and d. The second output from the beam splitter was directly detected using a New Focus 1811 (125-MHz bandwidth) photodetector. Its photocurrent allowed us to measure the dynamical response of the microcavity and measure the r.f. spectrum of the transmitted pump laser (see Supplementary Information).

Coupled pump power calibration. The input pump power (P_{in}) launched in the coupling waveguides is estimated from the input and output loss of our devices (12 ± 0.5 dB each). To ensure that both input and output losses were the same we calibrated them based on the nonlinear wavelength shift obtained by launching light from opposite sides of our device.

Optical modes. Three optical modes can be identified in the measured transmission spectrum. As indicated in Fig. 2a, we identified them as the fundamental symmetric mode in blue ($TE_1{}^+$), fundamental antisymmetric mode in red ($TE_1{}^-$), and the second-order symmetric mode in green ($TE_2{}^+$). This identification is obtained by comparing the measured free-spectral ranges with the simulated ones (see Supplementary Information). Further evidence that this identification is correct comes from the actual wavelength shift experienced by each of these modes. Although we show only the wavelength shift for the S ($TE_1{}^+$) and AS ($TE_1{}^-$) modes on Fig. 2, we extended the probe laser scanning window to measure the $TE_2{}^+$ mode resonance shift. As shown in the Supplementary Information, the $TE_2{}^+$ experiences a red-shift for attractive forces (symmetric pump) as expected for a symmetric mode.

Gap change calculation. As detailed in the Supplementary Information, the thermo-mechanical effect is negligible in our structure. Therefore the measured resonance shifts are due to a combination of the thermo-optic and optomechanical effects. The measured shift for S (+) and AS (−) modes is given by $\Delta\lambda^{\pm} = k^{\pm}_{om}\Delta g + k^{\pm}_{to}\Delta T$. Both $\Delta\lambda^{+}$ and $\Delta\lambda^{-}$ can be measured in our cavities and both k_{om} and k_{to} can be estimated from simulations, so the system of two equations can be solved for the gap (Δg) and temperature change (ΔT). The corresponding contributions to the net resonance shift can easily be calculated as $\Delta\lambda^{\pm}_{om} = k^{\pm}_{om}\Delta g$ and $\Delta\lambda^{\pm}_{to} = k^{\pm}_{to}\Delta T$. Details of this calculation can be found in the Supplementary Information.

Dynamical response. The temporal response of the resonator was investigated by analysing the time-domain response of the transmitted pump laser using a 1-GHz real-time oscilloscope and a r.f. spectrum analyser. Under the maximum power explored here (3 mW of coupled pump power) the measured relative oscillation amplitude in the pump transmission was $\delta P/P \approx 0.05$. By fitting a lorentzian lineshape to the pump resonance, we can map this power fluctuation to a drift of the resonance wavelength. For small pump detuning (<5 pm) the estimated wavelength drift is approximately 1.6 pm, which corresponds roughly to 7% of the pump resonance linewidth. By inspecting the r.f. spectrum of the transmitted pump laser we also identified the mechanical resonances responsible for the observed oscillations (see Supplementary Information). The mechanical quality factors for the fundamental mechanical mode was measured as $Q_m = 2$.

Vol 462|3 December 2009|**doi:10.1038/nature08520**

LETTERS

Half-precessional dynamics of monsoon rainfall near the East African Equator

Dirk Verschuren[1], Jaap S. Sinninghe Damsté[2,3], Jasper Moernaut[4], Iris Kristen[5], Maarten Blaauw[6], Maureen Fagot[1], Gerald H. Haug[7,8] & CHALLACEA project members*

External climate forcings—such as long-term changes in solar insolation—generate different climate responses in tropical and high latitude regions[1]. Documenting the spatial and temporal variability of past climates is therefore critical for understanding how such forcings are translated into regional climate variability. In contrast to the data-rich middle and high latitudes, high-quality climate-proxy records from equatorial regions are relatively few[2–4], especially from regions experiencing the bimodal seasonal rainfall distribution associated with twice-annual passage of the Intertropical Convergence Zone. Here we present a continuous and well-resolved climate-proxy record of hydrological variability during the past 25,000 years from equatorial East Africa. Our results, based on complementary evidence from seismic-reflection stratigraphy and organic biomarker molecules in the sediment record of Lake Challa near Mount Kilimanjaro, reveal that monsoon rainfall in this region varied at half-precessional (~11,500-year) intervals in phase with orbitally controlled insolation forcing. The southeasterly and northeasterly monsoons that advect moisture from the western Indian Ocean were strengthened in alternation when the inter-hemispheric insolation gradient was at a maximum; dry conditions prevailed when neither monsoon was intensified and modest local March or September insolation weakened the rain season that followed. On sub-millennial timescales, the temporal pattern of hydrological change on the East African Equator bears clear high-northern-latitude signatures, but on the orbital timescale it mainly responded to low-latitude insolation forcing. Predominance of low-latitude climate processes in this monsoon region can be attributed to the low-latitude position of its continental regions of surface air flow convergence, and its relative isolation from the Atlantic Ocean, where prominent meridional overturning circulation more tightly couples low-latitude climate regimes to high-latitude boundary conditions.

Mechanistic analyses of past climate change in the tropics have been hampered by the fragmentary or poorly dated nature of palaeoclimate data from key regions; but also by ambiguity about whether low-latitude ocean records are representative of climate history on nearby continents[5], and by the difficulty of extracting independent information on past temperature and precipitation from tropical climate archives (Supplementary Discussion 1). To fill this knowledge gap, we reconstructed the history of rainfall and drought near the Equator in East Africa, using two complementary hydrological proxies extracted from the sediment record of Lake Challa, a crater lake on the lower east slope of Mt Kilimanjaro (3° 19′ S, 37° 42′ E; Supplementary Discussion 2). In this region bordering the western Indian Ocean, the direct climatic influence of Atlantic Ocean circulation is relatively limited, and seasonal migration of tropical convective activity (the Intertropical Convergence Zone, ITCZ) spans the widest latitude range in the world (Fig. 1b). The Challa climate-proxy record can therefore provide unique information on how varying rainfall contributions from the northeasterly (Northern Hemisphere autumn–winter) and southeasterly (Northern Hemisphere spring–summer) monsoons have shaped climate history at the Equator. We find that long-term rainfall variability in equatorial East Africa is a hybrid of patterns typically documented for the northern and southern tropics. The timing of major wet and dry periods over the past 25,000 years has important implications for the nature of ITCZ migration at orbital timescales, and for the cause of glacial-era drought at the Equator.

The moisture-balance history of equatorial East Africa is revealed in seismic-reflection data of Lake Challa's bottom deposits, and the branched and isoprenoid tetraether (BIT) index of soil bacterial versus aquatic archaeal membrane lipids extracted from these sediments (Supplementary Methods 2, 3). Past lake-level fluctuations and events of basin-slope failure recorded by seismic stratigraphy (Fig. 2h) correlate tightly to variations in the water content and lithology of the cored sequence (Supplementary Methods 2). Two older units of inferred lowstand date to the early late-glacial period (L6: ~20,500–14,500 calendar years before present, ~20.5–14.5 cal. kyr BP) and the Younger Dryas (L5: ~12.9–12.0 cal. kyr BP), when most East African lakes are known to have experienced low levels[4]. The four younger, Holocene lowstand units are dated to ~8.0–6.7 (L4), ~5.9–4.7 (L3), ~3.6–3.0 (L2) and ~0.7–0.6 cal. kyr BP (L1). Seismic-reflection evidence for lake-wide mass wasting events or more local basin-slope failures suggest that the lowstand episodes of units L4, L3 and L1 started with abrupt lake-level decline (Supplementary Methods 2). Another mass-wasting event, dated to ~23.4 cal. kyr BP, lacks clear evidence of having been followed by an episode of basin-focused sedimentation, suggesting that a Heinrich-2 (H2) period drought affected equatorial East Africa but was short-lived. The unit of strongly focused sedimentation L6′ represents an extreme lowstand below the mean level of L6, and is dated to ~17.0–16.4 cal. kyr BP, that is, early into the Heinrich-1 (H1) chronozone (Fig. 2h).

The Challa BIT-index record of rainfall-induced surface run-off (Fig. 2g; Supplementary Methods 3) more precisely indicates that post-glacial intensification of the southeasterly Indian Ocean monsoon (driven by increasing summer insolation over North Africa; Fig. 2d) started from 16.5 ± 0.3 cal. kyr BP, and continued into the early

---

[1]Limnology Unit, Department of Biology, Ghent University, Ledeganckstraat 35, 9000 Gent, Belgium. [2]Faculty of Geosciences, Utrecht University, PO Box 80021, 3508 TA Utrecht, The Netherlands. [3]Department of Marine Organic Biogeochemistry, NIOZ Royal Netherlands Institute for Sea Research, PO Box 59, 1790 AB Den Burg, The Netherlands. [4]Renard Centre of Marine Geology, Department of Geology and Soil Science, Ghent University, Krijgslaan 281 S8, 9000 Gent, Belgium. [5]GeoForschungsZentrum Potsdam, Sektion 3.3 Klimadynamik und Sedimente, Telegrafenberg, D-14473 Potsdam, Germany. [6]School of Geography, Archaeology and Palaeoecology, Queen's University Belfast, BT9 6AX Belfast, UK. [7]Geological Institute, Department of Earth Sciences, ETH Zürich, CH-8092 Zürich, Switzerland. [8]DFG Leibniz Center for Earth Surface Process and Climate Studies, Potsdam University, D-14476 Potsdam, Germany.
*Lists of participants and affiliations appear at the end of the paper.

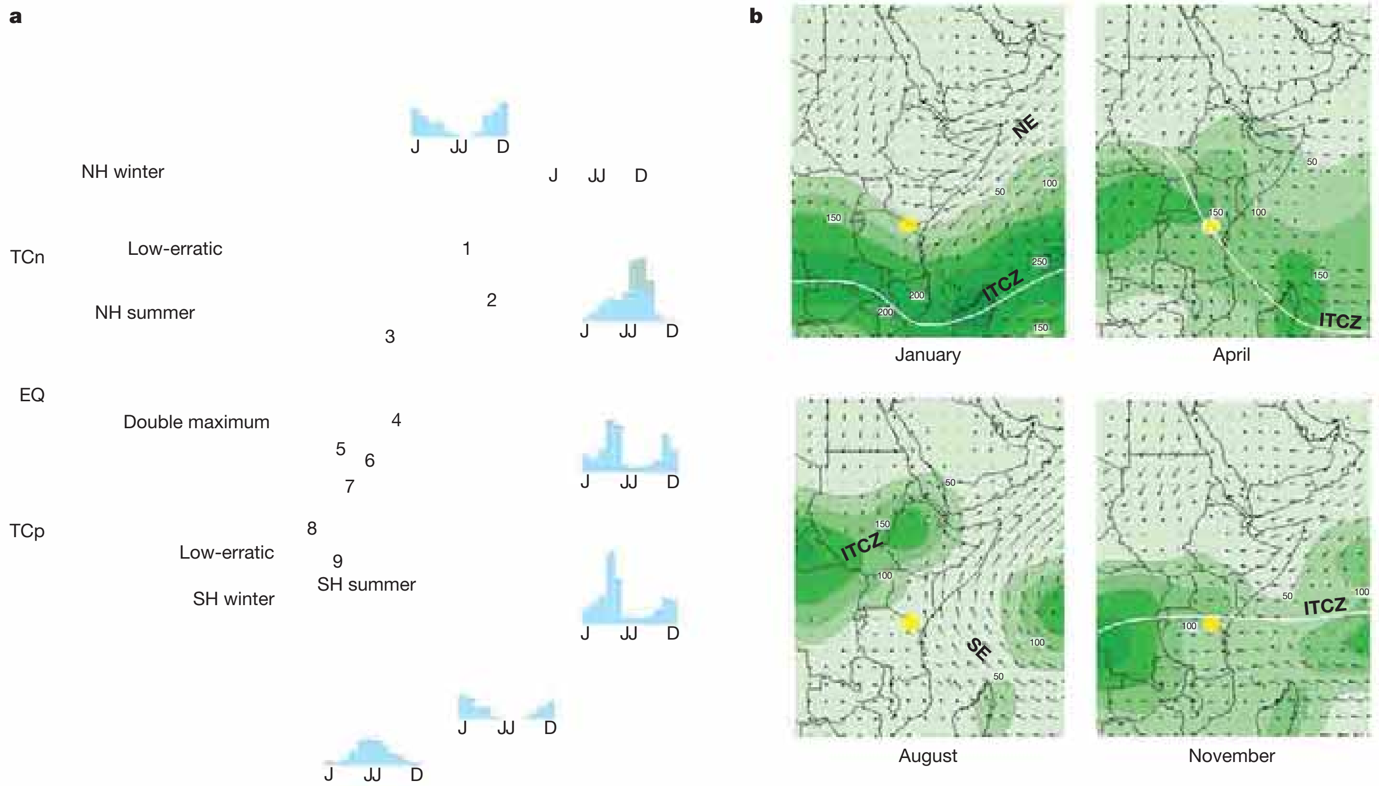


Figure 1 | Location and regional climate of the study site. a, The location of Lake Challa (yellow dot) in relation to the geographical distribution of seasonal rainfall regimes across Africa[4]. Insets, the principal rainfall regimes are illustrated by data on monthly rainfall (January through June/July to December; blue bars at scale intervals of 50 mm) together with the mean and daily range of temperature (red line and ochre shading at scale intervals of 5 °C) at seven representative climate stations; from top to bottom these are Algiers, Cairo, Addis Ababa, Nairobi, Dar Es Salaam, Gantsi and Cape Town. Numbered sites (1–9) refer to continental palaeodata in a latitudinal transect mostly bordering the western Indian Ocean, identified in Supplementary Fig. 5. NH, Northern Hemisphere; SH, Southern Hemisphere. EQ, Equator; TCn, Tropic of Cancer; TCp, Tropic of Capricorn. Boxed area shown magnified in **b**. **b**, Synoptic climatology of tropical East Africa for four representative months. The panels show the seasonal migration of the Intertropical Convergence Zone (ITCZ) and associated rain belts across the study region. Monthly precipitation (shaded green, with contours at 50 mm intervals) is based on gauge data over land, and satellite estimates over sea. Also indicated are wind fields for the 925 hPa pressure level (arrows), with wind speed proportional to arrow length (http://iridl.ldeo.columbia.edu).

Holocene, interrupted only by severe Younger Dryas aridity. Thus, monsoon circulation on the East African Equator started to intensify ~1,700 years before the Indian (that is, South Asian) and East Asian monsoons did so (at 14.8 cal. kyr BP: refs 6, 7 and Fig. 2c). The Asian monsoons are believed to have responded late to insolation forcing because the Arctic sea-ice expansion which cooled northern Eurasia during H1 initially weakened monsoonal moisture advection from the Indian and western Pacific oceans into southern Eurasia[8]. Also in the tropical Atlantic Ocean, H1-period slowdown of the meridional overturning circulation[9] and presumed southward shift of the ITCZ[10] delayed post-glacial strengthening of the West African monsoon to ~15 cal. kyr BP[11]. For East African regions receiving rainfall exclusively from the Indian Ocean, neither the source region of moist air (the southern Indian Ocean) nor its continental destination of surface flow convergence (the North African land mass) were strongly affected by H1 cooling[10], so moisture advection intensified as soon as enhanced Northern Hemisphere summer insolation created a sufficient land–sea temperature contrast between them[12]. Considering that tropical wetlands were the probable source of early late-glacial (pre-14.8 cal. kyr BP) CH_4 changes (Supplementary Discussion 3), eastern equatorial Africa must have been among those regions contributing to the rise in atmospheric CH_4 just before and during H1 (Fig. 2a). Throughout the Northern Hemisphere tropics, triggering of strong monsoon circulation was delayed pending favourable land and ocean thermal conditions, which initially remained under the control of climate events in the North Atlantic Ocean (Supplementary Discussion 4).

Our Challa BIT record further shows that drought encompassing the Younger Dryas in equatorial East Africa lasted from 13.3 to 11.7 cal. kyr BP, with most intense aridity during 12.4–12.8 cal. kyr BP, and that at the start of the Holocene the southeasterly monsoon regained its full strength in less than 200 years (Fig. 2g). Within dating error, this end to Younger Dryas drought in East Africa is coeval with corresponding signatures in records of the Indian[7], East Asian (ref. 6 and Fig. 2c), and tropical Atlantic monsoons[13], testifying to the rapid reorganization of low-latitude atmospheric circulation patterns following abrupt warming in the northern North Atlantic (Fig. 2b). Rapid resumption of ITCZ migration far into the Northern Hemisphere is also indicated by the abruptly intensifying southerly winds at ~11.7 cal. kyr BP across southern equatorial Africa (ref. 14 and Supplementary Discussion 5). The end of African Younger Dryas drought at Challa matches the tentative age assigned to the base of the current ice cap on Mt Kilimanjaro[15]. It thus appears that Mt Kilimanjaro's glacial-era ice cap (Supplementary Discussion 6), possibly in retreat since the start of regional post-glacial warming ~20 cal. kyr BP[16], was finally eradicated by severe Younger Dryas drought.

In its entirety, the 25,000-year Challa BIT record (Fig. 2g) implies relatively moist conditions in equatorial East Africa during the periods >25–20.5 (H2 excluded), 14.5–8.5 (the Younger Dryas excluded) and since 4.5 cal. kyr BP; and relative drought during the periods 20–16.5 and 8.5–4.5 cal. kyr BP. The Holocene part of this moisture-balance history differs markedly from that of sites in either northern tropical Africa, or southeastern tropical Africa from ~9° S (Supplementary Discussion 7). Today those regions experience a unimodal seasonal distribution of monsoon rainfall (Fig. 1a) in which the single wet season occurs during peak summer insolation over the corresponding hemisphere; and, like monsoon systems elsewhere[6,17,18], their long-term history reflects (more or less gradual) variation in monsoon strength that is controlled by precession-driven variation in Northern or

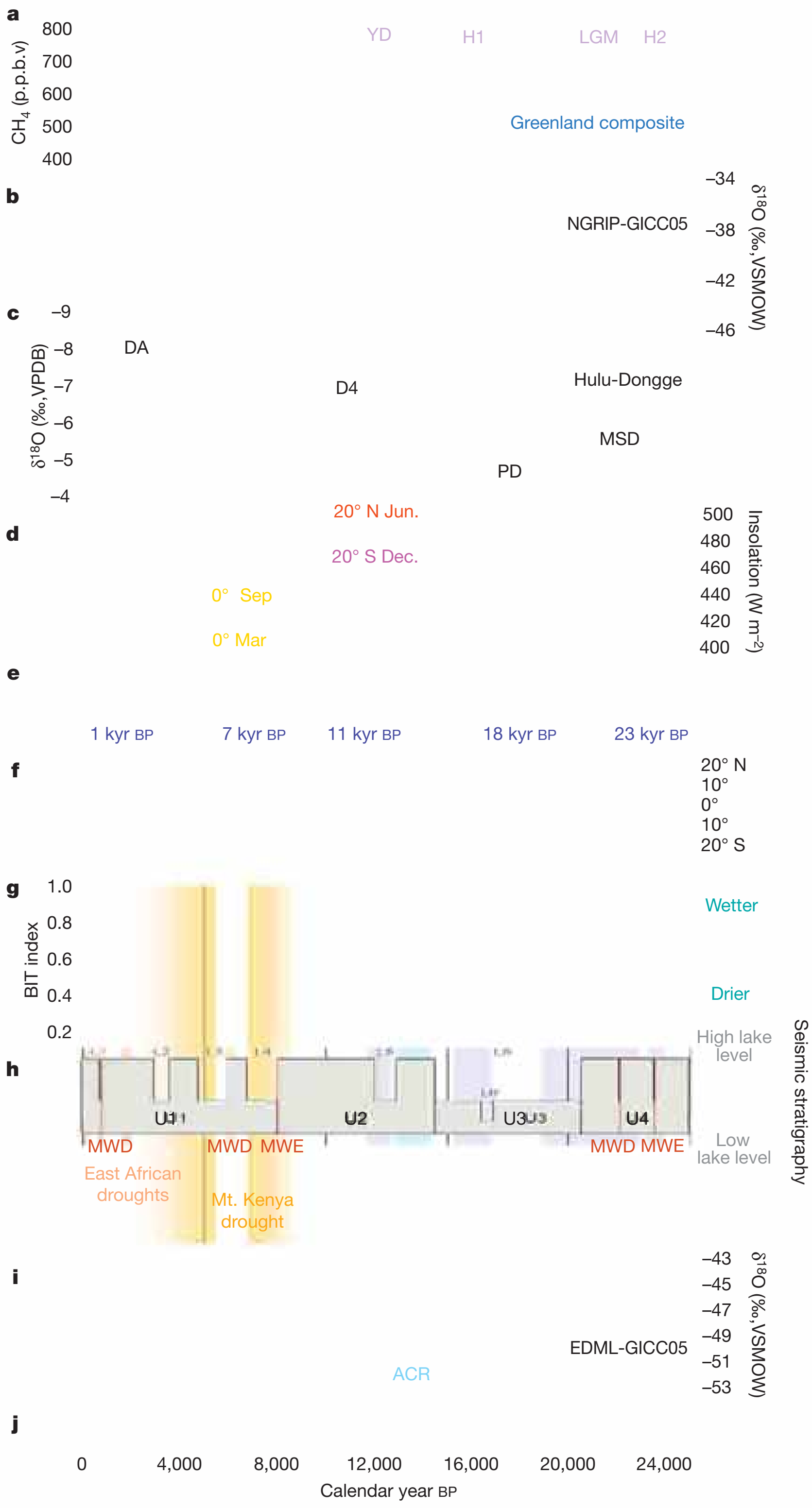


Figure 2 | Lake Challa climate-proxy data in context. Comparison of Lake Challa data (**g**, **h**) with selected other climate records (**a**–**c**, **i**), insolation (**d**) and schematic rainfall parameters (**e**, **f**). Relevant chronozones are shaded in purple and blue for northern and southern high-latitude influences respectively, and orange/pink for regional Holocene droughts. YD, Younger Dryas; ACR, Antarctic Cold Reversal; H1 and H2, Heinrich events 1 and 2; LGM, Last Glacial Maximum. **a**, Greenland composite (NGRIP-GISP2-GRIP) atmospheric CH$_4$ record[28]. **b**, Greenland (NGRIP) ice core δ^{18}O on the GICC05 timescale (ref. 29). **c**, Hulu/Dongge cave stalagmite δ^{18}O (ref. 6 and references therein); D4, DA, PD and MSD are the individual stalagmite samples making up this composite record. **d**, June and December insolation at 20° N/S, and March and September insolation on the Equator[30]. **e**, Monthly insolation[30] and schematic representation of seasonal rainfall in the Lake Challa region (January through December) at 1, 7, 11, 18 and 23 kyr BP, illustrating the joint effects on total annual rainfall of (1) a strengthened monsoon associated with a peak summer/winter insolation contrast between the hemispheres (which increases rainfall during either the long or short rain season) and (2) variation in local March or September insolation (with modest maxima causing the ensuing rain season to wither). **f**, Schematic representation of the approximate latitudinal extent of ITCZ migration in the western Indian Ocean at 1, 7, 11, 18 and 23 kyr BP, under the combined influence of low-latitude insolation forcing and (during glacial time) the equatorward displacement of the southern polar front; the line thickness of upward and downward arrows is proportional to the relative intensity of the southeasterly and northeasterly monsoons, respectively. **g**, Challa BIT-index rainfall reconstruction, individual data with three-point moving average. **h**, Challa lake-level reconstruction based on seismic-reflection stratigraphy (Supplementary Methods 1, 3). U1–U4, seismic units; L1–L6 and L6′, lowstand episodes; red marker horizons, basin-wide mass-wasting events (MWEs) and more local mass-wasting deposits (MWDs). **i**, Antarctica EDML ice core δ^{18}O record on the NGRIP-GICC05 timescale[28]. **j**, The Challa age–depth model (Supplementary Fig. 3) is derived from AMS ^{14}C dates on bulk organic carbon (black tick marks), corrected for an evolving lake-carbon reservoir age between 200 and 450 years determined at 19 depth intervals (red tick marks); the pink areas indicate sections of successful wiggle-match dating (Supplementary Methods 2).

Southern Hemisphere summer insolation (Fig. 2d). In equatorial East Africa, twice-yearly passing of ITCZ-associated rain belts (Fig. 1b) produces a bimodal seasonal rainfall pattern (Fig. 1a and Supplementary Discussion 2), and hence predicts a doubled frequency of precession-driven wet and dry periods as the southeasterly and northeasterly monsoons are enhanced in alternation. Clear manifestation of such 'double-frequency' (that is, ~11,500-year interval) wet–dry cycles in a palaeorecord may partly depend on the local prominence of dry seasons, as is the case in semi-arid southeastern Kenya (Supplementary Discussion 2). Over the past 25,000 years, equatorial East Africa experienced high rainfall when early-summer insolation peaked in either the northern or the southern tropics (that is, at ~22.5, 11.5 and 1.5 kyr BP; Fig. 2g, d), the summer/winter inter-hemispheric insolation gradient was large, and either the southeasterly or northeasterly monsoon was intensified[19]. Rainfall was reduced when the summer/winter inter-hemispheric insolation gradient was modest (at 17–18 and 6–7 kyr BP), and neither monsoon was particularly intensified (Supplementary Discussion 8). From an equatorial perspective, drought at those times may also have resulted because the coincident minimum in local insolation[20] during September (17–18 kyr BP) or March (6–7 kyr BP) caused the rain season which followed it to fail or wither (Fig. 2e and Supplementary Discussion 9).

Focusing on the Holocene, the Challa pattern of a short early-Holocene humid period followed first by mid-Holocene drought and then wetter late-Holocene conditions (Fig. 2g, h) has also been recorded at other African lakes located between the Equator and $\sim$8° S (Supplementary Discussion 7). Based on palaeovegetation data, Peyron *et al.*[21] located the transition area between a wetter (northern tropical) and drier (southern tropical) mid-Holocene climate regime between the Equator and 3° S, that is, just north of Lake Challa. Alternation of wet and dry climate at a half-precessional interval is also evident in the early part of the Challa BIT record, as local glacial drought was limited to a $\sim$4,000-year period between 20.5 and 16.5 cal. kyr BP (Fig. 2g). That in eastern (but not western) tropical Africa this 'Last Glacial Maximum (LGM) drought' was immediately preceded by a wet phase is well-documented[4]. Recently it has been recognized that in some regions of southeastern tropical Africa even the LGM was not particularly dry[22,23]. Importantly, we can establish that peak glacial drought in equatorial East Africa overlapped only partly with the LGM proper (Fig. 2g and Supplementary Discussion 10). For most of its duration it instead coincided with a modest summer/winter inter-hemispheric insolation contrast (Fig. 2d), as was the case during the mid-Holocene dry period.

Following mid-Holocene drought, rainfall in equatorial East Africa increased again from $\sim$6.8 cal. kyr BP (peaking between 4.5 and 2 cal. kyr BP), but was modulated by distinct multi-century droughts (Fig. 2g). The double-peaked wet period at $\sim$6.8–5.5 cal. kyr BP, which interrupts mid-Holocene drought in the BIT record and is represented in seismic stratigraphy by the inferred highstand separating L4 and L3 (Fig. 2h), matches the timing of a similarly double-peaked wet period interrupting mid-Holocene drought on Mt Kenya (ref. 24 and Supplementary Discussion 11). Further, the three most prominent dry episodes of the last two millennia recorded at Challa (2.0–1.6, 0.8–0.6 and 0.2–0.15 cal. kyr BP) match the timing of droughts known to have had broad regional impact in continental East Africa (ref. 25 and Supplementary Discussion 12). In summary, our Challa BIT data are also consistent with available regional palaeohydrological information at multi-decadal to century timescales. Notwithstanding our monitoring of the modern system (see Supplementary Methods 3), uncertainty remains regarding the exact process by which branched and isoprenoid tetraether lipids are transferred to and incorporated in Lake Challa sediments, and hence, about the precise relationship of the Challa BIT index to past rainfall variation in the Challa region at the multi-decadal resolution of our reconstruction. Our confidence in the truthfulness of the Challa BIT index as a hydrological proxy is based partly on its excellent agreement both with the (lower-resolution) Challa seismic-stratigraphic record of past lake-level fluctuations, and with diverse other high-quality palaeohydrological indicators from equatorial East Africa at a wide range of timescales (Figs 2g and Supplementary Fig. 7). The unique combination of adequate time coverage (25,000 years) and high resolution presented by the Challa BIT climate-proxy record, together with the long-term stability of the Challa sedimentary environment (Supplementary Discussion 2), are key to the climate-dynamical insights generated.

In regions bordering the western Indian Ocean, opposite Holocene moisture-balance trends were established in the northern and southern tropics, and a hybrid pattern near the Equator, from immediately after the Younger Dryas (Supplementary Discussion 7). The early end of the African humid period[11] near the Equator contrasts with its persistence until $\sim$6.0–5.0 cal. kyr BP in northeastern tropical Africa, related to Northern Hemisphere summer insolation control on Indian monsoon intensity[7]. During the LGM and the deglaciation period, northern ice-sheet influence and reduced Indian Ocean sea surface temperatures partly overrode low-latitude insolation influence on monsoon intensity and evaporation, creating more cross-equatorial uniformity in moisture-balance trends. Southeastern Africa enjoyed greater moisture than other tropical regions during the LGM proper, partly because of modest (1.5–2.0 °C) glacial cooling of the tropical Indian Ocean[26]. Also, the southern mid-latitude cooling associated with expanded circum-Antarctic sea ice impeded southward movement of the ITCZ during Southern Hemisphere summer (Fig. 2f), thereby compressing the tropical convective zone to a narrower latitudinal band[4]. These processes worked together with the orbitally enhanced northeasterly monsoon to break LGM-era drought in southeastern tropical Africa. They may also explain the high rainfall in equatorial East Africa immediately before and halfway into the LGM (Fig. 2g). Our equatorial data thus highlight the notion that orbital-scale ITCZ 'migration', or shifts in its mean annual latitudinal position, mainly relates to variation in how far high-latitude influences on lower mid-latitude climate allowed the ITCZ to move into the northern or southern hemisphere during summer (Fig. 2f); the cross-equatorial position of the region with twice-annual ITCZ passage, and hence bimodal rainfall regime, was probably not much affected (Supplementary Discussion 13).

The distinct timing of late-glacial drought on the Equator (20.5–16.5 cal. kyr BP) does not fit well with Northern Hemisphere glaciation being its primary driver. It is best attributed to partial failure of the short rain season due to modest local September insolation, perhaps exacerbated by weakening of the southeasterly monsoon due to northern cooling modestly affecting the North African land mass. Following establishment of interglacial temperatures (by 11.5 cal. kyr BP; ref. 16), low-latitude insolation clearly became the dominant driver of regional moisture balance, with strong north–south anti-phasing across the Equator revealing the signature of orbital precession[27].

On balance, low-latitude orbital forcing has exerted a dominant influence on climate in the western Indian Ocean throughout the past 25,000 years, including the LGM and deglaciation periods. High-latitude influences challenged this primacy only during the Younger Dryas, Heinrich events (to lesser extent), and perhaps also during some of the multi-century droughts that punctuated the relatively moist late Holocene (Fig. 2g and Supplementary Discussion 12). Important keys to the near-continuous prominence of low-latitude climate processes in this monsoon region are the low-latitude position of its continental regions of surface air flow convergence (in contrast to those of the Indian and East Asian monsoons) and its relative isolation from the Atlantic Ocean domain, where important latitudinal heat redistribution by the meridional overturning circulation more tightly coupled tropical climate history to high-latitude boundary conditions.

METHODS SUMMARY

A dense grid of echo-sounding and very-high-resolution seismic reflection profiles was acquired in 2003 using a 3.5 kHz Geopulse system. In 2003 and 2005 we recovered from a mid-lake location (Supplementary Fig. 1) gravity cores with intact sediment–water interface, a mini-Kullenberg piston core of 2.6 m length and three parallel UWITEC hammer-driven piston cores of 22 m. After cross-correlation of successive overlapping core sections, and excision of five turbidites (42, 36, 5, 6 and 14 cm thick) that consist largely of re-deposited crater slope material, the recovered sediment profile constituted a continuous, 20.82-m-long composite sequence of mostly finely laminated organic muds rich in diatom silica. Seismic-stratigraphic signatures of past lake-level fluctuation and mass-wasting events were linked to the lithology of the cored sediment sequence at the level of turbidites with clear seismic signature, and through patterns in sediment water content. The cored sediment sequence was dated using AMS ^{14}C dates on bulk organic carbon, corrected for an evolving lake-carbon reservoir age through combination with ^{210}Pb-dated sub-recent sediments, AMS ^{14}C dates on grass charcoal, and wiggle-match dating (Supplementary Methods 2).

For organic geochemistry, powdered and freeze-dried sediments (1–3 g dry mass) were extracted by Dionex accelerated solvent extraction with 9:1 dichloromethane/methanol by volume. The extracts were separated by Al_2O_3 column chromatography with 9:1 hexane/dichloromethane and 1:1 dichloromethane/methanol. The polar fraction was condensed by rotary evaporation, dissolved in 99:1 hexane/isopropanol, and filtered using a PTFE 0.4 μm filter before analysis of tetraether lipids by high-performance liquid chromatography/atmospheric-pressure chemical ionization mass spectrometry (HPLC/APCI-MS). Glycerol dialkyl glycerol tetraether lipids were detected by single ion monitoring of their $[M+H]^+$ ions (dwell time 234 ms), quantified by integration of peak areas, and used to calculate the BIT index (Supplementary Methods 3). All BIT analyses were run in duplicate; average error is <0.01 unit.

Received 10 January; accepted 15 September 2009.

1. Clement, A. C., Hall, A. & Broccoli, A. J. The importance of precessional signals in the tropical climate. *Clim. Dyn.* **22,** 327–341 (2004).
2. Partin, J. W., Cobb, K. M., Adkins, J. F., Clark, B. & Fernandez, D. P. Millennial-scale trends in west Pacific warm pool hydrology since the Last Glacial Maximum. *Nature* **449,** 452–455 (2007).
3. Cruz, F. W. *et al.* Orbitally driven east-west antiphasing of South American precipitation. *Nature Geosci.* **2,** 210–214 (2009).
4. Gasse, F., Chalié, F., Vincens, A., Williams, M. A. J. & Williamson, D. Climatic patterns in equatorial and southern Africa from 30,000 to 10,000 years ago reconstructed from terrestrial and near-shore proxy data. *Quat. Sci. Rev.* **27,** 2316–2340 (2008).
5. Trauth, M. H., Maslin, M. A., Deino, A. & Strecker, M. R. Late Cenozoic moisture history of East Africa. *Science* **309,** 2051–2053 (2005).
6. Wang, Y. J. *et al.* Millennial- and orbital-scale changes in the East Asian monsoon over the past 224,000 years. *Nature* **451,** 1090–1093 (2008).
7. Fleitmann, D. *et al.* Holocene ITCZ and Indian monsoon dynamics recorded in stalagmites from Oman and Yemen (Socotra). *Quat. Sci. Rev.* **26,** 170–188 (2007).
8. Vellinga, M. & Wood, R. A. Global climatic impacts of a collapse of the Atlantic thermohaline circulation. *Clim. Change* **54,** 251–267 (2002).
9. McManus, J. F., Francois, R., Gherardi, J. M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428,** 834–837 (2004).
10. Zhang, R. & Delworth, T. L. Simulated tropical response to a substantial weakening of the Atlantic thermohaline circulation. *J. Clim.* **18,** 1853–1860 (2005).
11. deMenocal, P. *et al.* Abrupt onset and termination of the African Humid Period: rapid climate responses to gradual insolation forcing. *Quat. Sci. Rev.* **19,** 347–361 (2000).
12. Kutzbach, J. E. Monsoon climate of the early Holocene: climate experiment with Earth's orbital parameters for 9000 years ago. *Science* **214,** 59–61 (1981).
13. Peterson, L. C., Haug, G. H., Hughen, K. A. & Röhl, U. Rapid changes in the hydrological cycle of the tropical Atlantic during the last glacial. *Science* **290,** 1947–1950 (2000).
14. Talbot, M. R., Filippi, M. L., Jensen, N. B. & Tiercelin, J.-J. An abrupt change in the African monsoon at the end of the Younger Dryas. *Geochem. Geophys. Geosyst.* **8,** doi:10.1029/2006GC001465 (2007).
15. Thompson, L. G. *et al.* Kilimanjaro ice core records: evidence of Holocene climate change in tropical Africa. *Science* **298,** 589–593 (2002).
16. Tierney, J. E. *et al.* Northern Hemisphere controls on tropical Southeast African climate during the last 60,000 years. *Science* **322,** 252–255 (2008).
17. Cruz, F. W. *et al.* Insolation-driven changes in atmospheric circulation over the past 116,000 years in sub-tropical Brazil. *Nature* **434,** 63–66 (2005).
18. Haug, G. H., Hughen, K. A., Sigman, D. M., Peterson, L. C. & Rohl, U. Southward migration of the intertropical convergence zone through the Holocene. *Science* **293,** 1304–1308 (2001).
19. Thevenon, F., Williamson, D. & Taieb, M. A 22-kyr BP sedimentological record of Lake Rukwa (8° S, SW Tanzania): environmental, chronostratigraphic and climatic implications. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **187,** 285–294 (2002).
20. Berger, A., Loutre, M. F. & Mélice, J. L. Equatorial insolation: from precession harmonics to eccentricity frequencies. *Clim. Past* **2,** 131–136 (2006).
21. Peyron, O., Jolly, D., Bonnefille, R., Vincens, A. & Guiot, J. Climate of East Africa 6000 ^{14}C yr BP as inferred from pollen data. *Quat. Res.* **54,** 90–101 (2000).
22. Garcin, Y., Vincens, A., Williamson, D., Guiot, J. & Buchet, G. Wet phases in tropical southern Africa during the last glacial period. *Geophys. Res. Lett.* **33,** L07703, doi:10.1029/2005GL025531 (2006).
23. Mumbi, C. T., Marchant, R., Hooghiemstra, H. & Wooller, M. J. Late Quaternary vegetation reconstruction from the Eastern Arc Mountains, Tanzania. *Quat. Res.* **69,** 326–341 (2008).
24. Barker, P. A. *et al.* A 14,000-year oxygen isotope record from diatom silica in two alpine lakes on Mt. Kenya. *Science* **292,** 2307–2310 (2001).
25. Verschuren, D. & Charman, D. J. in *Natural Climate Variability and Global Warming* (eds Battarbee, R. W. & Binney, H. E.) 189–231 (Wiley-Blackwell, 2008).
26. Pinot, S. *et al.* Tropical paleoclimates at the Last Glacial Maximum: comparison of Paleoclimate Modeling Intercomparison Project (PMIP) simulations and paleodata. *Clim. Dyn.* **15,** 857–874 (1999).
27. Ruddiman, W. F. What is the timing of orbital-scale monsoon changes? *Quat. Sci. Rev.* **25,** 657–658 (2006).
28. EPICA community members.. One-to-one coupling of glacial climate variability in Greenland and Antarctica. *Nature* **444,** 195–198 (2006).
29. Rasmussen, S. O. *et al.* A new Greenland ice core chronology for the last glacial termination. *J. Geophys. Res.* **111,** doi:10.1029/2005JD006079 (2006).
30. Berger, A. & Loutre, M.-F. Insolation values for the climate of the last 10 million years. *Quat. Sci. Rev.* **10,** 297–317 (1991).

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

Acknowledgements This research was sponsored by funding agencies in Belgium (FWO-Vlaanderen), Denmark (Danish National Research Council), Germany (DFG) and The Netherlands (NWO) through the European Science Foundation (ESF) EuroCORES programme, EuroCLIMATE. Fieldwork was conducted with permission of the Kenyan Ministry of Education, Science and Technology (MOEST) and the Tanzania Commission for Science and Technology (COSTECH). We thank A. Alcantara, M. Kohler, S. Lauterbach, R. Niederreiter and S. Opitz for assistance with core collection and processing; D. Engstrom for ^{210}Pb dating, A. Berger, D. Fleitmann, J. Kutzbach and Z. Liu for discussion, and E. Odada for support. J.M. acknowledges the support of the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen).

Author Contributions D.V. initiated and coordinated the research, and wrote the paper. J.S.S.D. was responsible for the production of the biomarker data, J.M. interpreted the seismic-reflection data, I.K. constructed the composite sediment sequence and depth scale, M.B. developed the sediment age model, and M.F. supervised core sampling and managed the project database. G.H.H. and all CHALLACEA members contributed their expertise to improving the discussed data sets and/or 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.V. (dirk.verschuren@UGent.be).

CHALLACEA project members Bas van Geel[1], Marc De Batist[2], Philip Barker[3], Mathias Vuille[4], Daniel J. Conley[5], Daniel O. Olago[6], Isla Milne[7], Birgit Plessen[8], Hilde Eggermont[9], Christian Wolff[8], Elizabeth Hurrell[3], Jort Ossebaar[10], Anna Lyaruu[1], Johannes van der Plicht[11], Brian F. Cumming[7], Achim Brauer[8], Stephen M. Rucina[12], James M. Russell[13], Edward Keppens[14], Joseph Hus[15], Raymond S. Bradley[16], Melanie Leng[17], Jens Mingram[8] & Norbert R. Nowaczyk[8]

[1]Institute for Biodiversity and Ecosystem Dynamics, Research Group Paleoecology and Landscape Ecology, Faculty of Science, Universiteit van Amsterdam, 1098 SM Amsterdam, The Netherlands. [2]Renard Centre of Marine Geology, Department of Geology and Soil Science, Ghent University, Krijgslaan 281 S8, 9000 Gent, Belgium. [3]Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, UK. [4]Atmospheric and Environmental Sciences, State University of New York, Albany, New York 12222, USA. [5]GeoBiosphere Centre, Department of Geology, Lund University, SE-223 62 Lund, Sweden. [6]Department of Geology, University of Nairobi, PO Box 30197, Nairobi, Kenya. [7]Paleoecological Environmental Assessment and Research Laboratory, Department of Biology, Queen's University, Kingston, ON K7L 3N6, Canada. [8]GeoForschungsZentrum Potsdam, Sektion 3.3 Klimadynamik und Sedimente, Telegrafenberg, D-14473 Potsdam, Germany. [9]Limnology Unit, Department of Biology, Ghent University, Ledeganckstraat 35, 9000 Gent, Belgium. [10]Faculty of Geosciences, Utrecht University, PO Box 80021, 3508 TA Utrecht, The Netherlands. [11]Center for Isotope Research, Groningen University, 9747 AG Groningen, The Netherlands. [12]Department of Palynology, National Museums of Kenya, Museum Hill, Nairobi, Kenya. [13]Geological Sciences, Brown University, Providence, Rhode Island 02912, USA. [14]Department of Geology, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. [15]Centre de Physique du Globe, Royal Meteorological Institute, 5670 Dourbes, Belgium. [16]Climate System Research Center, Department of Geosciences, University of Massachusetts, Amherst, Massachusetts 01003-9297, USA. [17]NERC Isotope Geosciences Laboratory, British Geological Survey, Keyworth, Nottingham NG12 5GG, UK.

nature
Vol 462|3 December 2009|**doi:10.1038/nature08553**

LETTERS

Common dependence on stress for the two fundamental laws of statistical seismology

Clément Narteau[1], Svetlana Byrdina[1,2], Peter Shebalin[1,3] & Danijel Schorlemmer[4]

Two of the long-standing relationships of statistical seismology are power laws: the Gutenberg–Richter relation[1] describing the earthquake frequency–magnitude distribution, and the Omori–Utsu law[2] characterizing the temporal decay of aftershock rate following a main shock. Recently, the effect of stress on the slope (the *b* value) of the earthquake frequency–magnitude distribution was determined[3] by investigations of the faulting-style dependence of the *b* value. In a similar manner, we study here aftershock sequences according to the faulting style of their main shocks. We show that the time delay before the onset of the power-law aftershock decay rate (the *c* value) is on average shorter for thrust main shocks than for normal fault earthquakes, taking intermediate values for strike-slip events. These similar dependences on the faulting style indicate that both of the fundamental power laws are governed by the state of stress. Focal mechanisms are known for only 2 per cent of aftershocks. Therefore, *c* and *b* values are independent estimates and can be used as new tools to infer the stress field, which remains difficult to measure directly.

The aftershock rate, Λ, is described by the Omori–Utsu law, $\Lambda(t) = k(t+c)^{-p}$, where t is the time from the main shock, k the productivity of the aftershock sequence, p the power-law exponent, and c the time delay before the onset of the power-law aftershock decay rate. The c value is strongly debated[4–10], as researchers[11,12] claim that c is close to zero and larger values are, at least partially, artefacts caused by lower data completeness after a main shock, especially during the main-shock coda. Investigations after the 2004 magnitude (M) 6.0 Parkfield, California, earthquake[13] and the 2004 $M = 6.6$ mid-Niigata Prefecture, Japan, earthquake[8] show that carefully identifying micro-earthquakes directly after main shocks give smaller but still non-zero estimates of the c value. We show for southern California and Japan that limiting the analysis to smaller main-shock and larger aftershock magnitudes results in non-zero c values that characterize a systematic behaviour of early aftershocks rather than a completeness artefact.

For southern California, we use an earthquake catalogue containing relocated events with focal mechanisms[14] from 1984 to 2003 and the Southern California Seismic Network (SCSN) catalogue since 2003. We select only high-quality events with a solution misfit <0.2 and a station distribution ratio ≥ 0.5. This subcatalogue contains more than 12,000 events. To identify main shocks, we adopt a time- and space-window algorithm[15,16] based on power-law relationships that reflect the hierarchical nature of the earthquake phenomenon (see Methods). The corresponding aftershocks with no focal mechanisms are extracted from the entire SCSN catalogue according to the same rules ($>$385,000 events between 1984 and 2007).

As for southern California, we use two different catalogues for Japan. We combine the F-Net catalogue with the JMA catalogue to isolate the focal mechanisms of main shocks since 1997 (F-Net) and their respective aftershocks (JMA). To eliminate swarms of seismicity in the different volcanic areas of Japan, we also eliminate all aftershock sequences for which the geometric average of the times from their main shocks is larger than 4 hours. Thus, we remove spatial clusters of seismicity for which no clear aftershock decay rate exist.

In order to avoid artefacts arising from overlapping records, we focus on aftershock sequences produced by intermediate-magnitude main shocks, disregarding the large ones. For the same reason, we consider only the larger aftershocks and we stack them according to the main-shock time to compensate for the small number of events in each sequence. Practically, two ranges of magnitude have to be defined for main shocks and aftershocks, $\left[M^{\mathrm{M}}_{\mathrm{min}}, M^{\mathrm{M}}_{\mathrm{max}}\right]$ and $\left[M^{\mathrm{A}}_{\mathrm{min}}, M^{\mathrm{A}}_{\mathrm{max}}\right]$, respectively. We set the minimum magnitude thresholds for main shocks to $M^{\mathrm{M}}_{\mathrm{min}} = 2.5$ as this is the overall completeness level of the relocated catalogue[3]. Then, we determine that $M^{\mathrm{M}}_{\mathrm{max}} = 4.5$ is the main-shock magnitude from which overlapping records start to systematically influence the earthquake frequency–magnitude distribution of $M \geq 1.8$ aftershocks in the southern Californian catalogue (see Methods). The magnitude range of aftershocks, $\left[M^{\mathrm{A}}_{\mathrm{min}}, M^{\mathrm{A}}_{\mathrm{max}}\right]$, is a free parameter that may be varied to produce different estimates of the c value. In all cases, we select only aftershock sequences for which the local magnitude of completeness is smaller than $M^{\mathrm{A}}_{\mathrm{min}}$, the minimum magnitude threshold for aftershocks[17]. For each main shock, this completeness magnitude is calculated within a circle of 15 km radius over the three preceding years using a goodness-of-fit method[18]. Hence, the $M^{\mathrm{A}}_{\mathrm{min}}$ value will not only determine the number of events that we consider in individual sequences but also the number of sequences that are taken into account (for example, taking $M^{\mathrm{A}}_{\mathrm{min}} = 2.4$ in southern California results in approximately 2,800 main shocks and more than 4,000 aftershocks during the first 12 hours).

The faulting style of main shocks is classified using rake angles, λ, in the Aki convention: $\lambda \approx 0^\circ$ or $\lambda \approx \pm 180^\circ$ for strike-slip events, $\lambda \approx 90^\circ$ for thrust events, and $\lambda \approx -90^\circ$ for normal events[3]. Here, we use both rake angles and a sliding window with a width of 60° and a step size of 5°. This results in 73 classes of main shocks. For each class and different intervals of magnitude, we stack aftershocks according to the time from their respective main shocks and we evaluate the c value by a maximum-likelihood method using continuous minimization by simulated annealing[4]. In addition, we evaluate the uncertainty of the maximum-likelihood estimates of the c value using a Monte Carlo approach and verify that all the observed behaviours are the same for the first and the second rake when they are treated independently. We also systematically perform a series of computations with different starting times of the optimization procedure to test the stability of the c value, and therefore eliminate potential artefacts related to catalogue incompleteness (see Methods).

For southern California, Fig. 1a, b shows the evolution of the c value with respect to the rake angle for aftershocks of magnitudes

---

[1]Institut de Physique du Globe de Paris (UMR 7154, CNRS, Univ. P7), 4 Place Jussieu, 75252 Paris Cedex 05, France. [2]Laboratoire de Géophysique Interne et Tectonophysique, UMR 5559, CNRS, Université de Savoie, IRD, 73376 Le Bourget-du-Lac Cedex, France. [3]International Institute of Earthquake Prediction Theory and Mathematical Geophysics, 84/32 Profsouznaya, Moscow 117997, Russia. [4]Department of Earth Sciences, University of Southern California, 3651 Trousdale Parkway, Los Angeles, California 90089, USA.

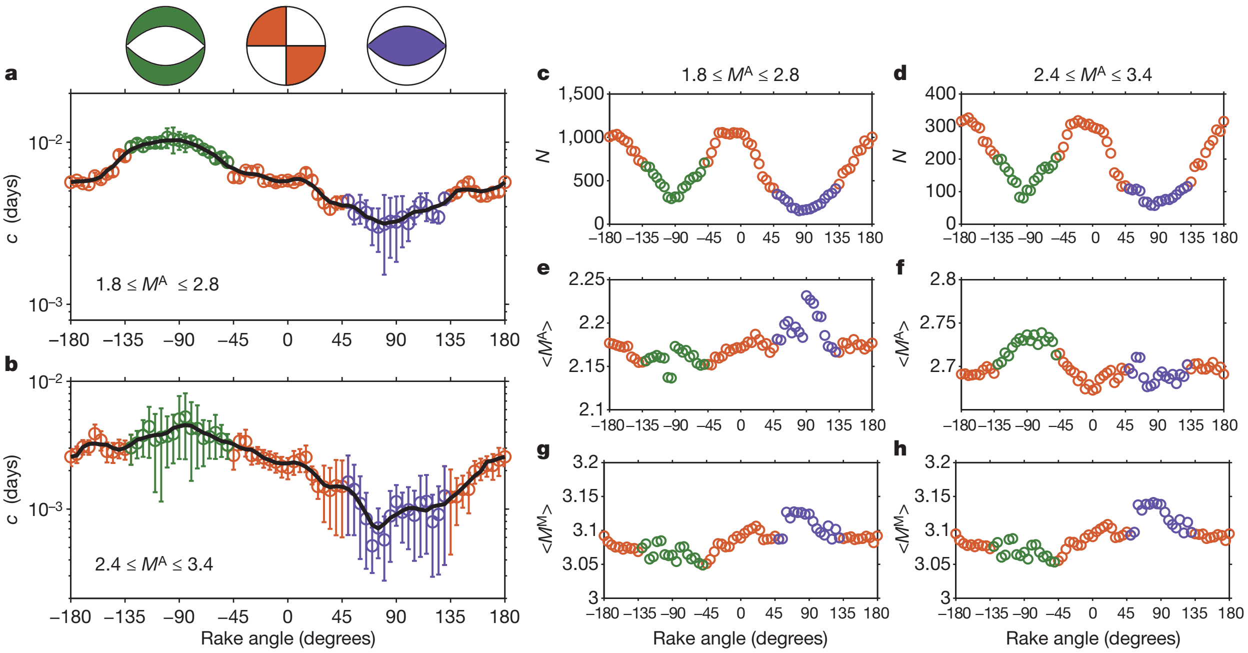


Figure 1 | Influence of faulting style on the time delay before the onset of the power-law aftershock decay rate in southern California. **a**, **b**, Logarithm of the *c* value with respect to rake angles for aftershock magnitude ranges of [1.8, 2.8] and [2.4, 3.4], respectively. Red, green and blue distinguish rake intervals of mainly strike-slip, normal and thrust events, respectively. Error bars are estimated by a Monte Carlo method and correspond to 16% and 84% quantiles of the maximum-likelihood estimates of *c* for 500 synthetic aftershock sequences at each point. **c**–**h**, Statistics of main shocks and aftershocks for both ranges of aftershocks. **c**, **d**, The number, *N*, of selected aftershocks. **e**, **f**, The average magnitude of selected aftershocks. **g**, **h**, The average magnitude of selected main shocks.

from 1.8 to 2.8 and from 2.4 to 3.4, respectively. For these intervals, Fig. 1c–h shows the numbers of selected aftershocks and the average magnitudes of selected main shocks and aftershocks for each rake angle. The uncertainty on the *c* value is strongly dependent on the number of events in the stack (Fig. 1c, d) but, even for high magnitude thresholds, a signal emerges. The *c* value for strike-slip earthquakes is higher than the *c* value for thrust earthquakes but lower than the *c* value for normal events. This means that the time delay before the onset of the power-law aftershock decay rate is shorter for thrust events, longer for normal ones and takes intermediate values for strike-slip events. The same behaviour is observed for the two magnitude ranges, despite a general shift in values that can be explained by a dependence of the *c* value on the average magnitude of aftershocks (see Methods). The differences of *c* values between thrust, strike-slip, and normal earthquakes cannot be explained by variations of the average magnitude of aftershocks (that is, the corresponding *b* value) because it does not exhibit a systematic pattern when plotted with respect to the rake angle (Fig. 1e, f). However, the average magnitude of main shocks reveals a similar behaviour for both magnitude ranges (Fig. 1g, h). Main shocks are larger for thrust earthquakes and smaller for normal earthquakes when compared to strike-slip events. This is a common feature in southern California that can be explained by differences in the earthquake frequency–magnitude distribution.

For Japan, to verify that missing aftershocks do not affect the estimation of *c* values, we also compare the completeness of the JMA catalogue with respect to the Peng catalogue[7] for 77 main shocks with magnitudes between 3.0 and 4.5 (see Supplementary Information). Then, as in southern California, we also found the same dependence of the *c* value on the rake angle for the same aftershock and main-shock magnitude ranges (Fig. 2a, b). Because the F-Net catalogue contains only few $M < 3.5$ events[3], the average magnitude of selected main shocks is higher than in southern California. As a result, the estimated *c* values are overall smaller.

In almost all cases, earthquake frequency–magnitude distributions follow a power-law relationship characterized by the *b* value, the exponent of the Gutenberg–Richter frequency–magnitude relation.

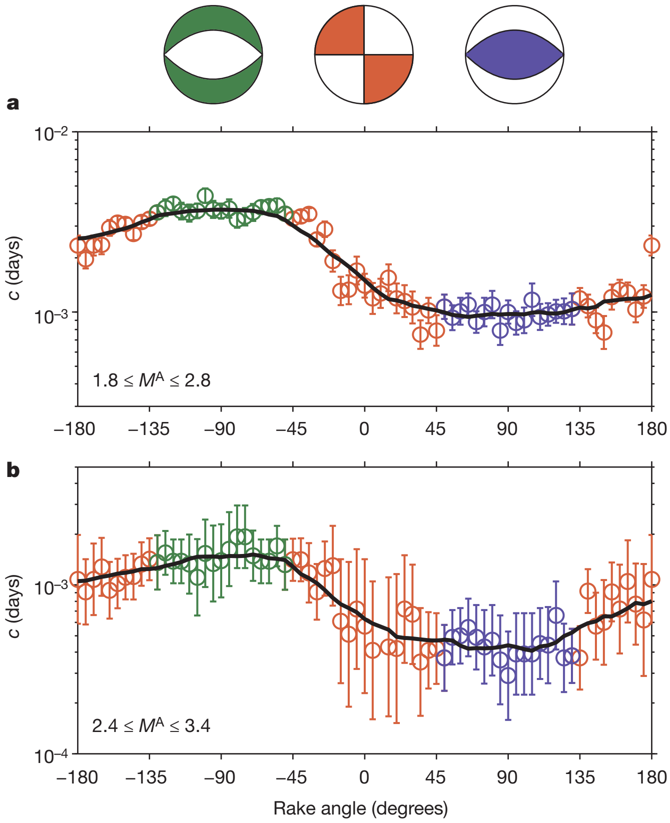


Figure 2 | Influence of faulting style on the time delay before the onset of the power-law aftershock decay rate in Japan. **a**, **b**, Logarithm of the *c* value with respect to rake angles of [1.8, 2.8] and [2.4, 3.4], respectively. From the comparison between the Peng[7] and the JMA catalogues, the optimization procedure starts 5×10^{-4} days after main shocks for $1.8 \le M^A \le 2.8$ aftershocks. Error bars are estimated by a Monte Carlo method and correspond to 16% and 84% quantiles of the maximum-likelihood estimates of *c* for 500 synthetic aftershock sequences at each point.

For the same catalogues that have been used to identify main shocks in Japan and southern California, the *b* value varies systematically for different styles of faulting. As shown in ref. 3, the *b* value for strike-slip events ($\sim 0.9 \pm 0.05$) is indeed higher than for thrust events ($\sim 0.7 \pm 0.05$) and lower than for normal events ($\sim 1.1 \pm 0.05$). Combining both observations, the *c* value and the *b* value exhibit a common dependence on focal mechanism that can be interpreted in terms of differential shear stress as follows.

The Mohr–Coulomb theory proposes that shear failure in a homogeneous medium subject to a uniform triaxial stress (σ_1, σ_2, σ_3) should occur on optimally oriented planes. These planes are parallel to the intermediate shear stress, σ_2, and form an angle Θ with σ_1, the maximum compressive stress. Then, the condition for faulting is governed by Amonton's law:

$$\tau \propto \mu \sigma_n$$

where μ is a coefficient of static friction, and σ_n and τ are the normal stress and the shear stress resolved on the fault surface. Assuming that one of the principal stresses is vertical and the other two lie in a horizontal plane, it is possible to express the angle Θ and the differential shear stress ($\sigma_1 - \sigma_3$) at which faulting may occur with respect to μ (ref. 19). In this context, the orientation of the principal shear stresses not only determines the faulting mechanism but also the amplitude of the differential shear stress. For example, the vertical stress corresponds to the maximum compressive stress for normal faulting and to the minimum compressive stress for thrust faulting. As a consequence, for the same coefficient of static friction, the differential shear stress should be much higher for thrust faulting than for normal faulting. Between these compressional and extensional regimes, strike-slip faulting should take an intermediate value of ($\sigma_1 - \sigma_3$). Hence, the relationship between the *c* value and the rake angle suggests that the time delay before the onset of the aftershock decay rate is positively correlated to the differential shear stress.

Considering that aftershocks result from a steplike perturbation of stress in the neighbourhood of a triggering event, various models for aftershock production predict that the amplitude of this perturbation controls the duration of the non-power-law regime of the aftershock decay rate over short time. See, for example, how an increasing stress step accelerates the onset of the power-law decay rate in rate-and-state friction models[20], in static fatigue models[4,21] and in damage mechanics models[22,23]. In addition, classical crack-growth experiments have shown that the amplitude of the stress perturbation is an increasing function of the remote applied stress[24]. The behaviour of the *c* value exposed in Figs 1a, b and 2a, b is an illustration of such relationships between the differential stress, the intensity of the stress redistribution and the shortness of the preliminary phase of an aftershock sequence. Nevertheless, there remains a high scatter in an individual aftershock sequence, most probably due to stress heterogeneity effects as well as measuring uncertainty. For this reason, the dependence of the *c* value on the differential shear stress can only be made through stacks of aftershocks.

The *b* value determines the ratio between the number of large and small earthquakes, and its variations can be associated with the ability of an earthquake rupture to propagate (lower *b* value) or not (higher *b* value) once nucleated. In this case, the positive correlation between the evolution of the *c* value and the *b* value across different stress regimes suggests a similar dependence on stress during earthquake propagation and aftershock production. Where the differential shear stress required to start sliding is higher, the porosity is reduced, stress interactions between the different parts of the fault zone are enhanced, and cracks can propagate faster in a wide range of directions[25,26]. Then, during the propagation of an earthquake, the characteristic time scales for failure, slip, or damage are shorter within process zones under higher confining stress. As a consequence, thrust earthquakes have a higher probability to propagate further than normal ones, consistent with the observation that the *b* value may be lower in a compressional regime than in an extensional one. If we assume that rupture terminates owing to structural and compositional heterogeneities, the same reasoning is again valid, and it takes a shorter time to nucleate aftershocks in higher-stress environments. As a result, the steady-state relaxation regime can be reached more rapidly and the *c* value is smaller for thrust earthquakes than for normal ones.

A common dependence on stress of earthquake frequency–magnitude distributions and aftershock decay rates offers new perspectives on earthquake-hazard analysis and earthquake physics, as it suggests that two independent aspects of seismicity can be used to constrain the state of stress in the brittle seismogenic crust. For example, the background seismicity and the quasi-systematic occurrence of aftershocks might be able to help in quantifying not only stress accumulation along major faults[27] but also the characteristic size of a potential earthquake[28].

METHODS SUMMARY

To identify main shocks, we deselect earthquakes of magnitude smaller than M which are within a $0.02 \times 10^{0.50M}$ kilometre radius circle during the first $0.125 \times 10^{0.55M}$ days after a magnitude M event[15,16]. Using the same spatial scaling, earthquakes that precede larger events by less than 12 hours are classified as potential foreshocks and removed from the catalogue of main shocks. This results in no overlapping aftershock sequences for the first 12 hours. For the selected main shocks, corresponding aftershocks with no rake angle are extracted from the entire SCSN catalogue (>385,000 events between 1984 and 2007) and the JMA catalogue (>1,100,000 events between 1997 and 2007). Practically, we select the smaller magnitude earthquakes occurring within 12 hours after a magnitude M main shock in a $0.02 \times 10^{0.50M}$ kilometre radius circle. Finally, only 2% of the selected aftershocks are also present in the catalogues of seismicity with focal mechanisms. Thus, data sets which are analysed here are different from the catalogues of seismicity that have been used to determine the relationship between the *b* value and the rake angle[3].

We investigate only earthquakes from particular magnitude ranges that can be shown to be complete even in early times of the aftershock sequence. The two major constraints are the following: (1) main shocks should be sufficiently small to have shorter coda waves such that subsequent aftershocks can be reliably detected; (2) aftershocks should be sufficiently big to ensure completeness at that particular magnitude level from the early times of an aftershock sequence. These magnitude ranges are determined from seismological and statistical constraints (see Methods).

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

Received 16 January; accepted 28 September 2009.

1. Gutenberg, B. & Richter, C. F. Frequency of earthquakes in California. *Bull. Seismol. Soc. Am.* **34,** 185–188 (1944).
2. Utsu, T. Aftershocks and earthquake statistics. *J. Fac. Sci. Hokkaido Univ. Ser. VII* **3,** 379–441 (1965).
3. Schorlemmer, D., Wiemer, S. & Wyss, M. Variations in earthquake-size distribution across different stress regimes. *Nature* **437,** 539–542 (2005).
4. Narteau, C., Shebalin, P. & Holschneider, M. Temporal limits of the power law aftershock decay rate. *J. Geophys. Res.* **107,** B2359, doi:10.1029/2002JB001868 (2002).
5. Shcherbakov, R., Turcotte, D. L. & Rundle, J. B. A generalized Omori's law for earthquake aftershock decay. *Geophys. Res. Lett.* **31,** L11613, doi:10.1029/2004GL019808 (2004).
6. Vidale, J. E., Peng, Z. & Ishii, M. Anomalous aftershock decay rates in the first hundred seconds revealed from the Hi-net borehole data. *Eos Trans. AGU* **85** (Fall Meet. Suppl.), abstr. S23C-07 (2004).
7. Peng, Z. G., Vidale, J. E., Ishii, M. & Helmstetter, A. Seismicity rate immediately before and after main shock rupture from high frequency waveforms in Japan. *J. Geophys. Res.* **112,** B03306, doi:10.1029/2006JB004386 (2007).
8. Enescu, B., Mori, J. & Miyasawa, M. Quantifying early aftershock activity of the 2004 mid-Niigata Prefecture earthquake (M_w6.6). *J. Geophys. Res.* **112,** B04310, doi:10.1029/2006JB004629 (2007).
9. Nanjo, K. Z. *et al.* Decay of aftershock activity for Japanese earthquakes. *J. Geophys. Res.* **112,** B08309, doi:10.1029/2006JB004754 (2007).
10. Kilb, D., Martynov, V. & Vernon, F. Aftershock detection as a function of time: results from the ANZA seismic network following the 31 October 2001 M_L 5.1 Anza, California, earthquake. *Bull. Seismol. Soc. Am.* **97,** 780–792 (2007).
11. Kagan, Y. Y. Short-term properties of earthquake catalogs and models of earthquake source. *Bull. Seismol. Soc. Am.* **94,** 1207–1228 (2004).

12. Lolli, B. & Gasperini, P. Comparing different models of aftershock rate decay: the role of catalog incompleteness in the first times after main shock. *Tectonophysics* **423,** 43–59 (2006).
13. Peng, Z. G., Vidale, J. E. & Houston, H. Anomalous early aftershock decay rate of the 2004 M_w 6.0 Parkfield, California, earthquake. *Geophys. Res. Lett.* **33,** doi:10.1029/2006GL026744 (2006).
14. Hauksson, E. Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in southern California. *J. Geophys. Res.* **105,** 13875–13903 (2000).
15. Gardner, J. & Knopoff, L. Is the sequence of earthquakes in southern California with aftershocks removed Poissonian? *Bull. Seismol. Soc. Am.* **5,** 1363–1367 (1974).
16. Reasenberg, P. Second-order moment of central California seismicity, 1969–1982. *J. Geophys. Res.* **90,** 5479–5495 (1985).
17. Schorlemmer, D. & Woessner, J. Probability of detecting an earthquake. *Bull. Seismol. Soc. Am.* **98,** 2103–2117 (2008).
18. Wiemer, S. & Wyss, M. Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the western United States, and Japan. *Bull. Seismol. Soc. Am.* **90,** 859–869 (2000).
19. Sibson, R. H. Frictional constraints on thrusts, wrench and normal faults. *Nature* **249,** 542–544 (1974).
20. Dieterich, J. A constitutive law for rate of earthquake production and its application to earthquake clustering. *J. Geophys. Res.* **99,** 2601–2618 (1994).
21. Scholz, C. Microfractures, aftershocks, and seismicity. *Bull. Seismol. Soc. Am.* **58,** 1117–1130 (1968).
22. Shcherbakov, R. & Turcotte, D. L. A damage mechanics model for aftershocks. *Pure Appl. Geophys.* **161,** 2379–2391 (2004).
23. Ben-Zion, Y. & Lyakhovsky, V. Analysis of aftershocks in a lithospheric model with seismogenic zone governed by damage rheology. *Geophys. J. Int.* **165,** 197–210 (2006).
24. Atkinson, B. K. Subcritical crack growth in geological materials. *J. Geophys. Res.* **89,** 4077–4114 (1984).
25. Amitrano, D. Brittle-ductile transition and associated seismicity: experimental and numerical studies and relationship with the *b*-value. *J. Geophys. Res. B* **108,** 2044, doi:10.1029/2001JB000680 (2003).
26. O'Connell, D. R. H., Ma, S. & Archuleta, R. J. Influence of dip and velocity heterogeneity on reverse- and normal-faulting rupture dynamics and near-fault ground motions. *Bull. Seismol. Soc. Am.* **97,** 1970–1989 (2007).
27. Narteau, C., Shebalin, P. & Holschneider, M. Loading rates in California inferred from aftershocks. *Nonlin. Process. Geophys.* **15,** 245–263 (2008).
28. Schorlemmer, D. & Wiemer, S. Microseismicity data forecast rupture area. *Nature* **434,** 1086 (2005).

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

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.N. (narteau@ipgp.jussieu.fr).

METHODS

The selection of main shock-aftershock sequences requires four parameters: two for the aftershock magnitude range $\left[M_{\mathrm{min}}^{\mathrm{A}}, M_{\mathrm{max}}^{\mathrm{A}}\right]$; one for the radius of influence of individual main shock, R; and one for the maximum main-shock magnitude, $M_{\mathrm{max}}^{\mathrm{M}}$. These parameter values are determined from seismological and statistical constraints. The radius R has to be as small as possible in order to limit the perturbations due to seismic noise, but large enough to capture a significant number of aftershocks within the first day. Practically, we take $R = 0.02$ km according to the relationship between the magnitude and the rupture length[29]. The lower aftershock magnitude threshold $M_{\mathrm{min}}^{\mathrm{A}} = 1.8$ is given by catalogue completeness[17] in southern California. The higher magnitude threshold $M_{\mathrm{max}}^{\mathrm{A}}$ is given from the $M_{\mathrm{min}}^{\mathrm{A}}$ value using a constant magnitude interval, Δ, determined according to the number of aftershocks in the catalogue. Because the results depend on the aftershock magnitude range, it is always better to take the Δ value as small as possible (here $\Delta = 1$). To determine $M_{\mathrm{max}}^{\mathrm{M}}$, we analyse one single magnitude range of aftershocks from 1.8 to 2.5 in southern California. Classes of main shocks are formed using a sliding window of width 0.5 that moves in 0.1 unit steps from $M_{\mathrm{min}}^{\mathrm{M}}$ to 5.3. For each class, we generate the corresponding list of aftershocks, and estimate $<M^{\mathrm{A}}>$, the average magnitude of these aftershocks (Supplementary Fig. 1). For an increasing main-shock magnitude, the $<M^{\mathrm{A}}>$ value is generally increasing if $M^{\mathrm{M}} > 4.5$ because smaller aftershocks are increasingly missing in the stacks. Below this threshold, the value of $<M^{\mathrm{A}}>$ oscillates around an equilibrium value, indicating that the impact of overlapping records on catalogue completeness is stationary; therefore, we set $M_{\mathrm{max}}^{\mathrm{M}} = 4.5$.

Without distinction of faulting style, we evaluate the c value separately for different magnitude ranges of aftershocks. Thus, we observe that the characteristic time before the power-law decay starts is continuously decreasing as the average magnitude of aftershocks is increasing (Supplementary Fig. 2). A similar relationship has been reported for individual earthquakes in California and Japan[5,9]. These observations explain the changes in mean c values for different aftershock magnitude ranges in Figs 1a, b and 2a, b (see also Supplementary Figs 9 and 10). Then, absolute variations of the mean c value can be entirely separated from the dependence of the c value on the rake angle.

For the maximum-likelihood estimates of c, the optimization procedure depends on two values: T_{comp}, the time at which the catalogue is complete (that is, all aftershocks are recorded), and T_{start}, the time at which we start the fit. It is evident that if $c < T_{\mathrm{comp}}$, the estimated c value cannot be considered as reliable. In the case $c < T_{\mathrm{start}}$, the error on the c value is large and increases exponentially with T_{start}. In the case $T_{\mathrm{start}} < T_{\mathrm{comp}} < c$, the c value may be overestimated because of catalogue incompleteness. Therefore, the preferred condition for the estimation of the c value is $T_{\mathrm{comp}} < T_{\mathrm{start}} < c$.

For Japan, we take advantage of a hand-picked catalogue[7] to determine values of c and T_{comp} that are not affected by missing aftershocks. For different aftershock magnitude ranges, we verify that $c > T_{\mathrm{comp}}$. Then, using the events listed in the JMA catalogue, we can use an increasing T_{start} value to converge towards a good estimation of the c values (that is, the c values obtained from the hand-picked catalogue). This convergence depends on the proportion of recorded events (see Supplementary Table 1, and Supplementary Fig. 7a–c) but also on their distribution over time (see Supplementary Fig. 4). For $2.4 \leq M^{\mathrm{A}} \leq 3.4$ aftershocks, we show that the JMA catalogue can be considered complete and we can take a very small value of T_{start}. For $1.8 \leq M^{\mathrm{A}} \leq 2.8$ aftershocks, it is recommended to take $T_{\mathrm{start}} = 5 \times 10^{-4}$ day because this is the time at which the proportion of recorded events switches rapidly from 0 to 1 (see Supplementary Fig. 4). Then, for these events also, the c value is larger for normal events, smaller for thrust events, and takes intermediate values for strike-slip events (Fig. 2a).

Most importantly, for any aftershock catalogue for which the level of completeness is not well defined, it is possible to use the starting time of the optimization procedure to estimate the stability of the c value and verify that missing aftershocks do not perturb the estimation of this parameter. In the present study, we perform such an analysis for southern California (Supplementary Fig. 8). We show that for $T_{\mathrm{start}} = 10^{-4}$ days (9 s), $T_{\mathrm{start}} = 5 \times 10^{-4}$ days (45 s), $T_{\mathrm{start}} = 10 \times 10^{-4}$ days (90 s) and $T_{\mathrm{start}} = 14 \times 10^{-4}$ days (2 min) there is no significant change in the estimation of c values for all classes of rake angles. As the catalogue of $M^{\mathrm{A}} \geq 1.8$ aftershocks is complete 2 min after a $M^{\mathrm{M}} < 4.5$ main shock[10], this stability in the c value clearly indicates that the missing aftershocks do not affect the estimation of this parameter in this particular case (that is, $T_{\mathrm{comp}} < T_{\mathrm{start}} < c$ for all considered parameters, including different rake angles).

29. Wells, D. L. & Coppersmith, K. J. New empirical relationships among magnitude, rupture length, rupture width, rupture area and surface displacement. *Bull. Seismol. Soc. Am.* **84**, 974–1002 (1994).

nature Vol 462|3 December 2009|doi:10.1038/nature08589

LETTERS

Regulation of adaptive behaviour during fasting by hypothalamic Foxa2

Jose P. Silva[1]†, Ferdinand von Meyenn[2]*, Jessica Howell[1,2]*, Bernard Thorens[3], Christian Wolfrum[2]
& Markus Stoffel[1,2]

The lateral hypothalamic area is considered the classic 'feeding centre', regulating food intake, arousal and motivated behaviour through the actions of orexin and melanin-concentrating hormone (MCH)[1–3]. These neuropeptides are inhibited in response to feeding-related signals and are released during fasting. However, the molecular mechanisms that regulate and integrate these signals remain poorly understood. Here we show that the forkhead box transcription factor Foxa2, a downstream target of insulin signalling[4–6], regulates the expression of orexin and MCH. During fasting, Foxa2 binds to MCH and orexin promoters and stimulates their expression. In fed and in hyperinsulinemic obese mice, insulin signalling leads to nuclear exclusion of Foxa2 and reduced expression of MCH and orexin. Constitutive activation of Foxa2 in the brain (*Nes-Cre/+;Foxa2T156A$^{flox/flox}$* genotype) results in increased neuronal MCH and orexin expression and increased food consumption, metabolism and insulin sensitivity. Spontaneous physical activity of these animals in the fed state is significantly increased and is similar to that in fasted mice. Conditional activation of Foxa2 through the T156A mutation expression in the brain of obese mice also resulted in improved glucose homeostasis, decreased fat and increased lean body mass. Our results demonstrate that Foxa2 can act as a metabolic sensor in neurons of the lateral hypothalamic area to integrate metabolic signals, adaptive behaviour and physiological responses.

We mapped the expression of Foxa2 in the central nervous system using specific antibodies. Strong neural immunoreactivity was exclusively observed in the lateral hypothalamic area (LHA), the subthalamic nucleus, the zona incerta, the substantia nigra and the ventral tegmental area. The related transcription factor Foxa1, which is not regulated by insulin signalling, showed a strikingly similar pattern of expression. Because orexin- and MCH-expressing neurons are known to be located in the LHA, the subthalamic nucleus and the zona incerta, we co-stained these regions in the mouse brain at 200-μm intervals with Foxa2 and MCH/orexin antibodies. Whereas MCH neurons are located in the anterior part of the LHA, orexin neurons are predominantly found in the posterior part of the LHA, the zona incerta and the subthalamic nucleus. All Foxa2 positive neurons co-stained either with MCH or with orexin in the anterior and posterior regions of the LHA. These results show that Foxa2 is expressed in two distinct neuronal cell populations, namely MCH and orexin neurons (Fig. 1a, b).

In the liver, Foxa2 activates genes encoding enzymes of beta oxidation and ketogenesis during fasting. This pathway is switched off by insulin in the fed state through Foxa2 phosphorylation at Thr 156 and nuclear exclusion (Supplementary Fig. 1)[5]. We investigated whether insulin can regulate the activity of Foxa2 in hypothalamic orexin- and MCH-expressing neurons by analysing the cellular localization of Foxa2 in fasted mice injected with either PBS or insulin. Whereas Foxa1 protein was always detected in the nucleus, Foxa2 immunoreactivity was cytosolic in insulin-treated mice. In contrast, PBS-treated mice showed mainly a nuclear localization of Foxa2 in the LHA and the subthalamic nucleus (Fig. 1c, d). Furthermore, immunoblotting of Foxa2 in hypothalamic nuclear and cytosolic extracts and immunostaining detected Foxa2 in nuclei of fasted mice and in the cytosol of fed mice (Fig. 1e, f). Foxa2 localization was unaffected by the nutritional state or by insulin in the substantia nigra and the ventral tegmental area (Supplementary Fig. 2a, b). The data in Fig. 1 show that the subcellular localization of Foxa2 is regulated by plasma insulin concentrations in MCH- and orexin-positive neurons of the hypothalamus.

We investigated the effect of feeding/fasting on messenger-RNA concentrations of MCH and orexin in C57Bl/6 mice. Hypothalamic MCH and orexin expression increased in fasted mice in comparison with fed mice, whereas Foxa2 concentrations did not change (Supplementary Fig. 3a–c). To investigate whether Foxa2 might be directly responsible for the feeding-dependent regulation of MCH and orexin, we measured promoter activities in reporter assays. *Pmch* and *Hcrt*, the respective genes encoding MCH and orexin, contain three potential forkhead DNA-binding sites in their proximal promoters. Constitutive active Foxa2 (Foxa2T156A) increased MCH reporter activity by a factor of ~5 and orexin reporter activity by a factor of ~8, and was unaffected by insulin. In contrast, activation of reporter activity by wild-type Foxa2 was inhibited by insulin. Co-transfection of an expression vector for wild-type Foxa2 and Akt2 in insulin-sensitive cells abolished Foxa2 reporter activity in a dose-dependent manner (Supplementary Fig. 3d–g).

To test further whether MCH and orexin expression is regulated by insulin, glucose or both, we cultured hypothalamus from C57Bl/6 mice for 4 h in medium containing either a low (5 mM) or a high (16 mM) concentration of glucose in the absence of insulin, or a low glucose concentration in the presence of 100 nM insulin. Western blotting of nuclear and cytosolic extracts showed nuclear localization of Foxa2 in hypothalamus that was cultured in high or low glucose concentrations without insulin. In contrast, Foxa2 was detected exclusively in cytosolic fractions when hypothalamus was cultured at a low glucose concentration in the presence of insulin (Supplementary Fig. 3h). Quantitative PCR measurements of MCH and orexin concentrations confirmed that glucose had no effect on the expression of these neuropeptides but that insulin had a strong inhibitory effect. Importantly, Foxa2 mRNA levels were similar in all conditions, suggesting the observed differences were not caused by changes in Foxa2 expression (Supplementary Fig. 3i–k).

---

[1]The Rockefeller University, Laboratory of Metabolic Diseases, 1230 York Avenue, New York, New York 10021, USA. [2]Swiss Federal Institute of Technology, ETH Zürich, Institute for Molecular Systems Biology, Wolfgang Pauli Strasse 16, 8093 Zürich, Switzerland. [3]Center of Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland. †Present address: Department of Neuroscience, The Scripps Research Institute, Scripps Florida, 130 Scripps Way, Jupiter, Florida 33458, USA.
*These authors contributed equally to this work.

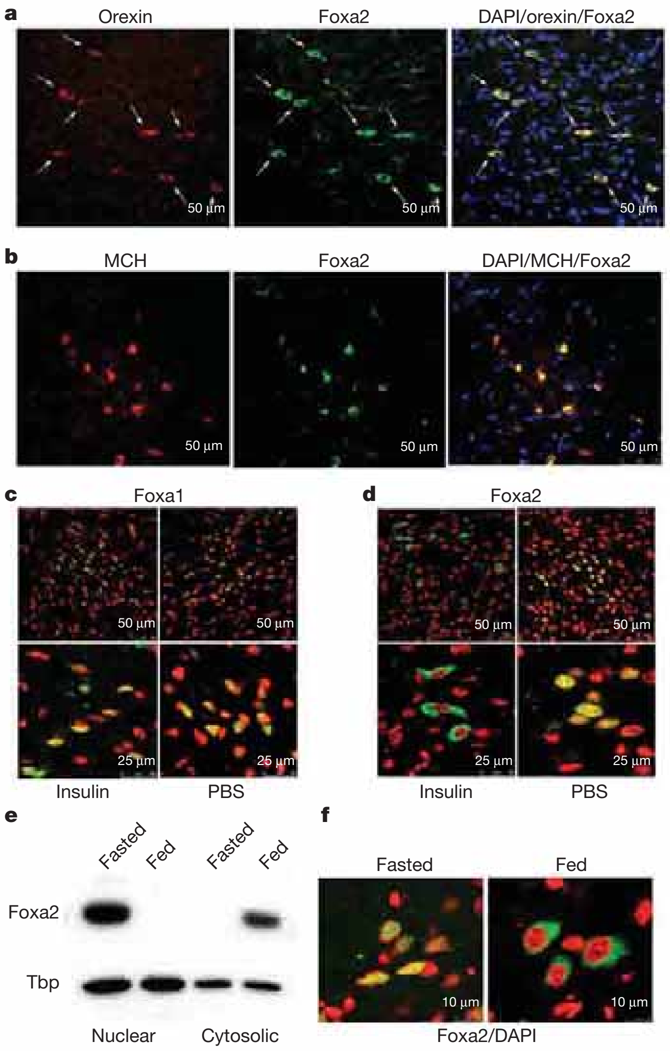


Figure 1 | Foxa2 is co-expressed with MCH and orexin in the lateral hypothalamus and its subcellular localization is regulated by insulin. **a, b**, Brain sections of fed mice were stained with antibodies against Foxa2 (green) and orexin (red) (**a**) or against Foxa2 (green) and MCH (red) (**b**) and counterstained with the nuclear dye 4′,6-diamidino-2-phenylindole (DAPI, blue). Arrows in **a** indicate neurons in which both Foxa2 and orexin co-localize. Co-localization of Foxa2 with orexin or MCH is demonstrated by the yellow staining of neurons. **c, d**, Immunofluorescence analysis to study the subcellular distributions of Foxa1 (**c**) and Foxa2 (**d**) was performed in fasted mice following tail-vein injection with either insulin (1 μg; $n = 3$) or PBS ($n = 3$). Sections were stained with antibodies against Foxa1 (green) or Foxa2 (green) and counterstained with DAPI (red). **e**, Immunoblot of nuclear and cytoplasmic hypothalamic extracts of fasted and fed wild-type C57Bl/6 mice. Detection of TATA-binding protein (Tbp) was used as a loading control. **f**, Immunohistochemical analysis of hypothalamic sections of fasted and fed C57Bl/6 mice stained with anti-Foxa2 antibodies (green) and DAPI (red).

To confirm the functionality of Foxa2-binding sites *in vivo*, we performed chromatin immunoprecipitation assays in hypothalamic extracts from fasted and fed mice. In fasted mice, DNA containing three putative Foxa2-binding sites in the MCH and orexin promoters was amplified from chromatin of immunoprecipitates using Foxa2 antibodies. In contrast, these DNA fragments could not be amplified from chromatin immunoprecipitation of fed mice (Supplementary Fig. 3l, m). Together these data demonstrate that Foxa2 activity in the hypothalamus is regulated by insulin signalling and that Foxa2 in the absence of insulin signalling is recruited to the MCH and orexin promoters to activate their expression.

We have previously demonstrated that in livers of obese mice, Foxa2 is permanently localized in the cytosol and is inactive. To test whether Foxa2 is also excluded from the nucleus of MCH- and orexin-expressing neurons, we investigated Foxa2 localization in the hypothalamus of C57Bl/6 mice that were fed a high-fat diet (HFD) for 12 weeks. Immunohistochemistry of Foxa2 in fasted mice revealed a cytosolic localization in neurons of the LHA (Fig. 2a). These results were confirmed by comparing expression of Foxa2 protein in nuclear and cytosolic extracts of fasted/fed mice on a chow diet or a HFD (Fig. 2b). The expression of MCH and orexin in fasted mice and fed mice on a HFD were similar to that in fed mice on a chow diet and were significantly lower in comparison with that in fasted mice on a chow diet. Expression of *Cpt1a*, *Mcad* (*Acadm*) and *Vlcad* (*Acadvl*), validated Foxa2 target genes, were also reduced in hypothalamus of fasted mice on a HFD relative to that of fasted mice on a chow diet (Fig. 2c, d), whereas Foxa2 levels were similar in all groups. This indicates that the observed changes in gene expression are not caused by differences in Foxa2 expression.

We used a genetic approach to examine the physiological effects of Foxa2 activation in the hypothalamus further. We generated conditional mutant mice with a constitutive active *Foxa2T156A* allele by introducing *loxP* sites upstream and downstream of exon 3, which encodes most of the open reading frame of Foxa2 and the 3′ untranslated region. An exon-3 duplication, harbouring the T156A mutation, was introduced

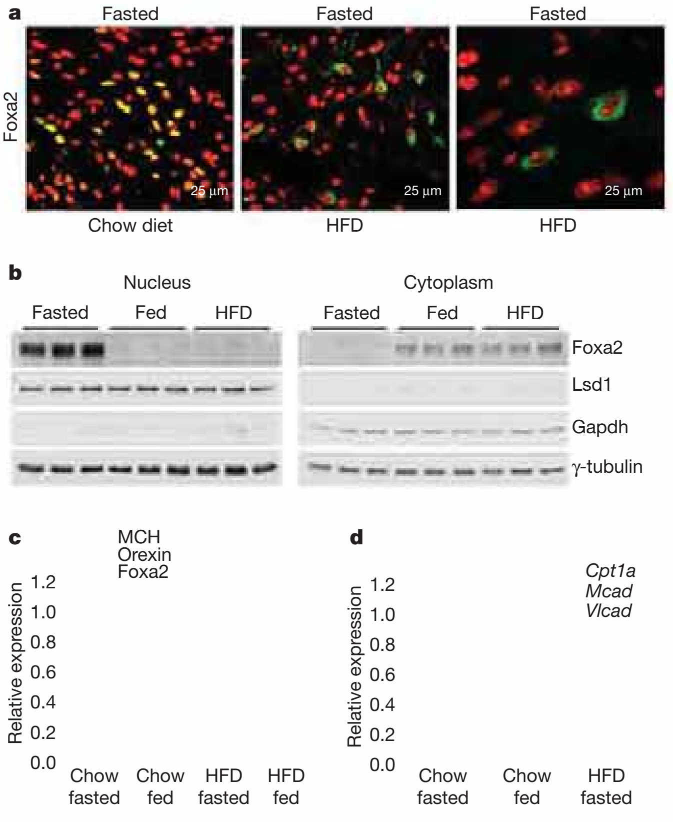


Figure 2 | Permanent inactivation of Foxa2 and reduced expression of MCH and orexin in hypothalamus of hyperinsulinemic, obese mice. **a**, Immunohistochemistry of hypothalamic brain sections of fasted C57Bl/6 mice that were either fed a chow diet or a HFD. Sections were stained with antibodies against Foxa2 (green) and counterstained with the DAPI (red). Yellow staining indicates nuclear co-localization of Foxa2. **b**, Immunoblot analysis of nuclear and cytoplasmic fractions of hypothalamus from fed or fasted C57Bl/6 mice and fasted HFD C57Bl/6 mice. Anti-γ-tubulin antibodies were used as a loading control, anti-Lsd1 (Lsd1 also known as Kdm1a) and Gapdh antibodies were used as nuclear and cytosolic markers, respectively. **c, d**, Measurement of relative gene expression levels of MCH, orexin and Foxa2 (**c**) and Foxa2 target genes *Cpt1a*, *Mcad* and *Vlcad* (**d**), in hypothalamus of chow or HFD C57Bl/6 mice that were either fasted or fed ($n = 5$–9 in each group; values are mean ± s.d.; $^{*}P \leq 0.05$, $^{**}P \leq 0.01$).

downstream of the floxed exon 3 (Supplementary Fig. 3a). Transgenic nestin-Cre (*Nes-Cre*) mice expressing Cre recombinase under the neuron-specific promoter nestin have been used successfully to achieve *loxP*-mediated recombination in the brain[7]. *Foxa2T156A*$^{flox/flox}$ mice were mated with *Nes-Cre/+* mice to produce progeny in which the floxed exon 3 was excised. Subsequent breeding segregated the *Cre* transgene to yield the *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$, *Foxa2T156A*$^{flox/flox}$ and *Nes-Cre/+* genotypes for further study. Southern blot analysis of multiple tissues of *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ mice showed that recombination had occurred in the brain, but not in other tissues (Supplementary Fig. 3b). We also sequenced Foxa2 transcripts from hypothalamus and liver ($n = 10$). Ninety per cent of the brain clones harboured the T156A mutation, whereas none of the liver clones showed the amino-acid substitution.

To test the functionality of the conditional *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ allele, we studied Foxa2 expression using immunohistochemistry. Hypothalamic Foxa2 immunoreactivity was nuclear in fed *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ mice. In contrast, Foxa2 was excluded from the nucleus in fed *Foxa2T156A*$^{flox/flox}$ mice (Fig. 3a). Intraperitoneal injection of superphysiological doses of insulin into fasted *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ and control mice showed nuclear Foxa2 localization in *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ mice, whereas Foxa2 in control animals was cytosolic (Supplementary Fig. 5a). In contrast, Foxa2 in the liver of *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ mice readily shuttled to the cytosol after insulin treatment, thereby confirming that nuclear Foxa2 localization was restricted to the hypothalamus under conditions of elevated insulin (Supplementary Fig. 5b). We also explored the effect of the *Foxa2T156A* allele on neuropeptide expression in hypothalamus of fed and fasted mice. Transcript levels of MCH and orexin were markedly increased in *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ mice in comparison with control mice when fed, whereas they were similar under conditions of fasting (Fig. 3b and data not shown). Expression of known Foxa2 targets was increased in *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ mice relative to control mice, whereas other neuropeptides were not regulated (Supplementary Fig. 5c, d).

We investigated the physiological consequences of central Foxa2T156A expression on energy homeostasis, food intake and spontaneous activity. Plasma glucose concentrations of fed mice were

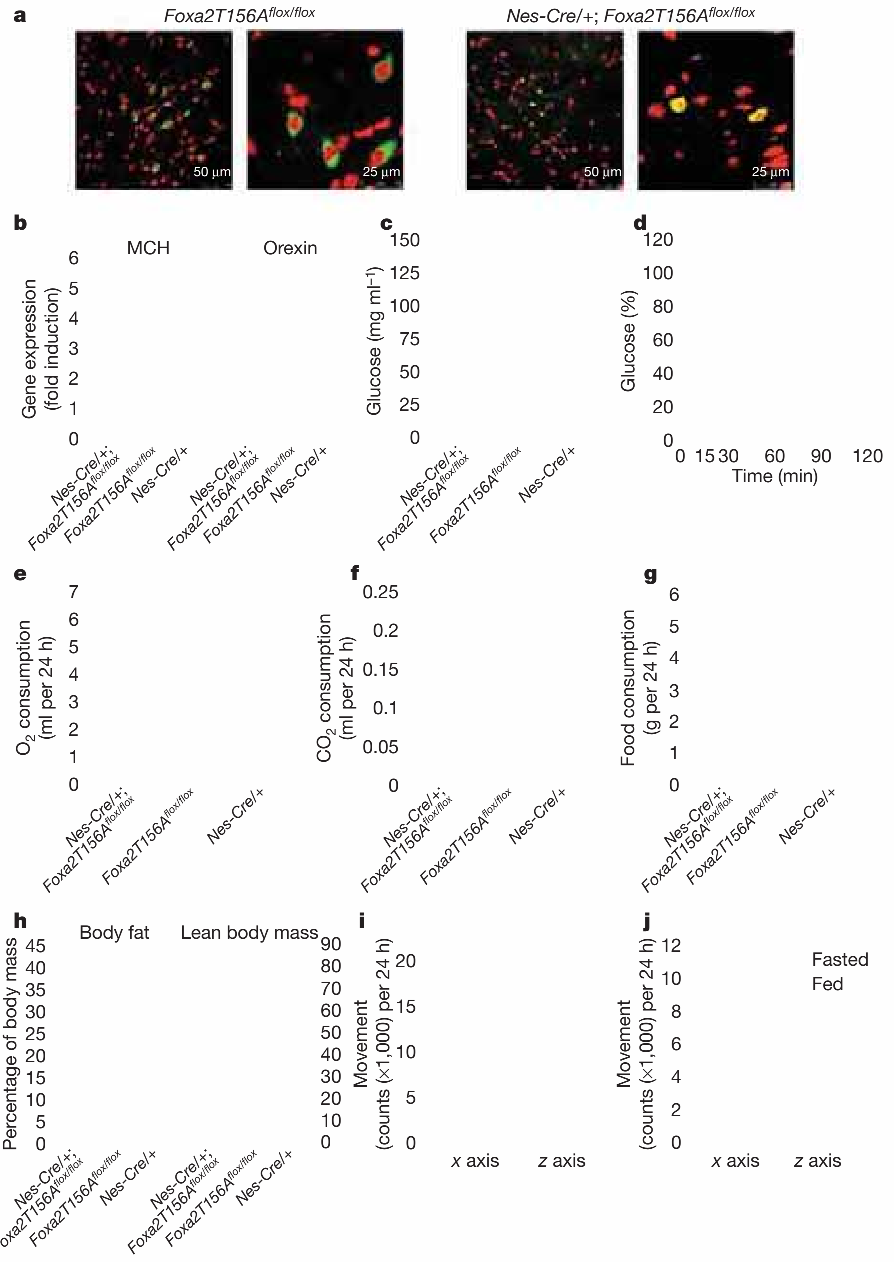


Figure 3 | Constitutive nuclear Foxa2T156A activates *Pmch* and *Hcrt* expression in fed mice. **a**, Immunofluorescence analysis of Foxa2 subcellular distribution in the lateral hypothalamus of fed *Foxa2T156A*$^{flox/flox}$ (control) and fed *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ mice. Foxa2 (green) is cytosolic in fed control mice and is nuclear in fed *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ mice ($n = 3$ in all groups). DAPI (red) was used as a nuclear counterstaining dye. **b**, Real-time reverse transcription–polymerase chain reaction analysis of MCH and orexin. Transcript levels were normalized to β-actin mRNA. The results are expressed as fold induction over mRNA of fed *Foxa2T156A*$^{flox/flox}$ mice (mean ± s.e.m.; $n = 5$ in all groups; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$). **c–h**, Metabolic and locomotion measurements in *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ mice, *Nes-Cre/+* mice and *Foxa2T156A*$^{flox/flox}$ mice at 8 weeks of age. **c**, Plasma glucose concentration. **d**, Time dependence of percentage of starting plasma glucose concentration during an insulin tolerance test. The areas under the curve between the 60-min and 90-min time points for *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ (blue), *Foxa2T156A*$^{flox/flox}$ (red) and *Nes-Cre/+* (green) mice are 7.9 ± 0.46, 9.8 ± 1.3 and 10.3 ± 1.1 mmol, respectively. For *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ vs *Foxa2T156A*$^{flox/flox}$, $P \leq 0.05$; for *Nes-Cre/+;Foxa2T156A*$^{flox/flox}$ vs *Nes-Cre/+*, $P \leq 0.001$. **e**, O_2 consumption. **f**, CO_2 production. **g**, Food intake. **h**, Body fat and lean body mass as measured by DEXA. **i**, Spontaneous physical activity assessed as the number (counts) of interruptions of photocells placed in the *y* and *z* axes (colour-coded as in **d**). **j**, Locomotion in fasted and fed C57Bl/6 mice. All values are the mean ± s.e.m.; $n = 5$–9 in each group; $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$).

decreased in *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ relative to controls (Fig. 3c). Glucose clearance from the circulation during an insulin-tolerance test was modestly increased in *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ mice (Fig. 3d). Furthermore, fasting plasma insulin and free fatty acids were reduced in these mice, indicating increased insulin sensitivity (Supplementary Fig. 5e, f). Plasma leptin levels were similar in *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ and control mice (data not shown). Measurements of metabolic rates revealed that *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ mice had increased O$_2$ consumption, CO$_2$ production and heat production relative to controls (Fig. 3e, f and data not shown). Food consumption was also increased in *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ animals (Fig. 3g). Dual-energy X-ray absorptiometry (DEXA) analyses revealed that *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ mice have reduced body fat and increased lean body mass (Fig. 3h). Interestingly, the *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ allele was associated with dramatic increases in spontaneous locomotor activity relative to control mice (Fig. 3i). The difference between the locomotor activity of *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ mice and that of *Foxa2T156A*$^{flox/flox}$ or *Nes-Cre*/+ mice was similar to the increase in movement of fasted wild-type mice relative to fed wild-type mice (Fig. 3j). The types of physical activity induced in *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ mice included searching as well as intense grooming, rearing and face-washing behaviour.

In addition to the genetic recombination of the conditional *Foxa2T156A* allele by nestin-regulated *Cre* expression, which results in the activation of Foxa2 in the central nervous system during late embryonic development and throughout adult live, we also used two approaches to induce nutrient-independent activation of Foxa2 in the lateral hypothalamus. We stereotactically injected adenovirus-expressing Cre recombinase (Ad-Cre) or control adenovirus (Ad-GFP) into the lateral hypothalamus of *Foxa2T156A*$^{flox/flox}$ mice and studied them metabolically. Constitutive activation of Foxa2 was confirmed by sequencing of hypothalamic Foxa2 transcripts and by immunoblotting of nuclear and cytosolic extracts from fed mice (Supplementary Fig. 6a). Expression levels of MCH and orexin, as well as beta-oxidation target genes, were increased in *Foxa2T156A*$^{flox/flox}$ mice treated with Ad-Cre relative to mice treated with Ad-GFP, whereas the expression of other neuropeptides was not affected (Supplementary Fig. 6b). Similar to the *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ mice, mice treated with Ad-Cre exhibited no differences in body mass relative to mice treated with Ad-GFP, but had increased food, water and O$_2$ consumption and CO$_2$ production. In addition, spontaneous physical activity was increased ~4-fold relative to control mice (Supplementary Fig. 6c–h). As a separate approach, we also injected C57Bl/6 mice with adenovirus expressing either *Foxa2T156A* under the control of a cytomegalovirus promoter or green fluorescent protein (GFP; control) in the LHA (Supplementary Fig. 7). Again, gene expression and physiological measurements were similar to those in the *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ mice and the *Foxa2T156A*$^{flox/flox}$ animals treated with Ad-Cre (Supplementary Fig. 7a–g). Together these data demonstrate that constitutive active Foxa2 in the LHA increases MCH and orexin expression and profoundly increases spontaneous physical activity.

To study further the effect on metabolism of constitutive Foxa2 activation in the brain, we examined *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ mice and *Nes-Cre*/+ and *Foxa2T156A*$^{flox/flox}$ control mice in all of which obesity had been induced with a HFD. The body mass of 16-week-old mice was indistinguishable between the groups (Fig. 4a). Blood glucose concentrations were decreased in *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ mice relative to control mice (Fig. 4b). Furthermore, plasma insulin and free-fatty-acid concentrations were decreased (Fig. 4c, d) and food and water consumption were increased (Fig. 4e, f) in mice with constitutive active Foxa2. Similar to chow-fed mice, HFD obese mice had increased metabolism as measured by enhanced O$_2$ consumption and CO$_2$ production. In addition, locomotor activity was increased in *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ relative to control mice (data not shown). DEXA measurements showed that the percentage body fat decreased in *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ mice relative to their *Nes-Cre*/+ and *Foxa2T156A*$^{flox/flox}$ littermates (Fig. 4g). In contrast, total lean body mass, mainly a measure of muscle mass, was increased in *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$ relative to controls (Fig. 4h).

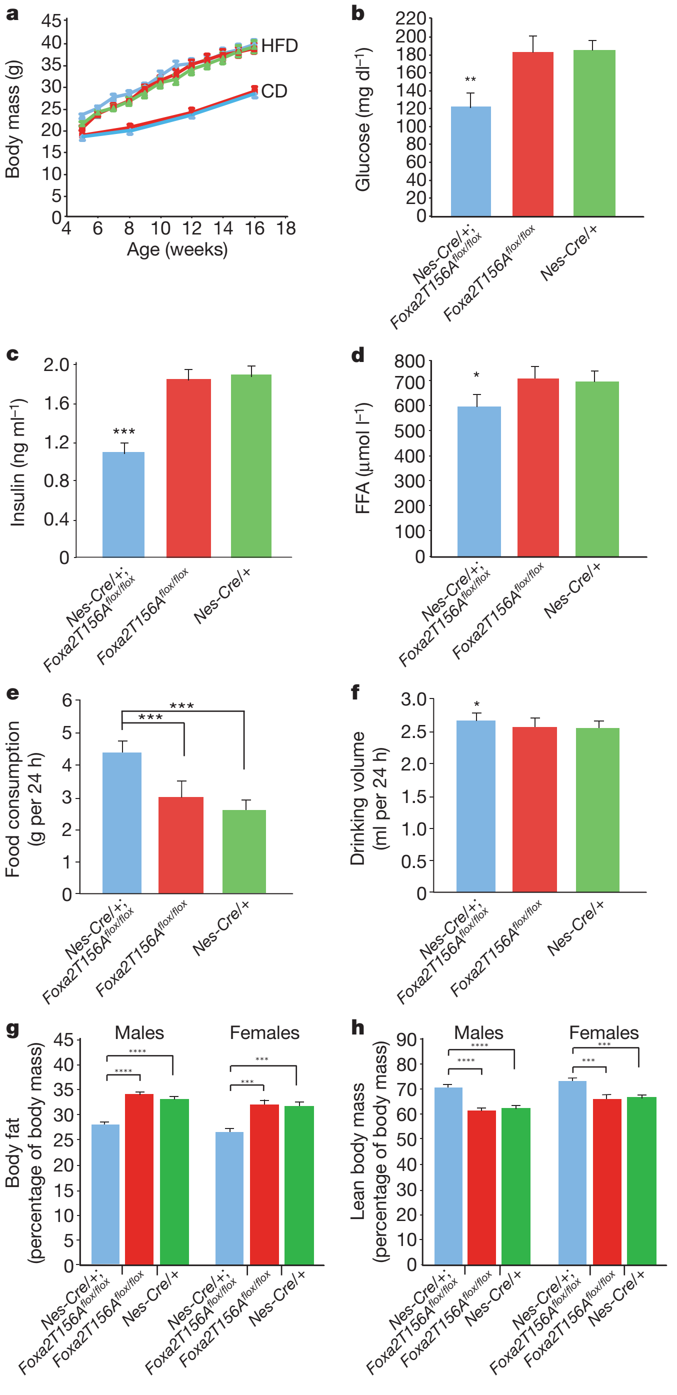


Figure 4 | Metabolic measurements in HFD, brain-specific *Foxa2T156A* mice. **a**, Body mass of male animals (blue, *Nes-Cre*/+;*Foxa2T156A*$^{flox/flox}$; red, *Foxa2T156A*$^{flox/flox}$; green, *Nes-Cre*/+; CD, chow diet). **b**, Plasma glucose concentration. **c**, Plasma insulin concentration (fasted). **d**, Plasma free fatty acid (FFA). **e**, Food consumption. **f**, Water consumption. **g**, Body fat. **h**, Lean body mass. Body fat and lean body mass were measured by DEXA. All values are the mean ± s.e.m.; $n = 5$–9 in each group; $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

Neurons in the lateral hypothalamus are particularly important for feeding, behavioural arousal and addiction[8]. Animals with lesions of the LHA exhibit hypophagia and decreased arousal and fail to react to homeostatic challenges such as fasting with appropriate feeding-related behavioural and physiologic responses[9]. Our studies suggest one molecular mechanism by which insulin controls the activity of the LHA. During fasting, Foxa2 acts as a transcriptional activator of *Pmch* and *Hcrt*, which in turn stimulates the classical fasting responses such as feeding, arousal and food-seeking behaviour. In the fed state, Foxa2 is phosphorylated by insulin/PI3K/Akt signalling, which leads to Foxa2 inactivation by nuclear exclusion. MCH and orexin expression is regulated by the same metabolic stimuli and molecular signalling pathway, thereby integrating metabolic and adaptive behavioural responses.

The combined activation of MCH- and orexin-expressing neurons may also explain why energy expenditure is increased but energy balance is not affected in *Nes-Cre/+;Foxa2T156A$^{flox/flox}$* mice. Increased food intake in *Nes-Cre/+;Foxa2T156A$^{flox/flox}$* mice is most probably mediated by increased MCH expression, because it has been shown to act as a potent positive regulator of food intake[2,10–12]. The lack of regulation of either plasma leptin or Npy, Pomc and Agrp expression levels in the hypothalamus of *Nes-Cre/+;Foxa2T156A$^{flox/flox}$* mice relative to control animals makes it unlikely that these peptides significantly contribute to the observed phenotype. Increased orexin expression probably mediates increased wakefulness and spontaneous physical activity, as pharmacological and genetic studies have shown that orexin mediates arousal, increased locomotor activity and searching behaviour[13,14]. Our studies show that fed *Nes-Cre/+;Foxa2T156A$^{flox/flox}$* mice have similar levels of orexin expression to fasted control mice. However, mice with constitutive active Foxa2 have slightly lower activity levels relative to fasted control mice, indicating that there are additional mechanisms accounting for the locomotor response of fasted mice.

As in the liver, Foxa2 in MCH- and orexin-expressing neurons is permanently inactive in hyperinsulinemic mice with diet-induced obesity[15]. Constitutive activation of brain Foxa2 in obese mice leads to improved glucose homeostasis, decreased fat and increased lean body mass linked to an increase in physical activity. Hence, our studies may offer a molecular explanation for the inverse associations of insulin resistance/hyperinsulinemia and physical inactivity/energy expenditure[16–18] that have been observed in humans[19–21]. Because the decline in daily physical activity as well as unbalanced food consumption is recognized as one major factor contributing to the current worldwide obesity epidemic, pharmacological inhibition of Foxa2 phosphorylation may improve levels of physical activity, overall health and longevity.

METHODS SUMMARY

Mouse and metabolic studies. We maintained all mouse models in the C57Bl/6 background and kept them on a 12-h:12-h light–dark cycle in a pathogen-free animal facility. Groups of mice were fed a HFD (TD93075, Harlan Teklad) for 12 weeks, starting at the age of 4 weeks. Locomotion (x, y and z axis), food and water intake, oxygen consumption, and CO_2 and heat production were simultaneously determined for four mice in separate cages per experiment during a 24-h period in an Oxymax metabolic chamber system (Columbus Instruments) at the ages of 4 and 8 weeks. We performed DEXA measurements (PIXImus, GE Healthcare) at ages of 8 and 12 weeks. The generation of the conditional mouse is described in Methods.

Nuclear translocation experiments. After fasting them for 24 h, we injected 4–6-week-old C57Bl/6 mice with 1 μg human insulin (Sigma-Aldrich) or PBS into their tail veins. Mice were killed 20 min after the injection. Brains were further processed and analysed as described in Methods.

Laser scanning confocal microscopy. We scanned sections using an inverted Leica TCS SPE. Scan were performed at a resolution of 1,012 × 1,012 and averaged three times. The optical-slice thickness was the same for each channel. Pictures were acquired and analysed with LAS AF software (Leica).

Metabolic measurements. Blood samples were taken from mice into non-heparinized capillary tubes. Insulin was quantified by a radioimmunoassay (Linco). Glucose was measured using a standard glucose sensor (Glucometer Elite, Bayer). Triglycerides, free fatty acids and cholesterol were measured by colorimetric assays (Wako Chemicals, Roche). Leptin levels were measured using the leptin ELISA kit (Crystal Chem) according to the manufacturer's instructions.

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

Received 25 June; accepted 19 October 2009.

1. Willie, J. T., Chemelli, R. M., Sinton, C. M. & Yanagisawa, M. To eat or to sleep? Orexin in the regulation of feeding and wakefulness. *Annu. Rev. Neurosci.* **24,** 429–458 (2001).
2. Shimada, M., Tritos, N. A., Lowell, B. B., Flier, J. S. & Maratos-Flier, E. Mice lacking melanin-concentrating hormone are hypophagic and lean. *Nature* **396,** 670–674 (1998).
3. Sakurai, T. *et al.* Orexins and orexin receptors: a family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell* **92,** 573–585 (1998).
4. Wolfrum, C., Besser, D., Luca, E. & Stoffel, M. Insulin regulates the activity of forkhead transcription factor Hnf-3beta/Foxa-2 by Akt-mediated phosphorylation and nuclear/cytosolic localization. *Proc. Natl Acad. Sci. USA* **100,** 11624–11629 (2003).
5. Wolfrum, C., Asilmaz, E., Luca, E., Friedman, J. M. & Stoffel, M. Foxa2 regulates lipid metabolism and ketogenesis in the liver during fasting and in diabetes. *Nature* **432,** 1027–1032 (2004).
6. Matsuzaka, T. *et al.* Crucial role of a long-chain fatty acid elongase, Elovl6, in obesity-induced insulin resistance. *Nature Med.* **13,** 1193–1202 (2007).
7. Tronche, F. *et al.* Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety. *Nature Genet.* **23,** 99–103 (1999).
8. Bernardis, L. L. & Bellinger, L. L. The lateral hypothalamic area revisited: neuroanatomy, body weight regulation, neuroendocrinology and metabolism. *Neurosci. Biobehav. Rev.* **17,** 141–193 (1993).
9. Hara, J. *et al.* Genetic ablation of orexin neurons in mice results in narcolepsy, hypophagia, and obesity. *Neuron* **30,** 345–354 (2001).
10. Gomori, A. *et al.* Chronic intracerebro-ventricular infusion of MCH causes obesity in mice. Melanin-concentrating hormone. *Am. J. Physiol. Endocrinol. Metab.* **284,** E583–E588 (2003).
11. Ludwig, D. S. *et al.* Melanin-concentrating hormone overexpression in transgenic mice leads to obesity and insulin resistance. *J. Clin. Invest.* **107,** 379–386 (2001).
12. Alon, T. & Friedman, J. M. Late-onset leanness in mice with targeted ablation of melanin concentrating hormone neurons. *J. Neurosci.* **26,** 389–397 (2006).
13. Hagan, J. J. *et al.* Orexin A activates locus coeruleus cell firing and increases arousal in the rat. *Proc. Natl Acad. Sci. USA* **96,** 10911–10916 (1999).
14. Hara, J. *et al.* Genetic ablation of orexin neurons in mice results in narcolepsy, hypophagia, and obesity. *Neuron* **30,** 345–354 (2001).
15. Byberg, L., Zethelius, B., McKeigue, P. M. & Lithell, H. O. Changes in physical activity are associated with changes in metabolic cardiovascular risk factors. *Diabetologia* **44,** 2134–2139 (2001).
16. Aadahl, M., Kjaer, M. & Jorgensen, T. Associations between overall physical activity level and cardiovascular risk factors in an adult population. *Eur. J. Epidemiol.* **22,** 369–378 (2007).
17. Assah, F. K., Brage, S., Ekelund, U. & Wareham, N. J. The association of intensity and overall level of physical activity energy expenditure with a marker of insulin resistance. *Diabetologia* **51,** 1399–1407 (2008).
18. Stolk, R. P. *et al.* Insulin and cognitive function in an elderly population. The Rotterdam study. *Diabetes Care* **20,** 792–795 (1997).
19. Young, S. E., Mainous, A. G. III & Carnemolla, M. Hyperinsulinemia and cognitive decline in a middle-aged cohort. *Diabetes Care* **29,** 2688–2693 (2006).
20. Komulainen, P. *et al.* Metabolic syndrome and cognitive function: a population-based follow-up study in elderly women. *Dement. Geriatr. Cogn. Disord.* **23,** 29–34 (2007).
21. Baranowska, B., Wolińska-Witort, E., Martyńska, L., Chmielowska, M. & Baranowska-Bik, A. Plasma orexinA, orexin B leptin neuropeptide Y (NPY) and insulin in obese women. *Neuroendocrinol. Lett.* **26,** 293–296 (2005).

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

Acknowledgements This study was supported by grants from SystemsX and the Swiss National Science Foundation (LiverX grant). We would like to thank S. Metref for technical support and expertise.

Author Contributions J.P.S.: project design, characterization of mice, transcriptional activation, preparation of manuscript; F.v.M.: immunohistochemistry, nuclear/cytosolic immunoblot analysis, characterization of mice; J.H.: nuclear/cytosolic immunoblot analysis; B.T.: hypothalamic adenovirus injections; C.W.: generation and characterization of mice; M.S.: project design, preparation of manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.S. (stoffel@imsb.biol.ethz.ch).

METHODS

Preparation of nuclear/cytosolic extracts. Hypothalami were dissected from fasted and fed 6-week-old C57Bl/6 male mice and snap frozen. Hypothalami from six fasted and six fed mice were pooled and dounced on ice in hypotonic lysis buffer (10 mM HEPES, pH 7.9; 1.5 mM $MgCl_2$; 10 mM KCl) containing 1 mM DTT and protease inhibitors (complete cocktail, Roche). Lysates were incubated on ice for 15 min and plasma membrane disruption was confirmed by trypan blue staining. Ten per cent NP40 was added to a final concentration of 0.6% and lysates were immediately vortexed for 10 s and centrifuged at 10,500*g* for 30 s at 4 °C. The supernatants (cytoplasmic fractions) were removed, and the nuclear pellets were resuspended in nuclear lysis buffer (10 mM HEPES, pH 7.9; 100 mM KCl; 3 mM $MgCl_2$; 0.1 mM EDTA) containing 1 mM DTT and protease inhibitors. Nuclear lysates were incubated on ice for 30 min; this was followed by the addition of one-tenth volume of 4 M $(NH_4)_2SO_4$ over 30 min. All samples (nuclear and cytoplasmic) were centrifuged at 16,000*g* for 10 min and supernatants were transferred to clean tubes. Protein concentration was determined by BCA assay (Sigma) and 30 μg of each lysate was separated by SDS–PAGE followed by western blot analysis.

Cryosectioning and immunofluorescence. Mice were CO_2 anaesthetized and intracardially perfused with 5 ml phosphate-buffered 2% paraformaldehyde, pH 7.4. Brains were post-fixed in phosphate-buffered 2% paraformaldehyde, pH 7.4, for 10 min at room temperature (24 °C), equilibrated in phosphate-buffered 30% sucrose at 4 °C for 24 h and frozen in Tissue-Tek OCT compound (Sakura). Brains were cryosectioned into 12-μm-thick coronal sections, which were stored at −20 °C for later use. Cryosections were permeabilized for 1 h at room temperature in 10 mM PBS-T (10 mM PBS, pH 7.4; 0.1% Triton X-100), then blocked with 1% BSA, 5% serum in PBS-T for 1 h at room temperature and incubated overnight at 4 °C with the primary antibody at a 1:50–1:300 dilution in 1% BSA, 5% serum in PBS-T. After three washes, sections were incubated secondary antibodies (Alexa Fluor 488, Invitrogen, or Alexa Fluor 568 conjugated, Molecular Probes) at 1:500 dilution in 1% BSA, PBS-T for 1 h at room temperature. Sections were counterstained with DAPI (Vector Laboratories) and mounted with Vectashield (Vector Laboratories). Primary antibodies were custom made or purchased from Abcam (Foxa1, ab23738; Foxa2, ab40874; Orexin A, ab35337) or Santa Cruz Biotechnology (pro-MCH, sc-14509).

Reverse transcription–polymerase chain reaction measurements of orexin and MCH transcripts. Total RNA was extracted from whole hypothalami of mice with the Trizol reagent (Invitrogen). Fifteen micrograms of total RNA were treated with DNaseI for 3 h at 37 °C using the DNA-free kit (Ambion). DNase-treated total RNA was precipitated with ethanol and run on a formaldehyde gel to rule out degradation. Complementary DNAs were generated with the SuperScript III First-Strand Synthesis System (Invitrogen) from 4 μg DNase-treated total RNA and further incubated for 20 min with RNase H. Orexin, MCH and β-actin cDNAs were measured in duplicate by real-time PCR on an ABI 7700 light cycler (Applied Biosystems). Samples were denatured at 94 °C for 15 min, followed by 40 cycles at 94 °C for 15 s and annealing/extension at 60 °C for 45 s. Specific primer and Taqman probe sequences were designed with the Primer Express 2.0 software (Applied Biosystems). Primer sequences were as follows: for orexin, CTGCCGTCTCTACGAACTGTTG and CGCTTTCCCAGAGTCAGGATA; for MCH, TTCAGAAGGAAGATACTGCAGAAAGA and CGCTCTCGTCGTTTTTGTATTG; for β-actin, GAGAAGCTGTGCTATGTTGCTC and AGGAAGAGGATGCGGCA. Taqman Probe sequences were as follows: for orexin, ACGGAGCTGGCAACCACGCTG; for MCH, CCGTTGTCGCCCCTTCTCTGGA; for β-actin, AGACTTCGAGCAGGAGATGGCCA. Standard curves for orexin, MCH and β-actin cDNAs were generated from a dilution series of one sample and used to calculate the dilutions of the cDNAs in the remaining samples. Orexin and MCH cDNA were normalized to β-actin cDNA.

Transfection and transactivation assays. Transactivation experiments were performed in the hepatoma cell line HepG2 and neuroblastoma cell line N2A. Cells were maintained in DMEM supplemented with 4.5 $g\,l^{-1}$ glucose, 10% FCS, 2 mM glutamine and 100 U penicillin/100 μg streptomycin. Cells were grown in 24-well plates to 70% confluency and transfected by using the transfection reagent Lipofectamine 2000 (Invitrogen) with 20–100 ng expression vector (FOXA1, FOXA2, FOXA2T156A), 100 ng firefly luciferase reporter vector (pGL2-Basic Vector or pGL2-Promoter Vector, Promega) harbouring an orexin or MCH promoter fragment and 100 ng pRL-TK *Renilla* luciferase control reporter vector (Promega) per 24 wells. Where indicated, a constitutively active Akt1 kinase expression vector[4] was included. MCH and orexin promoter fragments were cloned by PCR into the TOPO3 vector (Invitrogen) and further subcloned into the polylinker of the pGL2 basic and pGL2 promoter reporter vectors. Cells were grown for 48 h and assayed with the firefly luciferase and *Renilla* luciferase assay systems (Promega). Firefly luciferase activity was normalized for transfection efficiency by *Renilla* luciferase activity.

Chromatin immunoprecipitation analysis (ChIP). Four–six-week-old C57Bl/6 males were CO_2 anaesthetized and perfused with 1% phosphate-buffered paraformaldehyde, pH 7.4. Hypothalami were cut into small pieces and post-fixed in 1% phosphate-buffered paraformaldehyde, pH 7.4, for 15 min at room temperature. Glycine at a final concentration of 0.125 M was added to block further crosslinking. Hypothalami were washed twice with ice-cold PBS and homogenized on ice in 4 ml cell lysis buffer (10 mM Tris, pH 7.4; 10 mM NaCl; 0.2% NP-40) containing protease inhibitors (complete cocktail, Roche). The homogenates of five hypothalami were pooled and centrifuged at 5,500*g* for 5 min at 4 °C. The pellet was again homogenized in 450 μl nuclear lysis buffer (Chromatin Immunoprecipitation Assay Kit, Upstate Biotechnology) and the lysate was sonicated on a Fisher Sonic Dismembrator 550, five times for 20 s at 25% of its maximum power. This resulted in DNA fragments ranging in size from 200 to 1,000 bp as evidenced by gel electrophoresis. Homogenates were centrifuged at 16,000*g* for 10 min at 4 °C. The supernatant was diluted tenfold with ChIP dilution buffer (Chromatin Immunoprecipitation Assay Kit, Upstate Biotechnology). Samples were precleared with Protein A Agarose/Salmon Sperm DNA (Chromatin Immunoprecipitation Assay Kit, Upstate Biotechnology) for 45 min at 4 °C. Samples were centrifuged for 2 min at 1,000*g* and the supernatants were immunoprecipitated overnight at 4 °C on a wheel rotator with 10 μg Foxa2 antibody (ab5074, Abcam). Samples were incubated with Protein A Agarose/Salmon Sperm DNA for 1 h at 4 °C. Beads were washed for 5 min with 1 ml each of low-salt immune complex wash buffer, high-salt immune complex wash buffer, LiCl immune complex wash buffer and TE buffer (Chromatin Immunoprecipitation Assay Kit, Upstate Biotechnology). The DNA–protein complexes were eluted with 1% SDS and 0.1 M $NaHCO_3$ and reverse-crosslinked for 4 h at 65 °C with 0.2 M NaCl. DNA was extracted with phenol–chloroform, ethanol-precipitated using glycogen and resuspended in 10 mM Tris-Cl, pH 8.0. ChIP–PCR was carried out at annealing temperatures of 68 °C (orexin) and 65 °C (MCH) using 30 cycles. Orexin promoter fragments were amplified with the following primer pairs: GGTGGGGTAGGGGTTGGGGAGTA and AGAGGCCTTGGCTGGGACAC; GGACAGGGCGGCCATTCTTTG and CTTCCAGGGGCCAGGGGTTTAT. MCH promoter fragments were amplified with the following primer pairs: GGCGAGCCTGGTATCCTGTGG and ATGACTCACCCCGTAGAGCCTTTGT; AGGCGAGCCTGGTATCCTGTG and GTGGCCTGTTGTTTATTTCGTGTG; AGGCGAGCCTGGTATCCTGT and AAGCACTTGAAGGCCTTATCTTTT.

Generation of *Foxa2T156A* knock-in mice. *Foxa2T156A*$^{flox/flox}$ mice were generated using a targeting construct as depicted in Supplementary Fig. 4a. The short arm spans exons 1 and 2 (3.4 kb in length) and encompasses base-pair positions −2517 to +919, relative to the translation start site (+1). The long arm, which contained exon 3 with the T156A mutation was 4.7 kb in length, between base-pair positions +1034 and +5782, relative to the translation start site. Wild-type exon 3 was located between the long arm and the short arm and flanked by a *loxP* sequence. Furthermore a neo/tk cassette, flanked by *loxP* sites, was inserted upstream of wild-type exon 3. An outside probe as depicted in Supplementary Fig. 4a was used to visualize homologous recombination and Cre recombination events. The targeting vector was electroporated into R1 ES cells from mouse strain 129 (ref. 22) and neomycin-resistant colonies were recovered. The colonies were screened by Southern blotting using probe 1 (Supplementary Fig. 3a). Positive colonies were confirmed by Southern blotting with a probe located downstream of the 3′ homology arm. To delete the neo/tk cassette from the targeted allele, recombinant ES clones were transiently transfected with a plasmid expressing Cre recombinase. ES clones with partial recombination and deletion of the neo/tk cassette were injected into B6(D2B6F1) blastocysts to produce chimaeric mice. Germline transmission was confirmed by Southern blotting.

Foxa2T156A$^{flox/flox}$ mice were first backcrossed for seven generations to C57Bl/6J mice. To obtain brain-specific *Foxa2T156A* knock-in mice, *Foxa2T156A*$^{flox/flox}$ mice were mated with *Nes-Cre* mice (B6.Cg-Tg(Nes-cre)1Kln/J, Jackson Laboratories)[7] that had been backcrossed to C57Bl/6J mice for at least six generations. To confirm expression of the *Foxa2T156A* allele, *Foxa2T156A* transcripts were amplified from brain and liver using the primers AGCGGCCAGCGAGTTAAAGTATGC and CTGCCGGTAGAAAGGGAAGAGGTC spanning one intron of the mouse *Foxa2* gene. Reverse transcription–polymerase chain reaction products were cloned into the TOPO3 vector (Invitrogen) and sequenced.

Intrahypothalamic injections. *Foxa2T156A*$^{flox/flox}$ or C57Bl/6 mice were anaesthetized, and 150 nl of the recombinant adenovirus (Ad-GFP, Ad-Cre or Ad-Foxa2T156A[5]; 5×10^9 pfu μl^{-1}) were injected bilaterally in the lateral hypothalamus at a rate of 50 nl min^{-1}. The stereotactic coordinates for injection were determined from ref. 23 and were, relative to the bregma, anteroposterior by −0.4 mm, lateral by ±1.2 mm and dorsoventral by −5.2 mm. GFP expression in the lateral hypothalamus was confirmed by fluorescent microscopy.

Statistical analysis. Statistical analyses were done using a Student's *t*-test and the null hypothesis was rejected at the 0.05 level.

22. Nagy, A. Rossant, J. Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci. USA* **90**, 8424–8428 (1993).

23. Franklin, K. B. J&. Paxinos, G. *The Mouse Brain in Stereotaxic Coordinates* 3rd edn (Elsevier, 2007).

Vol 462 | 3 December 2009 | **doi:10.1038/nature08505**

Injury-induced mechanical hypersensitivity requires C-low threshold mechanoreceptors

Rebecca P. Seal[1]*, Xidao Wang[2]*, Yun Guan[3], Srinivasa N. Raja[3], C. Jeffery Woodbury[4], Allan I. Basbaum[2] & Robert H. Edwards[1]

Mechanical pain contributes to the morbidity associated with inflammation and trauma, but primary sensory neurons that convey the sensation of acute and persistent mechanical pain have not been identified. Dorsal root ganglion (DRG) neurons transmit sensory information to the spinal cord using the excitatory transmitter glutamate[1], a process that depends on glutamate transport into synaptic vesicles for regulated exocytotic release. Here we report that a small subset of cells in the DRG expresses the low abundance vesicular glutamate transporter VGLUT3 (also known as SLC17A8). In the dorsal horn of the spinal cord, these afferents project to lamina I and the innermost layer of lamina II, which has previously been implicated in persistent pain caused by injury[2]. Because the different VGLUT isoforms generally have a non-redundant pattern of expression[3], we used *Vglut3* knockout mice to assess the role of VGLUT3$^+$ primary afferents in the behavioural response to somatosensory input. The loss of VGLUT3 specifically impairs mechanical pain sensation, and in particular the mechanical hypersensitivity to normally innocuous stimuli that accompanies inflammation, nerve injury and trauma. Direct recording from VGLUT3$^+$ neurons in the DRG further identifies them as a poorly understood population of unmyelinated, low threshold mechanoreceptors (C-LTMRs)[4,5]. The analysis of *Vglut3*$^{-/-}$ mice now indicates a critical role for C-LTMRs in the mechanical hypersensitivity caused by injury.

In contrast to the more abundant isoforms VGLUT1 and 2, VGLUT3 is expressed by a small number of isolated neuronal populations[6]. Consistent with this general pattern, VGLUT1 and 2 occur at high levels throughout dorsal and ventral horns of the spinal cord[7,8], whereas VGLUT3 has a much more restricted distribution. In the dorsal horn, distinct bands within superficial laminae I and inner II immunostain specifically for VGLUT3 (Fig. 1a). Unilateral dorsal rhizotomy (L3–L5) eliminates VGLUT3 immunoreactivity in the ipsilateral, but not the contralateral dorsal horn, establishing the peripheral origin of these processes. Within inner lamina II, VGLUT3 immunoreactivity overlaps almost exclusively with a band of interneurons expressing the γ isoform of protein kinase C (PKC), and not with afferents that bind the lectin IB4 (Fig. 1b). This discrete projection suggests expression of the transporter by DRG neurons that convey specific sensory modalities.

To identify the DRG neurons expressing VGLUT3, we produced bacterial artificial chromosome (BAC) transgenic mice that express enhanced green fluorescent protein (EGFP) under the control of *Vglut3* regulatory sequences (VGLUT3 EGFP mice; Supplementary Fig. 1a) because the VGLUTs do not themselves generally localize to cell bodies. Consistent with the reliable expression from most BACs, VGLUT3 EGFP transgenic mice only label for EGFP in isolated cell populations in the brain known to express VGLUT3 (Supplementary Fig. 1b)[6]. In the spinal cord, EGFP expression resembles that of VGLUT3, with labelled fibres restricted to dorsal horn lamina I and the PKCγ layer of lamina II (Supplementary Fig. 2a), and unilateral L3–L5 dorsal rhizotomy eliminates EGFP immunoreactivity in the ipsilateral dorsal horn (Supplementary Fig. 2b). As EGFP in the BAC replaced *Vglut3* protein-coding sequences, we also double-stained for both EGFP and endogenous VGLUT3, and observed colocalization in the dorsal horn (Supplementary Fig. 3). Few if any EGFP$^+$ cell bodies are observed in the spinal cord of adult mice, consistent with the rare labelling for VGLUT3 by *in situ* hybridization (data not shown).

Taking advantage of the BAC transgenic mice, we examined the expression of VGLUT3 in adult sensory ganglia. Approximately 11% of trigeminal ganglion ($n = 205$ out of 1,916) and ~10% of L4/L5 DRG ($n = 237$ out of 2,260) neurons express EGFP, and these cells are small–medium in size (<30 μm; Fig. 1c), consistent with *in situ* hybridization for VGLUT3 in trigeminal ganglion (Fig. 1d). Furthermore, ~92% of EGFP$^+$ neurons express peripherin ($n = 207$ out of 226) (Fig. 1e), a marker of cells with unmyelinated axons. Indeed, very few (<1%) EGFP$^+$ cells ($n = 2$ out of 214) express the N52 antigen that labels DRG cells with myelinated axons. In general, peripherin$^+$ cells either bind the lectin IB4 or express neural peptides[9], but we find that only ~7% of EGFP$^+$ cells in DRG and trigeminal ganglion bind IB4 ($n = 19$ out of 279), and none express the peptides substance P or CGRP (Fig. 1e and data not shown). In addition, EGFP$^+$ cells do not express the primary heat sensor TRPV1 (Fig. 1e). VGLUT3$^+$ neurons thus represent a previously uncharacterized population of unmyelinated neurons that project to laminae I and II inner of the dorsal horn.

The expression of VGLUT3 by unmyelinated primary sensory neurons projecting to superficial layers of the dorsal horn suggested that these cells might convey pain or temperature sensation. To assess their role, we used mice lacking VGLUT3[10,11]. These animals exhibit sensorineural deafness and rare non-convulsive seizures, but are otherwise normal in terms of motor coordination, vision and mating behaviour, consistent with the highly restricted distribution of VGLUT3. Notably, VGLUT1 and 2 do not colocalize with VGLUT3 in spinal cord terminals (Supplementary Fig. 3), indicating that VGLUT3$^+$ afferents should not transmit any glutamatergic signal to the spinal cord in the absence of VGLUT3.

Vglut3$^{-/-}$ mice do not differ from wild-type littermates in their sensitivity to cold or heat (Fig. 2a, b). Furthermore, we observed no difference in the response to formalin. Because formalin acts through TRPA1 receptors[12] that are expressed by a subset of TRPV1$^+$ neurons (Fig. 2c), this result is consistent with the expression of VGLUT3 by sensory neurons distinct from those expressing TRPV1 (Fig. 1e). *Vglut3*$^{-/-}$ mice also do not differ from wild-type littermates in the threshold for withdrawal from stimulation with von Frey hairs of

---

[1] Departments of Physiology and Neurology, University of California, San Francisco School of Medicine, California 94143, USA. [2] Departments of Anatomy and Physiology, University of California, San Francisco School of Medicine, California 94158, USA. [3] Department of Anesthesiology and Critical Care Medicine, the Johns Hopkins University, School of Medicine, Baltimore, Maryland 21205, USA. [4] Department of Zoology and Physiology, University of Wyoming, Laramie, Wyoming 82071, USA.

*These authors contributed equally to the work.

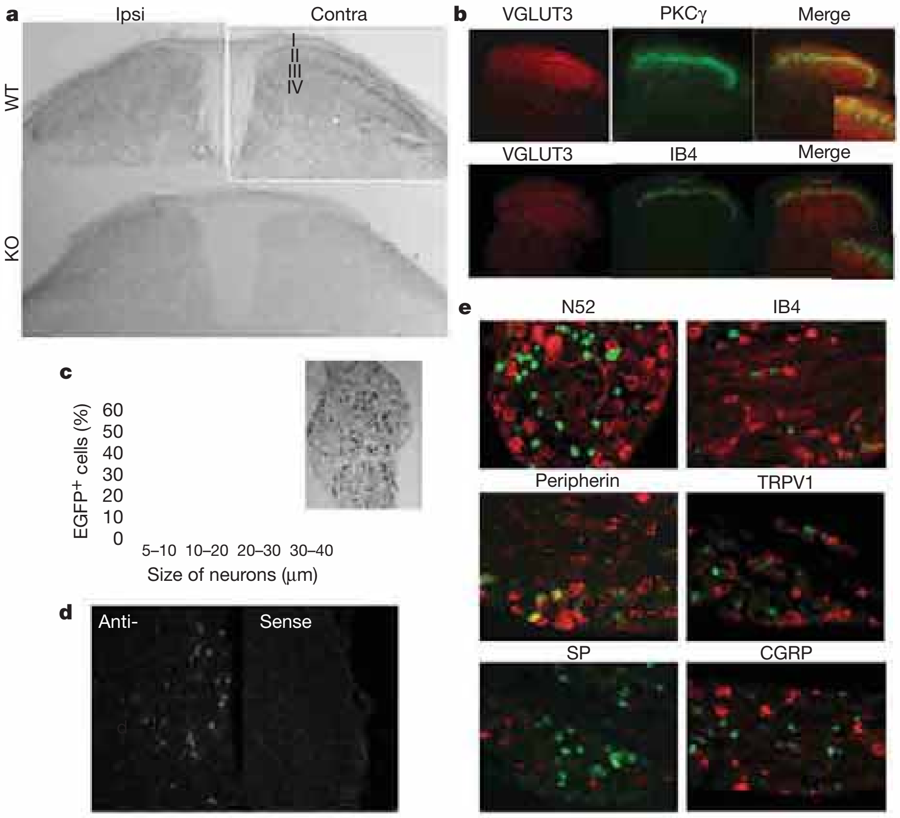


Figure 1 | VGLUT3 is expressed by a unique subset of small- and medium-sized DRG neurons that project to dorsal horn lamina I and the inner part of lamina II. a, Immunostaining shows VGLUT3 in laminae I and innermost II (arrowheads) of the dorsal horn in wild-type (WT) but not *Vglut3*$^{-/-}$ knockout (KO) mice. Dorsal rhizotomy abolishes VGLUT3 immunoreactivity in ipsilateral (ipsi) laminae I and II. Contra, contralateral. Scale bar, 50 μm. **b**, VGLUT3 immunofluorescence (red) in the dorsal horn overlaps to a large extent with interneurons that express PKCγ (green), but little if at all with the binding to IB4 (green). Scale bar, 50 μm. **c**, Immunoperoxidase labelling of a DRG section from a VGLUT3 EGFP mouse shows the size and distribution of EGFP$^+$ cells. Original magnification, ×20. **d**, *In situ* hybridization for VGLUT3 labels small- and medium-sized neurons in the trigeminal ganglion of wild-type mice. Hybridization with a sense probe confers no detectable signal. Scale bar, 30 μm. **e**, EGFP$^+$ DRG neurons (green) colocalize strongly with peripherin (92%), to a small extent with IB4 (7%) but rarely with N52 and not at all with TRPV1, substance P (SP) or CGRP. Arrowheads indicate colocalization. Scale bar, 30 μm.

increasing strength, suggesting normal sensitivity to light noxious mechanical stimuli (Fig. 2d). However, the knockout mice do show a significantly delayed response to the placement of a clip on the tail (Fig. 2e), as well as an increased threshold for tail withdrawal measured in the Randall–Selitto apparatus (Fig. 2f). Thus, in contrast to their normal temperature sensation, formalin sensitivity and von Frey hair threshold, VGLUT3 knockout mice show a selective albeit partial defect in the response to intense noxious mechanical stimuli.

To assess further the contribution of VGLUT3$^+$ primary afferents to acute mechanical pain, we recorded from dorsal horn wide dynamic range (WDR) neurons, which respond to innocuous as well as noxious mechanical stimuli[13]. WDR neurons from both wild-type and knockout mice fire at progressively higher frequencies in response to increasing mechanical force applied to the hindpaw (Fig. 2g). At the highest forces applied (6*g* and 9*g*), however, WDR neurons of *Vglut3*$^{-/-}$ mice show significantly less firing than those of wild-type animals. Furthermore, WDR neurons of knockout mice show significantly less response to pinch (a noxious stimulus), but do not differ in their response to brush (an innocuous stimulus) (Fig. 2h). The response of WDR neurons to mechanical stimuli thus depends on VGLUT3 only for the strongest mechanical stimuli, correlating with the specific behavioural defect of the knockout mice in acute mechanical pain sensation.

Because VGLUT3 seems to be important for acute mechanical pain, we also assessed the role of VGLUT3$^+$ primary afferents in persistent hypersensitivity states caused by inflammation, nerve injury and trauma. In a model of chronic, inflammatory pain produced by injection of carrageenan into the hindpaw[14], von Frey threshold drops markedly in wild-type mice, but remains unchanged in the knockout (Fig. 3a). However, knockout and wild-type mice show the same increase in sensitivity to radiant heat. *Vglut3*$^{-/-}$ mice thus exhibit a selective defect in mechanical hypersensitivity in this model of inflammatory pain. Using the model of spared nerve injury (SNI)[15] to assess neuropathic pain, wild-type animals develop a profound mechanical hypersensitivity ipsilateral to the SNI, whereas *Vglut3*$^{-/-}$ mice again show a much smaller reduction in mechanical threshold (Fig. 3b). In the Brennan model of post-surgical pain[16], knockout mice also exhibit a much smaller reduction in mechanical threshold than wild-type (Fig. 3c), and the frequency of paw withdrawal to light and intermediate von Frey stimuli is reduced (Supplementary Fig. 4). However, the two genotypes do not differ in thermal hyperalgesia (Fig. 3c). The loss of VGLUT3 thus causes a remarkably modality-specific defect in the mechanical hypersensitivity associated with inflammatory, neuropathic and post-operative pain.

The importance of VGLUT3$^+$ neurons for persistent mechanical pain produced by tissue injury may reflect their role in the induction of central sensitization, or in the expression of the sensitized state, such as enhancing or transmitting the sensation of pain provoked by an innocuous mechanical stimulus. To distinguish between these two possibilities, we used the model of capsaicin-induced secondary mechanical hyperalgesia. Capsaicin directly activates TRPV1$^+$ afferents that convey temperature sensation[17], but when injected into the ankle, produces a mechanical hypersensitivity on the adjacent plantar hindpaw[18]. Because TRPV1$^+$ afferents do not express VGLUT3 (Fig. 1e), and *Vglut3*$^{-/-}$ mice show a normal response to the acute injection of capsaicin (Supplementary Fig. 5c), the loss of VGLUT3 should not interfere with induction of the sensitized state, enabling us to assess the role of VGLUT3$^+$ afferents in the detection of innocuous mechanical stimuli as painful. As shown in Fig. 3d, the mechanical thresholds of wild-type animals drop markedly for at least 1 h after capsaicin injection, but the knockout mice show only a modest change. Because VGLUT3$^+$ afferents do not contribute to central sensitization in this model of tissue injury, they must be involved in transmitting the input that produces pain.

To characterize the properties of VGLUT3$^+$ cells, we recorded intracellularly from EGFP$^+$ DRG neurons of the BAC transgenic mice in an *ex vivo* preparation with the thoracic cord, DRG, nerves and skin intact[19]. With the exception of staining in the intermediolateral region of the thoracic cord, EGFP (and VGLUT3) expression appears similar at thoracic and lumbar levels (Supplementary Fig. 6). EGFP$^+$ cells exhibit large amplitude overshooting somal spikes, with

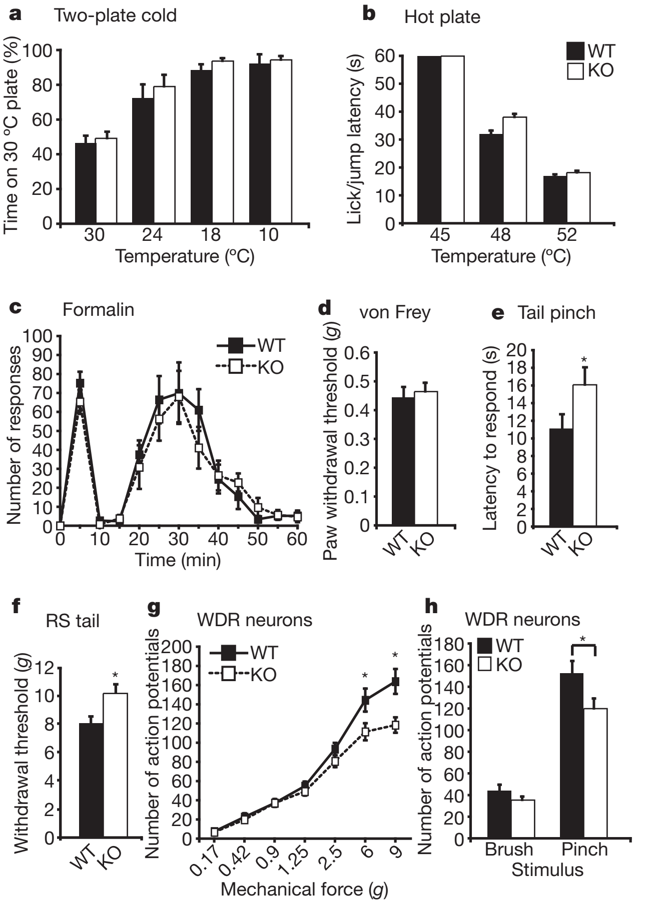


Figure 2 | *Vglut3*$^{-/-}$ mice show a selective defect in acute mechanical pain sensation to intense noxious stimuli. **a**, Using a two-plate preference chamber to assess cold sensitivity, *Vglut3*$^{-/-}$ and wild-type mice show the same preference for the 30 °C plate over 24, 18 or 10 °C plates ($n = 6$ mice for both genotypes). **b**, Using a hot plate to assess heat sensitivity, the latency of *Vglut3*$^{-/-}$ mice to exhibit nocifensive behaviours is similar to that of wild-type littermates at both 48 and 52 °C ($n = 8$). **c**, Both genotypes show a similar, biphasic response to 2% formalin injected into the hindpaw ($n \geq 15$). **d**, Mechanical thresholds measured using von Frey filaments do not differ between genotypes ($n \geq 10$). **e**, Measuring the acute mechanical pain sensitivity to tail clip, *Vglut3*$^{-/-}$ mice show a significantly longer latency to respond than wild-type (16 ± 2 versus 11 ± 1; $n \geq 17$), indicating a reduced sensitivity to acute mechanical pain. $^*P < 0.05$, Mann–Whitney U test. **f**, *Vglut3*$^{-/-}$ mice show a higher threshold than wild-type (10.2 ± 0.5 versus 8.1 ± 0.4 ; $n = 9$) to tail withdrawal on compression in the Randall–Selitto (RS) test. $^*P < 0.05$, Mann–Whitney U test. **g**, Single-unit extracellular recordings in the dorsal horn of the lumbar spinal cord show the response of WDR neurons to graded punctate mechanical stimulation of the plantar hindpaw receptive field. The genotypes differ specifically in response to 6 and 9 *g*. $^*P < 0.05$ by Fisher's LSD post-hoc test, $n = 50$–53 cells. **h**, The same group of WDR neurons in *Vglut3*$^{-/-}$ mice respond significantly less than those in wild-type to pinch ($^*P = 0.026$; Student's *t*-test), but not brush stimuli ($P > 0.05$; Student's *t*-test) applied to the receptive field for 3 s. $n = 10$ mice for both genotypes (**g**, **h**). Error bars indicate s.e.m.

an inflection on the falling phase, and an average conduction velocity of 0.6 ± 0.1 m s^{-1} ($n = 28$) (Fig. 4a, b and Supplementary Fig. 7a), consistent with the lack of myelination suggested by their immunoreactivity for peripherin and not N52 (Fig. 1e and Supplementary Fig. 6b). Of 28 cells characterized, all were exquisitely sensitive to the finest von Frey hair available (mechanical threshold ≤ 0.07 mN; Fig. 4c), responded better to slowly than rapidly moving tactile stimuli (Supplementary Fig. 7b), and adapted to stationary stimuli (Fig. 4c, d). Receptive fields were small (≤ 1 mm^2) and composed of 1–3 discrete spots. Using a series of controlled forces, we observed a trend to higher instantaneous firing rates for stronger stimuli (Fig. 4e and Supplementary Fig. 7c), with a significantly larger response to 200 mN

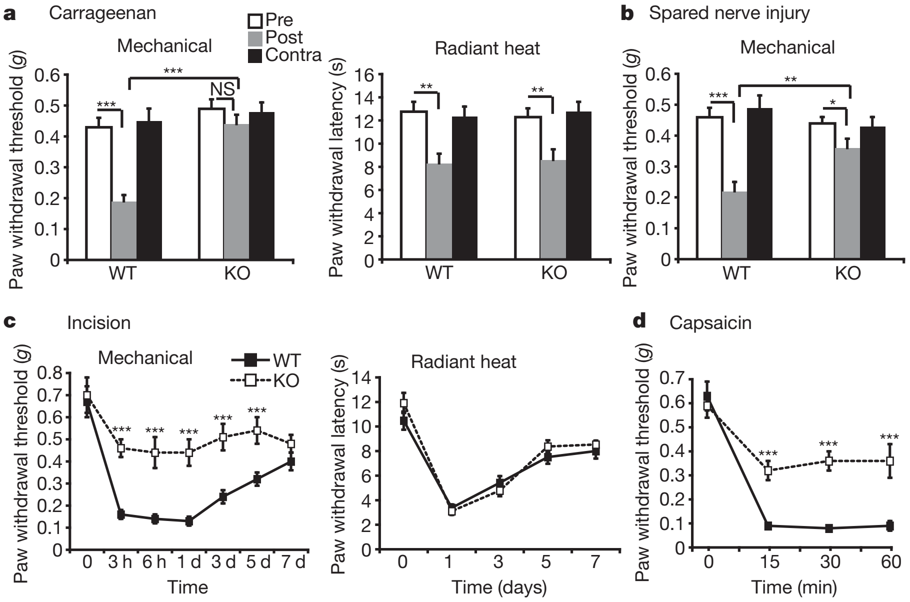


Figure 3 | *Vglut3*$^{-/-}$ mice show a profound, selective defect in the mechanical hypersensitivity produced by inflammation, nerve injury and trauma. **a**, *Vglut3*$^{-/-}$ mice and wild-type littermates were injected with carrageenan in the hindpaw. Tested 24 h later, both genotypes show a significant increase in sensitivity to radiant heat, but only wild-type mice show a significant decrease in mechanical threshold ($n \geq 10$ mice). NS, not significant. **b**, In the spared nerve injury model of neuropathic pain, wild-type mice exhibit a significant decrease in mechanical threshold (52%) 3 days after sciatic nerve section, but knockout animals exhibit only a modest reduction (18%) ($n \geq 9$ for both genotypes). $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ by one-way ANOVA with Tukey's honestly significant difference (HSD) post-hoc analysis for **a** and **b**. **c**, In wild-type mice, an incision made in the plantar hindpaw produces a robust decrease in mechanical threshold, but knockout mice show only a mild reduction for up to 5 days (d). Response to radiant heat does not differ between genotypes at all times tested ($n = 6$). **d**, In wild-type mice, capsaicin injection into the ankle produces a profound decrease in mechanical threshold measured in the hindpaw at 15, 30 and 60 min after injection, whereas knockout mice show only a modest decrease ($n = 7$). $^{***}P < 0.001$, one-way ANOVA with Newman-Keuls post-hoc analysis for **c** and **d**. Error bars indicate s.e.m.

than 5–25 mN ($P < 0.01$, $n = 5$–9), and to 50 and 100 mN than 5 mN ($P < 0.05$). Thermal sensitivity was assessed in a subset of the neurons, and all of these responded to cold (4 °C, $n = 28$) but not hot (52 °C, $n = 3$) solution. Controlled temperature ramps (from 31.5 °C to −4 °C) revealed a threshold for cold stimuli of 25 ± 1 °C ($n = 7$), whereas ramps from 31.5 to 50 °C elicited no response (Fig. 4f). Thus, VGLUT3 expression identifies a class of unmyelinated, low threshold mechanoreceptors (C-LTMRs)[4,5]. Reconstruction of central projections from seven of these cells injected with neurobiotin showed small, dense longitudinal columns extending 150–200 μm rostrocaudally in lamina II inner (Fig. 4g–i and Supplementary Fig. 8), with additional minor terminals in laminae III and I, consistent with the immunoreactivity for VGLUT3. Previous work has conflicted on the expression of IB4 by C-LTMRs[19,20], and a small number of EGFP$^+$ DRG neurons label with IB4 (Fig. 1e), but none of 25 physiologically identified C-LTMRs bound this lectin (Fig. 4j).

The results show that primary sensory neurons expressing VGLUT3 convey two distinct features of mechanical pain sensation. First, behavioural analysis of the knockout indicates that VGLUT3$^+$ cells contribute to acute pain in response to intense noxious mechanical stimuli, and recordings from WDR neurons support this role. Although C-LTMRs are very sensitive to low intensity mechanical stimulation, they do show a significantly greater response to high intensity stimuli, and may therefore participate in acute mechanical pain sensation as well. It nonetheless remains possible that other, less abundant VGLUT3$^+$ neurons contribute to the acute sensory phenotype observed in the knockout. Because nociceptive inputs from both high-threshold myelinated $A\delta$ and unmyelinated C-fibres terminate in lamina I (ref. 21), the acute mechanical pain sensation conveyed by VGLUT3$^+$ neurons may involve the projections to lamina I.

Second, VGLUT3$^+$ neurons contribute to the mechanical hypersensitivity that occurs in the setting of tissue or nerve injury. Previous work has described several mechanisms that may account for the hypersensitivity to innocuous stimuli that develops after injury. In models of inflammation, unmyelinated polymodal sensory neurons develop increased sensitivity to mechanical stimuli[22]. Large myelinated $A\beta$ neurons have also been implicated in chronic pain[23]: after injury, $A\beta$ neurons begin to express substance P, which increases the excitation of spinal cord neurons[24]. In contrast, the selective defect in mechanical hypersensitivity of the *Vglut3*$^{-/-}$, together with the sensitivity of C-LTMRs to light touch and the identification of most VGLUT3$^+$ afferents as C-LTMRs, now indicate that C-LTMRs are required to generate pain in response to innocuous mechanical stimuli. The loss of capsaicin-induced secondary mechanical hyperalgesia in the knockout further supports a role for C-LTMRs in expression rather than induction of the sensitized state. Consistent with this role,

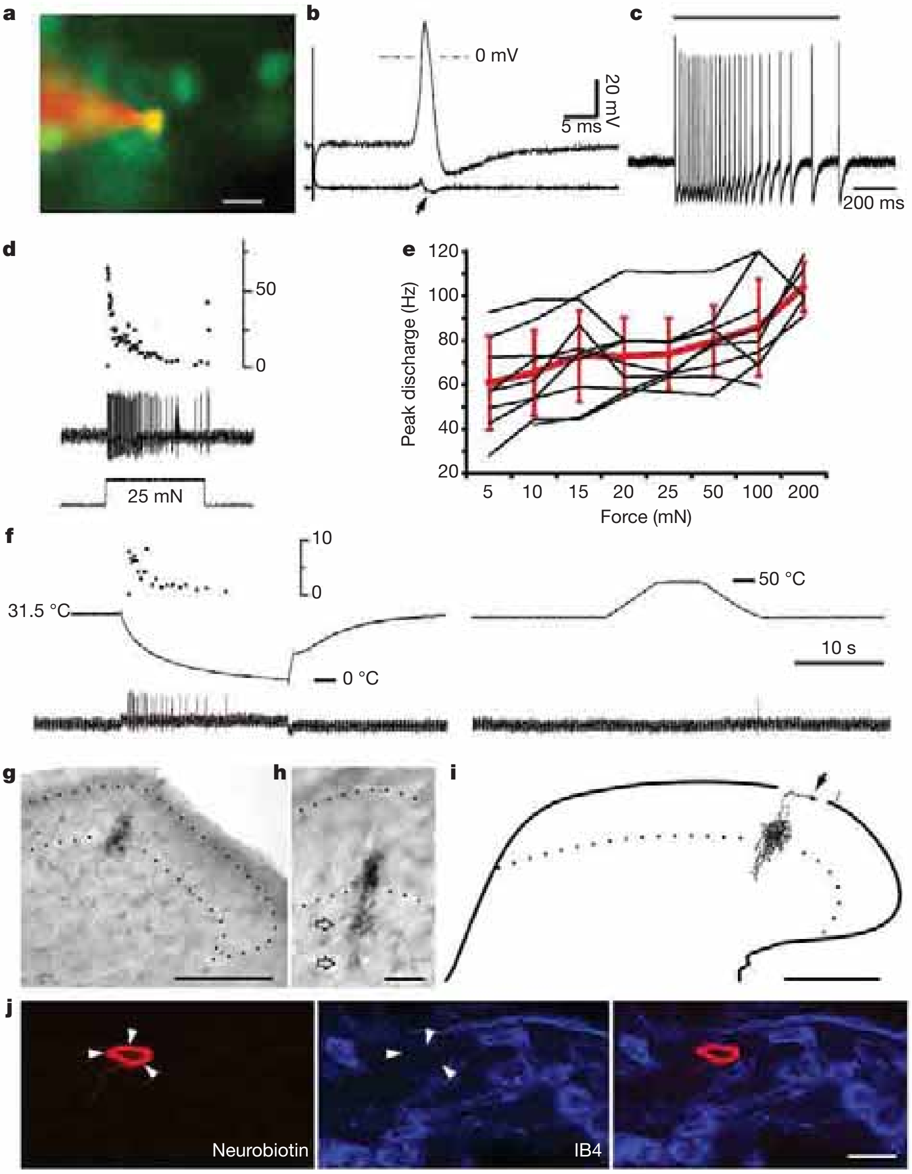


Figure 4 | VGLUT3 expression in DRG uniquely identifies C-LTMRs. **a**, An EGFP$^+$ cell is stained by Alexa 555 from an intracellular microelectrode (red) while recording from an *ex vivo* somatosensory preparation. **b**, Electrical stimulation of the dorsal cutaneous nerve evokes a high amplitude, overshooting action potential. The first derivative (bottom trace) reveals an inflected falling phase (arrow) typical for C-fibre spikes. **c**, Application of a 0.07-mN von Frey filament to the skin receptive field (duration indicated by bar) elicits a burst of spikes. **d**, Response to 25 mN force applied with a feedback-controlled stimulator (bottom trace) shows adaptation in the continued presence of the force (middle trace). Top trace shows the instantaneous frequency of spikes. **e**, Peak discharge rate of individual EGFP$^+$ neurons (black lines) to increasing force (5–200 mN) shows a small increase with stronger stimuli ($P < 0.01$ for 5–25 mN versus 200 mN by ANOVA with Bonferroni post-hoc correction, $n = 5$–9); red bars indicate mean ± s.d. **f**, Temperature ramps performed with a Peltier device (middle trace) evoke firing in an EGFP$^+$ neuron (bottom trace) on cooling (left) but not heating (right), Top trace (left) shows the instantaneous firing rate. **g**, Low-power photomicrograph of central terminals arborizing in lamina IIi; dotted lines indicate the substantia gelatinosa. **h**, Photomicrograph of different section at higher power shows arborization extending into lamina III (open arrows). **i**, Camera lucida reconstruction of arborization across four serial 50-μm sections, showing further input to lamina I (arrow); the ventral border of the substantia gelatinosa is indicated by a dotted line, and the border between grey and white matter by a solid line. **j**, Staining of a DRG section for IB4 (blue, middle) shows that the neurobiotin-labelled cell (red, left) is IB4-negative. Scale bars, 20 μm (**a**, **h**, **j**) and 100 μm (**g**, **i**). Physiological calibration: 20 mV and 5 ms (**b**), 200 ms (**c**).

VGLUT3$^+$ C-LTMRs terminate in the inner layer of lamina II, which does not receive acute nociceptive input[25–27], but rather contains interneurons expressing PKCγ that, like VGLUT3, are required for injury-associated mechanical hypersenstivity[2]. As the ablation of DRG neurons expressing the tetrodotoxin-insensitive Na$^+$ channel Nav1.8 produces a similar phenotype[28], VGLUT3$^+$ neurons may also express this channel.

In the absence of injury, we are unable to detect a behavioural role for VGLUT3 in the transmission of innocuous mechanical information. This probably reflects the presence of large, myelinated fibres that also convey the sensation of light touch. Human psychophysical studies have implicated C-LTMRs specifically in acute pleasant touch[29], but this behaviour would be difficult to assess in rodents. The expression of mechanical hypersensitivity after injury may thus involve a change in the sensation conveyed by C-LTMRs, from pleasant touch to pain.

METHODS SUMMARY

Guinea-pig antibodies to VGLUT3 and TRPV1 were generous gifts from H. Hioki and T. Kaneko (VGLUT3) and D. Julius (TRPV1).

Mouse genetics. *Vglut3*$^{-/-}$ mice and wild-type littermates were backcrossed to C57Bl/6 for at least eight generations. One-to-three-month-old animals were used for histology, behaviour and electrophysiology. The VGLUT3 EGFP BAC transgene was generated by homologous recombination in bacteria as previously described[30]. The EGFP complementary DNA and SV40 polyadenylation sequences were amplified by PCR from IRES2–EGFP (BD Biosciences) and inserted into BAC RP2488H21 that contains the mouse *Vglut3* gene (Supplementary Fig. 1a). Fertilized B6SJL oocytes were injected with purified VGLUT3 EGFP BAC DNA, implanted into pseudopregnant females and brain slices from transgenic offspring screened by immunocytochemistry for transgene expression.

Electrophysiology. Single-unit extracellular recordings were made from single WDR neurons in the L4–L5 dorsal horn in response to graded von Frey monofilaments, brush and pinch stimuli, each applied for 3 s to the plantar hindpaw receptive field. The experimenter was blinded to genotype, and statistical analysis was performed by two-way mixed-model analysis of variance (ANOVA) with Fisher's protected least significant difference (LSD) post-hoc test or Student *t*-test. Intracellular recordings were made from EGFP$^+$ neurons in an *ex vivo* preparation with dorsal skin, nerve and spinal cord intact[19]. Action potentials evoked with a suction electrode on the nerve bundle were used to calculate conduction velocity. To determine the response properties of recorded neurons, calibrated von Frey hairs, feedback-controlled mechanical stimulator, artificial cerebrospinal fluid solutions (warm or cold) and feedback-controlled Peltier stimulator were applied to the skin receptive field. Full methods are available in the Supplementary Information.

Received 14 May; accepted 17 September 2009.
Published online 15 November; corrected 3 December 2009 (see full-text HTML version for details).

1. Yoshimura, M. & Jessell, T. Amino acid-mediated EPSPs at primary afferent synapses with substantia gelatinosa neurones in the rat spinal cord. *J. Physiol. (Lond.)* **430**, 315–335 (1990).
2. Malmberg, A. B., Chen, C., Tonegawa, S. & Basbaum, A. I. Preserved acute pain and reduced neuropathic pain in mice lacking PKCγ. *Science* **278**, 279–283 (1997).
3. Fremeau, R. T. Jr, Voglmaier, S., Seal, R. P. & Edwards, R. H. VGLUTs define subsets of excitatory neurons and suggest novel roles for glutamate. *Trends Neurosci.* **27**, 98–103 (2004).
4. Bessou, P. & Perl, E. R. Response of cutaneous sensory units with unmyelinated fibers to noxious stimuli. *J. Neurophysiol.* **32**, 1025–1043 (1969).
5. Iggo, A. & Kornhuber, H. H. A quantitative analysis of non-myelinated cutaneous mechano-receptors. *J. Physiol.* **198**, 113passim (1968).
6. Gras, C. *et al.* A third vesicular glutamate transporter expressed by cholinergic and serotoninergic neurons. *J. Neurosci.* **22**, 5442–5451 (2002).
7. Todd, A. J. *et al.* The expression of vesicular glutamate transporters VGLUT1 and VGLUT2 in neurochemically defined axonal populations in the rat spinal cord with emphasis on the dorsal horn. *Eur. J. Neurosci.* **17**, 13–27 (2003).
8. Li, J. L., Fujiyama, F., Kaneko, T. & Mizuno, N. Expression of vesicular glutamate transporters, VGluT1 and VGluT2, in axon terminals of nociceptive primary afferent fibers in the superficial layers of the medullary and spinal dorsal horns of the rat. *J. Comp. Neurol.* **457**, 236–249 (2003).
9. Snider, W. D. & McMahon, S. B. Tackling pain at the source: new ideas about nociceptors. *Neuron* **20**, 629–632 (1998).
10. Seal, R. P. *et al.* Sensorineural deafness and seizures in mice lacking vesicular glutamate transporter 3. *Neuron* **57**, 263–275 (2008).
11. Gras, C. *et al.* The vesicular glutamate transporter VGLUT3 synergizes striatal acetylcholine tone. *Nature Neurosci.* **11**, 292–300 (2008).
12. McNamara, C. R. *et al.* TRPA1 mediates formalin-induced pain. *Proc. Natl Acad. Sci. USA* **104**, 13525–13530 (2007).
13. Chung, J. M., Kenshalo, D. R. Jr, Gerhart, K. D. & Willis, W. D. Excitation of primate spinothalamic neurons by cutaneous C-fiber volleys. *J. Neurophysiol.* **42**, 1354–1369 (1979).
14. Kayser, V. & Guilbaud, G. Local and remote modifications of nociceptive sensitivity during carrageenin-induced inflammation in the rat. *Pain* **28**, 99–107 (1987).
15. Shields, S. D., Eckert, W. A. III & Basbaum, A. I. Spared nerve injury model of neuropathic pain in the mouse: a behavioral and anatomic analysis. *J. Pain* **4**, 465–470 (2003).
16. Brennan, T. J., Vandermeulen, E. P. & Gebhart, G. F. Characterization of a rat model of incisional pain. *Pain* **64**, 493–502 (1996).
17. Caterina, M. J. *et al.* Impaired nociception and pain sensation in mice lacking the capsaicin receptor. *Science* **288**, 306–313 (2000).
18. Treede, R. D., Meyer, R. A., Raja, S. N. & Campbell, J. N. Peripheral and central mechanisms of cutaneous hyperalgesia. *Prog. Neurobiol.* **38**, 397–421 (1992).
19. Albers, K. M., Woodbury, C. J., Ritter, A. M., Davis, B. M. & Koerber, H. R. Glial cell-line-derived neurotrophic factor expression in skin alters the mechanical sensitivity of cutaneous nociceptors. *J. Neurosci.* **26**, 2981–2990 (2006).
20. Liu, Q. *et al.* Molecular genetic visualization of a rare subset of unmyelinated sensory neurons that may detect gentle touch. *Nature Neurosci.* **10**, 946–948 (2007).
21. Christensen, B. N. & Perl, E. R. Spinal neurons specifically excited by noxious or thermal stimuli: marginal zone of the dorsal horn. *J. Neurophysiol.* **33**, 293–307 (1970).
22. Schaible, H. G., Ebersberger, A. & Von Banchet, G. S. Mechanisms of pain in arthritis. *Ann. NY Acad. Sci.* **966**, 343–354 (2002).
23. Woolf, C. J. & Doubell, T. P. The pathophysiology of chronic pain–increased sensitivity to low threshold A beta-fibre inputs. *Curr. Opin. Neurobiol.* **4**, 525–534 (1994).
24. Neumann, S., Doubell, T. P., Leslie, T. & Woolf, C. J. Inflammatory pain hypersensitivity mediated by phenotypic switch in myelinated primary sensory neurons. *Nature* **384**, 360–364 (1996).
25. Light, A. R., Trevino, D. L. & Perl, E. R. Morphological features of functionally defined neurons in the marginal zone and substantia gelatinosa of the spinal dorsal horn. *J. Comp. Neurol.* **186**, 151–171 (1979).
26. Boada, M. D. & Woodbury, C. J. Myelinated skin sensory neurons project extensively throughout adult mouse substantia gelatinosa. *J. Neurosci.* **28**, 2006–2014 (2008).
27. Neumann, S., Braz, J. M., Skinner, K., Llewellyn-Smith, I. J. & Basbaum, A. I. Innocuous, not noxious, input activates PKCγ interneurons of the spinal dorsal horn via myelinated afferent fibers. *J. Neurosci.* **28**, 7936–7944 (2008).
28. Abrahamsen, B. *et al.* The cell and molecular basis of mechanical, cold, and inflammatory pain. *Science* **321**, 702–705 (2008).
29. Löken, L. S., Wessberg, J., Morrison, I., McGlone, F. & Olausson, H. Coding of pleasant touch by unmyelinated afferents in humans. *Nature Neurosci.* **12**, 547–548 (2009).
30. Yang, X. W., Model, P. & Heintz, N. Homologous recombination based modification in *Escherichia coli* and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nature Biotechnol.* **15**, 859–865 (1997).

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

Acknowledgements We thank D. Bautista, D. Julius and members of the Basbaum and Edwards laboratories for advice, and D. Woodbury and C. Cassidy from the Woodbury laboratory for help with experiments. The work was supported by the Johns Hopkins Blaustein Pain Research Fund (Y.G.) and by grants from NARSAD (R.P.S., R.H.E.) and NIH (R.P.S., Y.G., S.N.R., C.J.W., A.I.B. and R.H.E.).

Author Contributions R.P.S. created the *Vglut3*$^{-/-}$ and VGLUT3 EGFP BAC transgenic mice; X.W. and R.P.S. performed behaviour and histology experiments; Y.G. performed *in vivo* spinal cord recordings; and C.J.W. and R.P.S. performed *ex vivo* recordings. R.H.E., R.P.S. and A.I.B. wrote the manuscript with contributions from C.J.W., Y.G., S.N.R. and X.W.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.H.E. (robert.edwards@ucsf.edu).

nature Vol 462|3 December 2009|doi:10.1038/nature08586

LETTERS

Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis

Zasha Weinberg[1,2], Jonathan Perreault[2], Michelle M. Meyer[2] & Ronald R. Breaker[1,2,3]

Estimates of the total number of bacterial species[1–3] indicate that existing DNA sequence databases carry only a tiny fraction of the total amount of DNA sequence space represented by this division of life. Indeed, environmental DNA samples have been shown to encode many previously unknown classes of proteins[4] and RNAs[5]. Bioinformatics searches[6–10] of genomic DNA from bacteria commonly identify new noncoding RNAs (ncRNAs)[10–12] such as riboswitches[13,14]. In rare instances, RNAs that exhibit more extensive sequence and structural conservation across a wide range of bacteria are encountered[15,16]. Given that large structured RNAs are known to carry out complex biochemical functions such as protein synthesis and RNA processing reactions, identifying more RNAs of great size and intricate structure is likely to reveal additional biochemical functions that can be achieved by RNA. We applied an updated computational pipeline[17] to discover ncRNAs that rival the known large ribozymes in size and structural complexity or that are among the most abundant RNAs in bacteria that encode them. These RNAs would have been difficult or impossible to detect without examining environmental DNA sequences, indicating that numerous RNAs with extraordinary size, structural complexity, or other exceptional characteristics remain to be discovered in unexplored sequence space.

Conserved secondary structures of RNAs can be identified by phylogenetic comparative sequence analysis[18,19], whereby nucleotides and structures important for RNA function are revealed by identification of conserved sequences and nucleotide covariation (for example, see Supplementary Fig. 1). We used this approach to identify over 75 new structured RNAs from bacteria or archaea. Among these are new riboswitch classes that sense tetrahydrofolate, *S*-adenosylhomocysteine and *S*-adenosylmethionine, and c-di-GMP, and other candidate *cis*-regulatory and ncRNAs (unpublished data). On the basis of available sequence data, several of these RNAs are present only in specific environments or in phyla with few available genome sequences (Supplementary Table 1). Here we report a special subset of new-found RNA structures that are exceptional, either because they are extremely large and structurally complex or because they are produced in unusually high amounts.

We identified two RNA structures (GOLLD and HEARO) that are among the largest complex bacterial ncRNAs known (Fig. 1). GOLLD (Giant, Ornate, Lake- and Lactobacillales-Derived) RNA is particularly striking because it represents the third-largest highly structured bacterial RNA discovered to date, ranking only behind 23S and 16S rRNAs. The structural complexity of GOLLD RNA (Fig. 2a), as quantified by the number of multistem junctions and pseudoknots, is similar to most self-splicing group II introns[20]. Also, as observed in large ribozymes[18–20], some regions of GOLLD RNA can adopt a diversity of complex folds (Supplementary Fig. 2).

We identified GOLLD RNAs by searching environmental sequences collected from Lake Gatún, Panama[21], and representatives were subsequently identified in eight cultivated organisms distributed among three bacterial phyla. GOLLD RNAs are frequently located adjacent to tRNAs, and in three cases, a tRNA is predicted inside a variable region in GOLLD RNA itself (Fig. 2a and Supplementary Discussion).

In *Lactobacillus brevis* ATCC 367 and other organisms, GOLLD RNA resides in an apparent prophage. We therefore monitored GOLLD RNA transcription in *L. brevis* cultures grown with mitomycin C, an antibiotic that commonly induces prophages to lyse their hosts[22]. Increased GOLLD RNA expression correlates with bacteriophage particle production, and DNA corresponding to the GOLLD RNA gene is packaged into phage particles (Fig. 2b). Furthermore, most *L. brevis* GOLLD RNA transcripts made during bacteriophage production closely bracket the entire span of conserved sequences and structural elements as determined by mapping of the 5′ and 3′ termini (Supplementary Fig. 3). Thus, expression of the entire noncoding RNA presumably is important for the bacteriophage lytic process.

HEARO (HNH Endonuclease-Associated RNA and ORF) RNAs (Fig. 3a) often carry an embedded ORF that usually is predicted to code for an HNH endonuclease. This enzyme is commonly exploited by a variety of mobile genetic elements to achieve DNA transposition[23]. Thus HEARO RNA and its associated ORF together might constitute a mobile genetic element. The number of HEARO RNAs

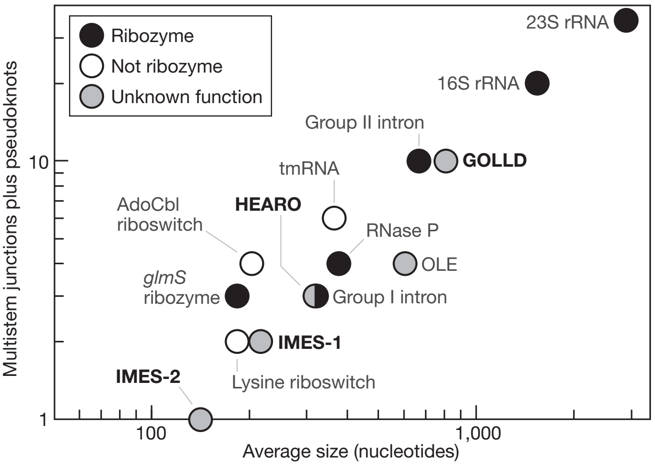


Figure 1 | Size and structural complexity of new-found RNAs compared to the ten largest known bacterial ncRNAs with complex structures. Structural complexity is represented by the number of multistem junctions plus pseudoknots (see Methods for details). RNAs described in this report are in bold type. HEARO and Group I ribozyme symbols overlap. Narrowly distributed RNAs (present in only one bacterial class) are not included.

---

[1]Howard Hughes Medical Institute, [2]Department of Molecular, Cellular and Developmental Biology, [3]Department of Molecular Biophysics and Biochemistry, Yale University, Box 208103, New Haven, Connecticut 06520-8103, USA.

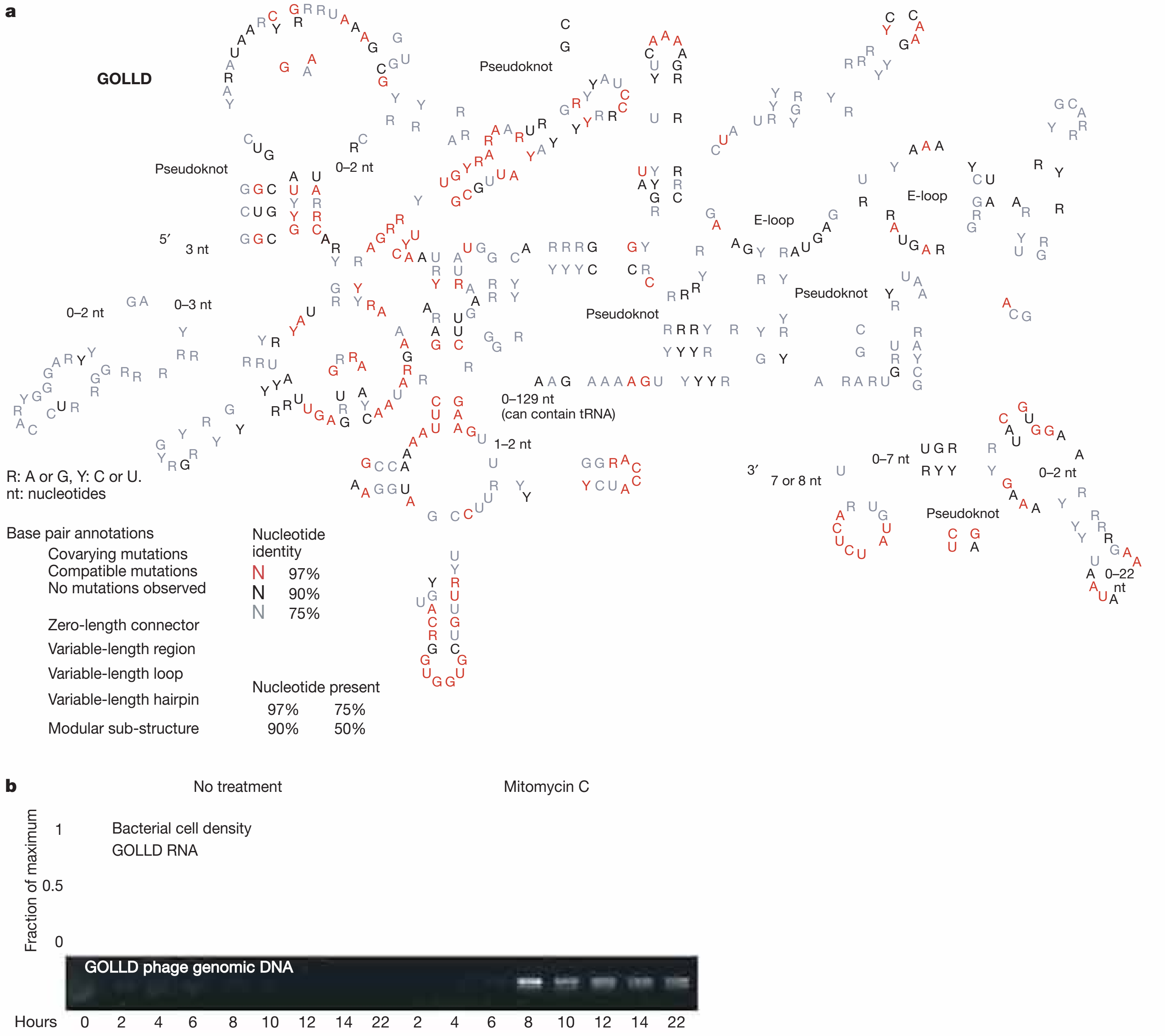


Figure 2 | GOLLD RNAs. a, Simplified consensus sequence and secondary structure model for the most common architecture of GOLLD RNAs. Annotated 5′ and 3′ ends reflect *L. brevis* transcripts observed by RACE experiments (Supplementary Fig. 3). **b**, Phage induction and expression of GOLLD RNA. Experimental details are presented in the Methods.

encoded by bacterial genomes varies widely. A total of 42 HEARO RNAs are predicted in *Arthrospira maxima* CS-328 (Supplementary Data), and most of these RNAs seem to represent recent duplications (Supplementary Fig. 4). When *A. maxima* HEARO sequences are aligned, it is apparent that the elements are highly conserved in sequence, whereas their flanking sequences show no conservation (Supplementary Fig. 5).

In some instances, homologues of the sequences flanking the consensus sequence can be identified in related bacterial species wherein the HEARO element is absent. These observations allow us to map putative integration events (Fig. 3b, Supplementary Fig. 6), which are consistent with a requirement for integration immediately upstream of the sequence ATGA or GTGA. Self-splicing group I and group II introns frequently carry ORFs coding for endonucleases, and the combined action of the protein enzyme and ribozyme components permit transposition with a reduced chance for genetic disruption at the integration site[23,24]. The similarity in gene association between these RNAs indicates that HEARO RNAs may also process themselves. However, self-splicing could not be demonstrated using protein-free assays (unpublished data), and therefore HEARO may have a different function.

We observed expression of HEARO RNA from *Exiguobacterium sibiricum* (Supplementary Fig. 7), although we have not yet determined whether these RNAs undergo unusual processing *in vivo*. Structural probing experiments *in vitro* (Supplementary Fig. 8) show that an *A. maxima* HEARO RNA adopts most of the secondary structure features predicted from comparative sequence analysis data. Therefore, these RNAs may not require protein factors to form the folded state required for their biological function, just as some large ribozymes can form their active states without the obligate participation of proteins.

Four unusually abundant RNA structures were identified in marine environmental sequences (IMES) and designated IMES-1 through IMES-4 (Supplementary Fig. 9). The first three correspond to several noncoding RNA classes recently identified independently[5], though our findings support different structural models (Supplementary Discussion). Expression of RNAs is often quantified relative to 5S rRNA[25], which is among the most abundant of bacterial

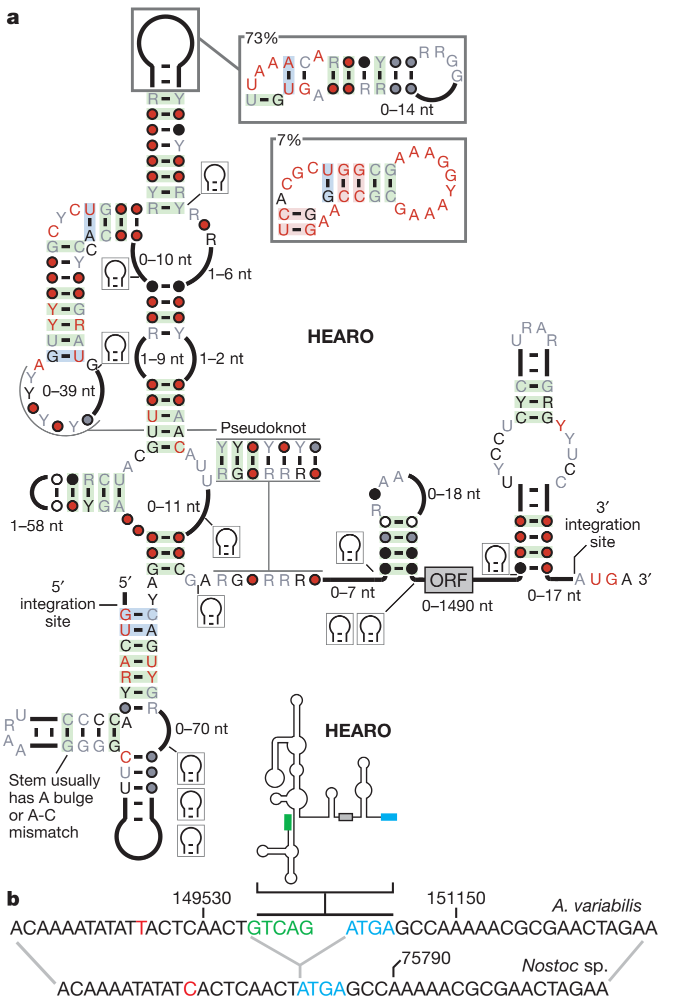


Figure 3 | HEARO RNAs. a, Consensus sequence and secondary structure model for HEARO RNAs. Annotations are as described in the legend to Fig. 2a. **b,** Typical sequence signature of HEARO genomic integration (see also Supplementary Fig. 6). (Top) HEARO element and flanking sequence in *Anabaena variabilis* ATCC 29413, plasmid C (NC_007412.1). Green text designates DNA corresponding to the first five nucleotides of conserved HEARO RNA. Blue text designates DNA corresponding to the conserved RUGA motif at each integration site. (Bottom) Homologous genome sequence lacking the HEARO element from *Nostoc* sp. PCC 7120, plasmid pCC7120beta (NC_003240.1). Red nucleotides identify positions that vary between the two genomes.

RNAs. Notably, metatranscriptome sequences collected near Station ALOHA[5,26] (Pacific Ocean) showed that all IMES RNAs are exceptionally abundant (Supplementary Table 2). IMES-1 and IMES-2 RNAs are over five- and over twofold more abundant than 5S rRNA, respectively.

Moreover, we find that IMES-1 RNA is also highly expressed in bacteria from another marine environment, in Block Island Sound (Atlantic Ocean), though not as abundantly as found in Station ALOHA samples (Supplementary Fig. 10). The high amounts of IMES-1 and IMES-2 RNAs are extremely rare for bacterial ncRNAs[25], and only 6S RNA and total tRNAs are known to outnumber 5S rRNAs[27]. Moreover, other than SprD[28] and OxyS[29] RNAs, all RNAs whose abundance is comparable to even the lower IMES-1 levels at Block Island Sound were reported by the early 1970s[25,27].

Although we have identified numerous other noncoding RNAs in our searches (for example, see Supplementary Table 1 and Supplementary Fig. 11), examples of ncRNAs with conserved sequence and structural complexity comparable to GOLLD and HEARO RNAs or with expression levels comparable to IMES RNAs are exceedingly rare. With few exceptions, these highly complex or abundant RNAs were discovered decades ago. One exception, OLE RNA[16], is a complex-folded RNA recently discovered by conducting similar phylogenetic comparative sequence analysis using DNA sequence data from cultured bacteria. This RNA is found in bacteria that can live under anaerobic conditions and that are commonly extremophilic. Thus GOLLD, HEARO and OLE RNAs are members of a select group of large and complex-folded RNAs whose mysterious functions have an impact on specialized groups of bacteria.

Only recently has sufficient DNA sequence data from cultured organisms been made available such that GOLLD and HEARO RNAs can be detected in a few disparate species, and IMES RNAs are not found at all within genome sequences derived from known bacteria. However, among the environmental sequences used to identify GOLLD and IMES RNAs, perhaps as much as 10 to 30 percent of bacterial cells in the relevant environment use these RNAs (Supplementary Table 3). Given that most bacterial species are extremely uncommon[1–3], more RNAs with extraordinary characteristics likely remain undiscovered in rarer bacteria. Thus, improvements in sequencing technologies, cultivation methods, bioinformatics and experimental approaches are poised to yield a far greater spectrum of biochemical functions for large ncRNAs from bacterial, archaeal and phage genomes.

METHODS SUMMARY

RNA motifs were discovered using a computational pipeline based on an early version of a method to cluster intergenic regions by sequence similarity[17]. The amounts of RNA expression in metatranscriptome data were established by the use of covariance model searches to identify IMES RNA and 5S RNA variants. Additional details on the sequence search and alignment methods are provided in the Methods.

Information on oligonucleotides, bacterial cultures and RNA analyses is detailed in the Methods. GOLLD RNA expression was established by treating *L. brevis* cultures with mitomycin C (0.5 μg ml^{-1}) to induce bacteriophage production. GOLLD RNA was detected by northern analysis and transcripts mapped by 5′-RNA-ligase-mediated rapid amplification of cDNA ends (RLM-RACE) and 3′-RACE. Bacteriophages were detected from supernatant by PCR. IMES-1 RNA detection and quantification was achieved using northern analysis of RNA samples isolated from bacteria collected by filtering ocean water. HEARO RNA was detected *in vivo* using RT–PCR of total RNAs isolated from cultured *E. sibiricum* cells.

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

Received 20 July; accepted 15 October 2009.

1. Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95,** 6578–6583 (1998).
2. Curtis, T. P., Sloan, W. T. & Scannell, J. W. Estimating prokaryotic diversity and its limits. *Proc. Natl Acad. Sci. USA* **99,** 10494–10499 (2002).
3. Bent, S. J. & Forney, L. J. The tragedy of the uncommon: understanding limitations in the analysis of microbial diversity. *ISME J.* **2,** 689–695 (2008).
4. Yooseph, S. *et al.* The Sorcerer II Global Ocean Sampling expedition: expanding the universe of protein families. *PLoS Biol.* **5,** e16 (2007).
5. Shi, Y., Tyson, G. W. & DeLong, E. F. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459,** 266–269 (2009).
6. Gelfand, M. S., Mironov, A. A., Jomantas, J., Kozlov, Y. I. & Perumov, D. A. A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet.* **15,** 439–442 (1999).
7. Rivas, E. & Eddy, S. R. Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* **2,** 8 (2001).
8. Wassarman, K. M., Repoila, F., Rosenow, C., Storz, G. & Gottesman, S. Identification of novel small RNAs using comparative genomics and microarrays. *Genes Dev.* **15,** 1637–1651 (2001).
9. Barrick, J. E. *et al.* New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl Acad. Sci. USA* **101,** 6421–6426 (2004).
10. Yao, Z. *et al.* A computational pipeline for high-throughput discovery of *cis*-regulatory noncoding RNA in prokaryotes. *PLOS Comput. Biol.* **3,** e126 (2007).
11. Weinberg, Z. *et al.* Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.* **35,** 4809–4819 (2007).

12. Meyer, M. M. *et al.* Identification of candidate structured RNAs in the marine organism '*Candidatus* Pelagibacter ubique'. *BMC Genomics* **10,** 268 (2009).
13. Montange, R. K. & Batey, R. T. Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.* **37,** 117–133 (2008).
14. Roth, A. & Breaker, R. R. The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.* **78,** 305–334 (2009).
15. Barrick, J. E., Sudarsan, N., Weinberg, Z., Ruzzo, W. L. & Breaker, R. R. 6S RNA is a widespread regulator of eubacterial RNA polymerase that resembles an open promoter. *RNA* **11,** 774–784 (2005).
16. Puerta-Fernandez, E., Barrick, J. E., Roth, A. & Breaker, R. R. Identification of a large noncoding RNA in extremophilic eubacteria. *Proc. Natl Acad. Sci. USA* **103,** 19490–19495 (2006).
17. Tseng, H. H., Weinberg, Z., Gore, J., Breaker, R. R. & Ruzzo, W. L. Finding non-coding RNAs through genome-scale clustering. *J. Bioinform. Comput. Biol.* **7,** 373–388 (2009).
18. Michel, F. & Westhof, E. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216,** 585–610 (1990).
19. Pace, N. R., Thomas, B. C. & Woese, C. R. in *The RNA World* 2nd edn (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) Ch. 4, 113–141 (Cold Spring Harbor Laboratory Press, 1999).
20. Toor, N., Keating, K. S. & Pyle, A. M. Structural insights into RNA splicing. *Curr. Opin. Struct. Biol.* **19,** 260–266 (2009).
21. Rusch, D. B. *et al.* The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **5,** e77 (2007).
22. Raya, R. R. & Hébert, E. M. in *Bacteriophages: Methods and Protocols* Vol. 1 (ed. Clokie, M. R. J.) (Humana, 2009).
23. Stoddard, B. L. Homing endonuclease structure and function. *Q. Rev. Biophys.* **38,** 49–95 (2005).
24. Lambowitz, A. M. & Zimmerly, S. Mobile group II introns. *Annu. Rev. Genet.* **38,** 1–35 (2004).
25. Wassarman, K. M., Zhang, A. & Storz, G. Small RNAs in *Escherichia coli*. *Trends Microbiol.* **7,** 37–45 (1999).
26. Frias-Lopez, J. *et al.* Microbial community gene expression in ocean surface waters. *Proc. Natl Acad. Sci. USA* **105,** 3805–3810 (2008).
27. Wassarman, K. M. 6S RNA: a regulator of transcription. *Mol. Microbiol.* **65,** 1425–1431 (2007).
28. Pichon, C. & Felden, B. Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains. *Proc. Natl Acad. Sci. USA* **102,** 14249–14254 (2005).
29. Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L. & Storz, G. A. Small, Stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* **90,** 43–53 (1997).

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

Acknowledgements We thank N. Carriero and R. Bjornson for assisting our use of the Yale Life Sciences High Performance Computing Center (NIH grant RR19895-02), T. Gruczka for advice and assistance in ocean water collection, J. Yang for assistance with the analysis of the dct-1 motif, D. Rodrigues for *E. sibiricum*, D. Bryant for *A. maxima* genomic DNA and P. O'Donoghue, M. Hammond, N. Sudarsan, S. Li, J. Barrick, Z. Yao, W. L. Ruzzo and E. Tseng for advice. J.P. and M.M.M. were supported by postdoctoral fellowships from the Canadian Institutes of Health Research and National Institutes of Health, respectively. R.R.B. is a Howard Hughes Medical Institute Investigator.

Author Contributions Z.W. and R.R.B. conceived the study and R.R.B. supervised the research. Z.W. created bioinformatics scripts and prepared RNA sequence alignments. J.P. conducted GOLLD and IMES RNA experiments. M.M.M. conducted GOLLD RACE and HEARO RNA experiments. Z.W. and R.R.B. wrote the manuscript, and all authors participated in editing.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.R.B. (ronald.breaker@yale.edu).

METHODS

Detection and alignment of homologous RNA sequences. New classes of structured bacterial RNAs were identified using an updated method to cluster intergenic regions on the basis of sequence similarity that is related to a method published recently[17]. Similar to our earlier method[10], CMfinder[30] was used to infer secondary structures from the clustered intergenic regions by simultaneously aligning on the basis of conserved sequence and secondary structure features. The identified structures are used subsequently in covariance model-based homology searches[31–33] to find additional examples that are used by CMfinder to refine its initial alignment. Motifs were scored using a variety of statistics as described previously[10], and by inferring a phylogenetic tree using subroutines in Pfold[34], then using pscore[35]. To find all homologues of the new RNA classes, we used various homology search strategies that were developed previously[11]. The set of genome sequences searched were RefSeq[36] version 32, and environmental sequences from acid mine drainage[37], soil and whale fall[38], human gut[39,40], mouse gut[41], gutless sea worms[42], sludge communities[43], termite hind gut[44], marine samples at various depths near Station ALOHA[45] and the global ocean survey[21,46].

Protein-coding genes were annotated using several sources. For RefSeq sequences, we used the RefSeq annotation. For global ocean survey sequences, we used proteins predicted previously[4]. For certain sequences[40,44,45], genes were predicted using the MetaGene program[47] with default parameters. For the remainder, we used predictions downloaded from IMG/M[48]. Genes were classified into protein families based on the Conserved Domain Database[49] version 2.08.

Multiple sequence alignments were developed using methods established previously[11]. Conservation statistics reflected in the consensus diagrams were calculated on the basis of protocols established previously, and the following description is adapted from our previous report[11]. To reduce bias caused by nearly redundant sequences, sequences were weighted using Infernal's implementation of the GSC algorithm. These weights were used to calculate nucleotide frequencies at each position in the alignment. Base pairs whose weighted frequency of non-canonical base pairs (neither Watson–Crick nor G–U) exceeded 10% were not classified as covarying, and are not shaded in consensus diagrams. Sequences in which both positions of the base pair are missing (deleted) or either nucleotide was uncertain (for example, was the degenerate nucleotide code 'N') were not counted in the frequency. Base paired positions annotated as showing covariation (shaded green) meant that at least two Watson–Crick or G–U pairs were observed at the given base pair position and these two pairs differed at both nucleotides (for example, A–U and C–G differ at both positions). If pairs were detected that differ at only one position (for example, A–U and G–U), this is classified as compatible mutation. The method was varied for GOLLD RNA. GOLLD RNA is very large, and yet is mostly present in environmental sequence scaffolds that are relatively short. Thus, most of the detected GOLLD RNAs are fragmentary. We did not count parts of sequences that are not present in a particular sequence fragment in any of the statistics. Also, the diagram of GOLLD RNA in Fig. 2a is based on the most common structure observed, but statistics for the highly conserved 3′ half (Supplementary Discussion) are based on all GOLLD RNAs, not just those with the most common structure.

RNA detection in metatranscriptome sequences. To determine if IMES RNAs are transcribed, we performed homology searches using accelerated covariance model searches implemented in RAVENNA[31–33] on marine transcript sequences collected from Station ALOHA, in the open Pacific Ocean[5,26]. To accommodate the short transcript reads, we performed searches in 'local' mode, which allows for partial matches to a query RNA model. To guard against false positives, we manually inspected predicted homologues, and used stringent *E*-value thresholds. Searches performed on known RNA motifs used as queries the relevant 'seed' alignments in Rfam version 9.0 (ref. 50).

Estimates of RNA distributions in genome and metagenome sequences. For the distribution analysis presented as Supplementary Table 1, the DNA sequences corresponding to data of the project 'Microbial Community Genomics at the HOT/ALOHA' were downloaded from the CAMERA website (http://camera.calit2.net). The specific files downloaded were JGI_SMPL_HF10_10-07-02_fasta.gz, JGI_SMPL_HF130_10-06-02_fasta.gz, JGI_SMPL_HF200_10-06-02_fasta.gz, JGI_SMPL_HF4000_12-21-03_fasta.gz, JGI_SMPL_HF500_10-06-02_fasta.gz, JGI_SMPL_HF70_10-07-02_fasta.gz and JGI_SMPL_HF770_12-21-03_fasta.gz. The sequences were searched in the same way as the metatranscriptome sequences, with an *E*-value cutoff of 10^{-4}. When we aligned and inspected the hits, several contained long polynucleotide repeats, which are a common source of spurious low *E*-values; these hits were discarded. (As noted above, metatranscriptome hits were also inspected, but fit the expected consensus well, so no hits were discarded from those samples.)

Calculation of RNA class sizes and structural statistics. We enumerated bacterial ncRNAs with a complex structure (based on comparative or experimental data) that were present in more than one bacterial class. RNA classes were derived based on Rfam[50] version 9.1 and NONCODE[51]. Sizes are the average reported by Rfam for the 'seed' alignment, except as follows. All rRNA statistics are based on the *E. coli* model[52]. The RNase P size is the average of sequences in the two bacterial models in Rfam. Group II intron and HEARO RNA sizes were calculated as the average of RNA length minus their embedded ORF length. Because many HEARO RNAs are found in incomplete or environmental genomes, its ORFs are not well annotated. To avoid noise from misannotations (where typically the start codon is annotated upstream of the true start codon), we subtracted the entire variable-length region that can contain the ORF. Consequently, HEARO RNA sizes might be slightly underestimated. Group II intron and ORF sizes were derived from a previous study[53]. Conserved structures of known RNAs were taken from the literature for rRNAs as above[52], group II introns[54,55], OLE[16], RNase P RNAs[56], tmRNAs[57], group I introns[18] and riboswitches[58]. At least two consecutive Watson–Crick base pairs were required to define pseudoknots.

Although many quantitative definitions of structural complexity are possible, our use of multistem junctions and pseudoknots is equally applicable to a wide variety of RNAs for which comparative analysis or biochemical experiments are possible. Definitions based on other tertiary interactions, for example, would only be appropriate for RNAs that have been the subject of many detailed biochemical experiments.

Phylogenetic tree inference. The phylogenetic tree for HEARO RNAs was inferred as follows. First, from the HEARO multiple sequence alignment, we extracted the region 5′ to the point at which the ORF is sometimes inserted. This resulting alignment was converted to sequential PHYLIP format using an in-house script, and used as input to PhyML[59]. PhyML version 3.0 was run with the command line:

```
phyml -i hearo-5prime.phylip --rand_start --n_rand_starts 10 -d nt -q -m
GTR -f e -t e -v e -a e -s SPR -o tlr
```

Support for nodes in the resulting phylogenetic tree were calculated using the -b -4 option. The tree was drawn using the drawtree command from the PHYLIP package version 3.66, written by J. Felsenstein.

Oligonucleotides. The sequences of oligonucleotides used in this study are shown below. Probes used for northern hybridizations of environmental RNA use IUPAC symbols for degenerate nucleotides.

RT–PCR for HEARO RNA, forward primer 5′-ATCATACAGGTAGAGAATGGAAGGTGACAATG-3′, reverse primer 5′-CGTCCGGTTGATAAACGATGTGACCAATC-3′. PCR to generate template for Ama-1-29 RNA (forward primer carries T7 RNA polymerase promoter), forward primer 5′-CCAAGTAATACGACTCACTATAGGTCGTCGATAGTCAGCACCCCCGG-3′, reverse primer 5′-CACGTAAAACTCCTGGGAGGGTTGG-3′. Overlapping primers used to generate a fragment of *Agrobacterium tumefaciens* 5S rRNA as positive control in Northern hybridization experiments (forward primer carries T7 RNA polymerase promoter), forward primer 5′-TAATACGACTCACTATAGGCGACCTGGTGGTCATCGCGGGGCGGCTGCACCCGTTCCCTTTCCG-3′, reverse primer 5′-CCTACTCTCCCGCGTCTTGAGACGAAGTACCATTGGCGCTGGGGCGTTTCACGGCCGTGTTCGGAATGGGAACGGGT-3′. Three overlapping primers used to generate synthetic IMES-1 RNA as positive control in Northern hybridization experiments (forward primer carries T7 RNA polymerase promoter), first primer 5′-TAATACGACTCACTATAGGTAATTTTCGACTAGTGACCAACTGCAGACGGAAGATCCTAGAGAAAAATTAAAGGAAGAGACCAAAGGGTGAAAGCAT-3′, second primer 5′-GGAAGAGACCAAAGGGTGAAAGCATTTATAAGAGTCGATGATAAAAAACAGCTTATAAATCCACCAAGAATACAAGAGAAAGTATTCAAGGAG-3′, third primer 5′-AAAACAGAGTCTAGCTCTGTTTCTTTAGTTCGAGGTCCTTAGGGGTCTAAAGTGAGATTTTCTTTAGCTCCTTGAATACTTTCTCTTG-3′. Overlapping primers used to generate synthetic IMES-2 RNA as positive control in Northern hybridization experiments (forward primer carries T7 RNA polymerase promoter), forward primer 5′-TAATACGACTCACTATAGGAAATGAATTAAGAGGCAACTCTTAACTGACCATCTGGGGAAAAACCGAGAGGTTCAAGCCCAGAGGGCAGAAAACTCTAC-3′, reverse primer 5′-TTTTGGCACGTCGTTTTATGCGCTACCGCCATGCCTTCCACTCTACTATTTTAGCGCTACTCTGTAGAGTTTTCTGCCCTCTGGGC-3′. Northern probe for IMES-1 RNA, 5′-ARGKGTCNDRAGTGAGATTYTCTTTAGCNCCTTGRNKDNTWTCTCTTNHNNYCTTGGTGG-3′. Northern probe for IMES-2 RNA, 5′-TTTGGYWCGTCGTTKTANGCGCTACCG-3′. Alternate Northern probe for IMES-3 RNA, 5′-ARYTSCGATCCAACYNRARRGTTGTGGACGATCTSA-3′. Northern probe for IMES-4 RNA, 5′-AAWYTRMTTAYTAGGTTTGCGTGTAATAA-3′. PCR of GOLLD gene in *L. brevis*, forward primer 5′-GGTTAAAAAAAAGCCGCCT-3′, reverse primer 5′-AGATTAACAGATTGAGAATACATCCG-3′. PCR of non-phage DNA in *L. brevis* (negative control), forward primer 5′-GACTGTAAAGATTGGTATTAATGGTTTC-3′, reverse primer 5′-TTAGAGCGTTGCAAAGTGCA-3′. PCR of phage DNA in *L. brevis* at a different locus from the GOLLD gene, forward primer 5′-ATTCCCGCCGTGC-3′, reverse primer 5′-CTGCTGCATCCATCTCA-3′. Primers used to generate dsDNA template for *in vitro* transcription of GOLLD 5′ RLM-RACE linker, forward primer 5′-TTTCTACTCCTTCAGTCCATGTCAGTGTCCTCGTGCTCCAGTCGCCTATAGTGAGTCGTATTA-3′, reverse primer 5′-TAATACGACTCACTATAGG-3′. Synthetic RNA linker used in GOLLD 5′ RLM RACE, 5′-CGACUGGAGCACGAGGACACUGACAUGGACUGAAGGAGUAGAAA-3′, GOLLD 5′ RLM RACE reverse transcription 5′-CCGTTACCCGCGTTACGCTTAGACCAC-3′, GOLLD 5′ RLM RACE PCR, forward primer 5′-GACTGGAGCACGAGGACACTGA-3′, reverse primer 5′-CCGGTTCGTTTCCAGCTTAACGCCTTC-3′, GOLLD 3′ RACE reverse transcription, 5′-GCGGTCACGCTTACTTAGCCCTCACTGAAATTTTTTTTTTTTTTTTT-3′, GOLLD 3′ RACE PCR, forward primer 5′-GAACGGGTGGAACTCCTTCCACCG-3′, reverse primer 5′-GCGGTCACGCTTACTTAGCCCTCACTGAA-3′.

***In vitro*-transcribed RNAs.** RNAs corresponding to 5S rRNA, IMES-1 and IMES-2 were *in vitro*-transcribed to use as standards or markers. The template DNA for the *in vitro* transcription was assembled from overlapping oligonucleotides (see sequences above). The RNA sequences expected to result from the transcription reactions are as follows. Lowercase g's represent G nucleotides that were added to improve transcription yield.

5S rRNA: 5′-ggCGACCUGGUGGUCAUCGCGGGGCGGCUGCACCCGUUCCCUUUCCGAACACGGCCGUGAAACGCCCCAGCGCCAAUGGUACUUCGUCUCAAGACGCGGGAGAGUAGG-3′.

IMES-1 RNA: 5′-ggUAAUUUUCGACUAGUGACCAACUGCAGACGGAAGAUCCUAGAGAAAAAUUAAAGGAAGAGACCAAAGGGUGAAAGCAUUUAUAAGAGUCGAUGAUAAAAAACAGCUUAUAAAUCCACCAAGAAUACAAGAGAAAGUAUUCAAGGAGCUAAAGAAAAUCUCACUUUAGACCCCUAAGGACCUCGAACUAAAGAAACAGAGCUAGACUCUGUUUU-3′.

IMES-2 RNA: 5′-ggAAAUGAAUUAAGAGGCAACUCUUAACUGACCAUCUGGGGAAAAACCGAGAGGUUCAAGCCCAGAGGGCAGAAAACUCUACAGAGUAGCGCUAAAAUAGUAGAGUGGAAGGCAUGGCGGUAGCGCAUAAAACGACGUGCCAAAA-3′.

Bacteria and growth conditions. *L. brevis* ATCC 367 was grown in MRS broth (Becton Dickinson) at 28 °C. *E. sibiricum* 255-15 was grown in half-strength tryptic soy broth (Becton Dickinson) at 37 °C.

Phage induction experiments. *L. brevis* cultures were grown in MRS broth, and samples were taken at the indicated time points starting from cells in exponential phase at attenuance $D_{600\,\mathrm{nm}}$ (often called optical density) of 0.15 to 0.2. For cultures treated with mitomycin C, the time points are relative to the addition of mitomycin C at 0.5 μg ml^{-1}. Each sample was centrifuged to isolate cells for RNA extraction and to recover supernatant for phage detection. RNA was isolated from pelleted cells using TRIzol LS (Invitrogen) in accordance with the manufacturer's instructions. Supernatant was treated with DNase RQ1 (Promega) for 1 h at 37 °C to eliminate naked genomic DNA, followed by proteinase K (Roche) and EDTA treatment 30 min at 37 °C followed by 30 min at 55 °C to degrade DNase molecules and phage capsids. Proteinase K was heat-inactivated at 96 °C for 10 min. One microlitre of a 1/10 dilution was used to deliver phage DNA for PCR templates. To ensure phage identity of the DNA, two separate regions were amplified, and a bacterial genomic DNA region was used as a negative control against host DNA (see Oligonucleotides).

Collection of water samples and extraction of RNA. Shore water samples were collected in Long Island Sound at Lighthouse Point Park (41° 15′ 6.3″ N, 72° 54′ 14.5″ W) at 12 p.m. on 26 April 2009. Off-shore water samples were collected in Block Island Sound at coordinates 41° 19′ 17″ N, 71° 32′ 11″ W between 11:30 a.m. and 12:00 p.m. on May 21, 2009. The off-shore sample was collected from 15–20 m depth (total depth was 27 m) using a sealed container whose seal was broken when it reached the specified depth. Because we did not have filtering equipment at the sampling sites, there was a delay of approximately 30 min for the shore sample, and 3 h for the off-shore sample between collection and commencement of filtration. We obtained essentially identical results in Northern hybridization experiments using water that was filtered roughly 10 h after collection as with water that was filtered 3 h after collection.

Bacterial cells were collected from the water sample by vacuum filtration onto a 47 mm diameter, 0.22 μm pore Durapore membrane filter (Millipore part #GVWP04700), with the use of a 37 mm diameter, 1.6 μm 'APFA'-type glass-fibre prefilter (Millipore part #APFA03700). After filtration the filters were stored at −80 °C. To extract RNA, lysozyme from chicken egg white (Sigma) at 1 mg ml^{-1} was applied directly to the filter, until the filter was covered (covering required 300 to 500 μl). The cells were subjected to three freeze/thaw cycles. TRIzol LS reagent was added directly to the filter and re-applied repeatedly to fully suspend cellular material. The TRIzol solution was collected and subsequent steps for RNA isolation followed the manufacturer's instructions.

GOLLD and IMES-1 RNA analysis by Northern hybridization. Northern analysis of GOLLD RNA was conducted on RNA (2 μg per lane) separated using a denaturing (6% formaldehyde) 1% agarose gel electrophoresis in a running buffer of 20 mM MOPS (free acid), 8 mM sodium acetate, 1 mM EDTA (final pH 7.0). Northern blots for IMES-1 were conducted on RNA (1 μg per lane) extracted from ocean water that was separated by denaturing (8 M urea) 6% polyacrylamide gel electrophoresis (PAGE). RNAs in the resulting gels were blotted by capillarity action to a Hybond-N+ membrane (Amersham Biosciences) and hybridization was conducted with 5′ radiolabelled oligonucleotides in Rapid-hyb buffer (GE Healthcare) with hybridization times ranging from 2 to 16 h and temperatures ranging from 42 to 45 °C. Bands were quantified using a Storm PhosphorImager and ImageQuant software (Molecular Dynamics).

Standards for quantification were created by probing an *in vitro*-transcribed IMES-1 RNA with an IMES-1-specific probe or a 5S rRNA-specific probe (see '*In vitro*-transcribed RNAs' section above for RNA sequences). The amount of *in vitro*-transcribed RNA applied to the gel ranged from 0.001 to 1 pmol. Total RNA isolated from ocean water samples was hybridized to the same membrane, and hybridized with the same probes.

Because the standards and total RNA analysis lanes were part of the same gel and membrane, their intensities can be directly compared. Thus, for example, lane 7 of the IMES-1 analysis image (Supplementary Fig. 10) appears to contain ~0.01 pmol IMES-1 RNA. Our estimates assume that the *in vitro*-transcribed IMES-1 and 5S rRNA standards anneal to the probes with similar efficiencies to homologous RNAs found in environmental samples. The *in vitro*-transcribed 5S rRNA sequence is based on the α-proteobacterial *A. tumefaciens* sequence, because the species containing IMES-1 RNAs were predicted to be related to proteobacteria, and probably to α-proteobacteria. Sample 7 appears to contain ~0.2 pmol 5S rRNA.

GOLLD 5′-RLM RACE. A total of 10 μg RNA isolated from *L. brevis* 8 h after the addition of mitomycin C was treated with tobacco acid pyrophosphatase (Epicentre Biotechnologies) to remove any 5′ pyrophosphate or triphosphate in a total volume of 20 μl for 1 h at 37 °C according to the manufacturer's instructions. The reaction was terminated by phenol/chloroform extraction and the RNA was recovered by precipitation with ethanol.

The RNA was resuspended in deionized water and ligated using T4 RNA ligase 1 (New England Biolabs) to an RNA linker (0.25 μg, see Oligonucleotides) in a total volume of 20 μl incubated at 37 °C for 1 h according to the manufacturer's instructions. The reaction was terminated and the RNA recovered as described above. The RNA linker was transcribed *in vitro* from a DNA template constructed by primer extension (see Oligonucleotides for primer sequences). The RNA was resuspended in deionized water and reverse transcription performed using a GOLLD specific primer (see Oligonucleotides) with Superscript-II reverse transcriptase (Invitrogen) in a total volume of 20 μl for 1.5 h at 42 °C according to the manufacturer's instructions. The reaction was terminated by heating at 65 °C for 20 min and 1 μl used as template for PCR amplification with *Taq* polymerase (New England Biolabs) (see Oligonucleotides for primer sequences). PCR products were cloned using a TOPO-TA cloning kit (Invitrogen) and transformed into TOP10 cells (Invitrogen). Plasmids from 12 of the resulting colonies were purified (Qiagen) and sequenced (Keck DNA sequencing resource, Yale University).

The initial RLM-RACE experiments produced sequences with additional bases between the expected linker sequence and the genomic sequence from *L. brevis* resulting from run-on transcription of the *in vitro*-transcribed RNA-linker[60]. Although the 5′ extents of the transcripts seemed clear, the entire RLM-RACE was repeated as above using a synthetic linker purchased from the Keck Biotechnology Resources Laboratory with a 2′ protecting TOM group, deprotected[61] through treatment with 1 M tetrabutylammonium fluoride in tetrahydrofuran (Sigma) and subsequently purified by denaturing 6% PAGE. An additional 11 sequences were obtained and the combined results of both experiments are reported. Some significantly shorter sequences resulted from the second RACE experiment where the RNA was probably more degraded owing to sample handling and additional freeze-thaw cycles between the two experiments. However, the dominant 5′ ends in the second experiment match the location determined from the first experiment.

GOLLD 3′ RACE. A total of 10 μg RNA isolated from *L. brevis* 8 h after the addition of mitomycin C was polyadenylated using *E. coli* poly(A) polymerase (New England Biolabs) according to the manufacturer's instructions. The reaction was terminated and the RNA was recovered as described above. The polyadenylated RNA was resuspended in water and reverse transcription was conducted using Superscript II reverse transcriptase (Invitrogen) in a total volume of 20 μl at 42 °C for 1.5 h according to the manufacturer's instructions. The reaction was terminated by heating at 65 °C for 20 min and 1 μl was subsequently used as template for PCR using *Taq* polymerase (New England Biolabs) (see Oligonucleotides for primer sequences). PCR products were cloned and their DNA sequenced as described above.

RT–PCR analysis of HEARO RNA. *E. sibiricum* cultures were harvested during both stationary and logarithmic phase growth. The equivalent of 1 ml of cell culture at an D_{600} of 3 were pelleted and resuspended in 100 μl of 3 mg ml^{-1} lysozyme in TE buffer (10 mM Tris-HCl (pH 7.5 at 23 °C), 1 mM EDTA). Cell lysis was facilitated by several freeze-thaw cycles before isolating RNA with 1 ml of TRIzol using the manufacturer's protocol. DNA was removed through digestion with DNase RQ1 (Promega) at 37 °C for 1 h. Approximately 2.5 μg of total RNA was used as template for reverse transcription at 55 °C for 1.5 h in a volume of 20 μl using SuperScript III reverse transcriptase (Invitrogen) primed by random DNA hexamers supplied by the vendor according to the manufacturer's instructions. Negative control samples included for each analysis were prepared using identical conditions but without enzyme addition. One microlitre of each reverse transcription reaction was used to deliver cDNA templates for PCR.

In-line probing of HEARO (Ama-1-29 RNA). DNA corresponding to the Ama-1-29 RNA was amplified from *Arthrospira maxima* genomic DNA by PCR, and the resulting DNA product was used as template for *in vitro* transcription using T7 polymerase to produce RNA (see Oligonucleotides for primer sequences). The RNA was purified by 6% denaturing PAGE, extracted from the gel slice in 200 mM NaCl, 10 mM Tris-HCl (pH 7.5 at 25 °C), 1 mM EDTA and precipitated with ethanol. The RNA was resuspended in deionized water and separate aliquots of the RNA were 5′ or 3′ ^{32}P-labelled.

For the 5′ labelling, 10 pmol of RNA was dephosphorylated using rAPid alkaline phosphatase (Roche) according to the manufacturer's instructions. The phosphatase reaction was terminated by heating at 95 °C for 3 min and the dephosphorylated RNA was subsequently 5′ ^{32}P-labelled in 5 mM $MgCl_2$, 25 mM CHES (pH 9.0 at 25 °C), 3 mM DTT using 40 μCi of [γ-^{32}P]-ATP and 20 U T4 polynucleotide kinase (New England Biolabs) and incubated at 37 °C for 1 h. The RNA was purified by denaturing PAGE and recovered from the gel as described above.

For the 3′ labelling, 50 pmol of RNA was incubated in 50 mM Tris-HCl (pH 7.8 at 25 °C), 10 mM $MgCl_2$, 10 mM DTT, 2 mM ATP, 10% DMSO with 150 μCi of [5′-^{32}P]-pCp, and 50 U of T4 RNA ligase 1 (New England Biolabs) at 4 °C for 40 h. RNA was purified and recovered as described above. In-line probing reactions were prepared with radiolabelled RNAs as noted and analysed by denaturing 10% PAGE essentially as described previously[62].

30. Yao, Z., Weinberg, Z. & Ruzzo, W. L. CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics* **22,** 445–452 (2006).
31. Weinberg, Z. & Ruzzo, W. L. Sequence-based heuristics for faster annotation of non-coding RNA families. *Bioinformatics* **22,** 35–39 (2006).
32. Eddy, S. R. & Durbin, R. RNA Sequence Analysis Using Covariance Models. *Nucleic Acids Res.* **22,** 2079–2088 (1994).
33. Klein, R. J. & Eddy, S. R. RSEARCH: finding homologs of single structured RNA sequences. *BMC Bioinformatics* **4,** 44 (2003).
34. Knudsen, B. & Hein, J. Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Res.* **31,** 3423–3428 (2003).
35. Yao, Z. *Genome scale search of noncoding RNAs: bacteria to vertebrates.* Dissertation, Univ. of Washington (2008).
36. Pruitt, K., Tatusova, T. & Maglott, D. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33,** D501–D504 (2005).
37. Tyson, G. W. *et al.* Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428,** 37–43 (2004).
38. Tringe, S. G. *et al.* Comparative metagenomics of microbial communities. *Science* **308,** 554–557 (2005).
39. Gill, S. R. *et al.* Metagenomic analysis of the human distal gut microbiome. *Science* **312,** 1355–1359 (2006).
40. Kurokawa, K. *et al.* Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* **14,** 169–181 (2007).
41. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444,** 1027–1031 (2006).
42. Woyke, T. *et al.* Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443,** 950–955 (2006).
43. Martín, H. G. *et al.* Metagenomic analysis of two enhanced biological phosphorus removal (EBPR) sludge communities. *Nature Biotechnol.* **24,** 1263–1269 (2006).
44. Warnecke, F. *et al.* Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450,** 560–565 (2007).
45. Konstantinidis, K. T. *et al.* Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Appl. Environ. Microbiol.* **75,** 5345–5355 (2009).
46. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304,** 66–74 (2004).
47. Noguchi, H., Park, J. & Takagi, T. MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res.* **34,** 5623–5630 (2006).
48. Markowitz, V. M. *et al.* IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* **36,** D534–D538 (2008).
49. Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res.* **33,** D192–D196 (2005).
50. Gardner, P. P. *et al.* Rfam: updates to the RNA families database. *Nucleic Acids Res.* **37,** D136–D140 (2009).
51. Liu, C. *et al.* NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Res.* **33,** D112–D115 (2005).
52. Gutell, R. R., Larsen, N. & Woese, C. R. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.* **58,** 10–26 (1994).
53. Dai, L. & Zimmerly, S. Compilation and analysis of group II intron insertions in bacterial genomes: evidence for retroelement behavior. *Nucleic Acids Res.* **30,** 1091–1102 (2002).
54. Boudvillain, M. & Pyle, A. M. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. *EMBO J.* **17,** 7091–7104 (1998).
55. Toor, N., Hausner, G. & Zimmerly, S. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* **7,** 1142–1152 (2001).
56. Haas, E. S., Brown, J. W., Pitulle, C. & Pace, N. R. Further perspective on the catalytic core and secondary structure of ribonuclease P RNA. *Proc. Natl Acad. Sci. USA* **91,** 2527–2531 (1994).
57. Zwieb, C., Wower, I. & Wower, J. Comparative sequence analysis of tmRNA. *Nucleic Acids Res.* **27,** 2063–2071 (1999).
58. Barrick, J. E. & Breaker, R. R. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.* **8,** R239 (2007).
59. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).
60. Cazenave, C. & Uhlenbeck, O. C. RNA-template-directed RNA synthesis by T7 polymerase. *Proc. Natl Acad. Sci. USA* **91,** 6972–6976 (1994).
61. Wu, T., Ogilvie, T. T. & Pon, R. T. Prevention of chain cleavage in the chemical synthesis of 2′ silylated oligoribonucleotides. *Nucleic Acids Res.* **17,** 3501–3517 (1989).
62. Regulski, E. E. & Breaker, R. R. in *Methods in Molecular Biology* Vol. 419 *Post-Transcriptional Gene Regulation* (ed. Wilusz, J.) (Humana, 2008).

nature Vol 462|3 December 2009|doi:10.1038/nature08599

LETTERS

In vitro reconstitution of an abscisic acid signalling pathway

Hiroaki Fujii[1,2]*, Viswanathan Chinnusamy[1,2]*, Americo Rodrigues[3], Silvia Rubio[3], Regina Antoni[3], Sang-Youl Park[1], Sean R. Cutler[1], Jen Sheen[4], Pedro L. Rodriguez[3] & Jian-Kang Zhu[1,2]

The phytohormone abscisic acid (ABA) regulates the expression of many genes in plants; it has critical functions in stress resistance and in growth and development[1–7]. Several proteins have been reported to function as ABA receptors[8–13], and many more are known to be involved in ABA signalling[3,4,14]. However, the identities of ABA receptors remain controversial and the mechanism of signalling from perception to downstream gene expression is unclear[15,16]. Here we show that by combining the recently identified ABA receptor PYR1 with the type 2C protein phosphatase (PP2C) ABI1, the serine/threonine protein kinase SnRK2.6/OST1 and the transcription factor ABF2/AREB1, we can reconstitute ABA-triggered phosphorylation of the transcription factor *in vitro*. Introduction of these four components into plant protoplasts results in ABA-responsive gene expression. Protoplast and test-tube reconstitution assays were used to test the function of various members of the receptor, protein phosphatase and kinase families. Our results suggest that the default state of the SnRK2 kinases is an autophosphorylated, active state and that the SnRK2 kinases are kept inactive by the PP2Cs through physical interaction and dephosphorylation. We found that in the presence of ABA, the PYR/PYL (pyrabactin resistance 1/PYR1-like) receptor proteins can disrupt the interaction between the SnRK2s and PP2Cs, thus preventing the PP2C-mediated dephosphorylation of the SnRK2s and resulting in the activation of the SnRK2 kinases. Our results reveal new insights into ABA signalling mechanisms and define a minimal set of core components of a complete major ABA signalling pathway.

Several ABA receptors have been reported[8–13], although many of them remain unconfirmed[15,16]. Recently, a family of novel START domain proteins, known as PYR/PYLs (also known as RCARs), were identified as ABA receptors. Several of the PYR/PYLs were shown to interact with and inhibit clade-A PP2Cs[11–13]. The PP2Cs (ABI1, ABI2, HAB1 and PP2CA/AHG3) negatively regulate ABA responses[13]. In contrast, a subfamily of ABA-activated SnRK2s are positive regulators of ABA signalling[17–21]. Through unknown mechanisms, the inhibition of the negatively acting PP2Cs leads to the successful activation of a subfamily of SnRK2 kinases (SnRK2.2, SnRK2.3 and SnRK2.6 in *Arabidopsis*), which phosphorylate the basic leucine zipper (bZIP) transcription factors called ABFs/AREBs[22,23]. The ABFs bind to ABA-responsive promoter elements (ABRE) to induce the expression of ABA-responsive genes[1].

The present study was aimed at defining the core components of the ABA response pathway that are both necessary and sufficient for ABA perception, signalling, and finally ABA-responsive gene expression. It has been suggested that ABA-dependent phosphorylation of ABF2 at amino-acid residues S26, S86, S94 and T135 is important for stress-responsive gene expression in *Arabidopsis*[23]. We used transient activation analysis with protoplasts from the *snrk2.2/2.3/2.6* triple mutant to determine the role of ABF2 phosphorylation and its dependence on SnRK2s for ABA-responsive gene expression. We have shown previously that the *snrk2.2/2.3/2.6* triple mutant is deficient in ABA responses[21]. As expected, transfection of *snrk2.2/2.3/2.6* protoplasts with ABF2 did not induce *RD29B–LUC* (luciferase reporter gene driven by the ABA-responsive *RD29B* promoter) expression even in the presence of ABA, but co-transfection of ABF2 with SnRK2.6 resulted in the induction of *RD29B–LUC* in an ABA-dependent manner (Fig. 1a). Furthermore, ABF2 with alanine substitutions at all of the four phosphorylation sites was inactive, whereas aspartic acid substitutions at these sites led to a constitutively active ABF2, resulting in the induction of *RD29B–LUC* expression even without ABA treatment (Fig. 1a). Co-transfection of Ala-substituted ABF2 with SnRK2.6 led to only a very low level of *RD29B–LUC* induction (Fig. 1a). Replacement of lysine 50, a conserved residue critical for ATP-binding and kinase activity, with asparagine (K50N) inactivates SnRK2.6 in phosphorylation assays *in vitro* (H.F. and J.-K.Z., unpublished observations). Co-transfection of ABF2 with SnRK2.6^{K50N} did not induce *RD29B–LUC* expression (Fig. 1a), demonstrating that the kinase activity is necessary for ABF2 activation. Transfection of ABF2 alone in wild-type protoplasts induced a low level of *RD29B–LUC* expression under ABA treatment, which is consistent with the presence of a low basal level of endogenous ABA-signalling components in the protoplasts (Supplementary Fig. 1a). These results show that SnRK2.6 mediates ABF2 activation in an ABA-dependent manner, and that ABF2 phosphorylation is sufficient for the induction of *RD29B–LUC* expression by ABA.

We next tested the effect of ABI1 and PYR1 on the induction of *RD29B–LUC* expression by ABA. Transfection of ABI1 together with ABF2 and SnRK2.6 resulted in inhibition of *RD29B–LUC* expression (Fig. 1b and Supplementary Fig. 1a). This shows that ABI1 negatively regulates the activation of *RD29B–LUC* expression that is dependent on SnRK2.6 and ABF2. Addition of PYR1 together with ABI1, SnRK2.6 and ABF2 enabled the ABA-dependent induction of *RD29B–LUC* expression (Fig. 1b and Supplementary Fig. 1a). However, addition of PYR1^{P88S}, which is defective in interaction with and inhibition of PP2Cs[12], did not enable the ABA-dependent induction of *RD29B–LUC* expression. The dominant *abi1-1* mutation (G180D) disrupts the interaction between ABI1 and PYR1 (ref. 12). Like the wild-type ABI1, ABI1^{G180D} also inhibited the effect of SnRK2.6 and ABF2 on *RD29B–LUC* expression in response to ABA, but this antagonistic effect could not be overcome by expression of PYR1 (Fig. 1b and Supplementary Fig. 1a). This suggests that the ABI1^{G180D} mutant protein retains the inhibitory activity but can no longer be regulated. Thus reconstitution with PYR1, ABI1, SnRK2.6 and ABF2 is sufficient

---

[1]Department of Botany and Plant Sciences, University of California at Riverside, Riverside, California 92521, USA. [2]Center for Plant Stress Genomics and Technology, 4700 King Abdullah University of Science and Technology, Thuwal 23955-6900, Kingdom of Saudi Arabia. [3]Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-UPV, ES-46022 Valencia, Spain. [4]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.
*These authors contributed equally to this work.

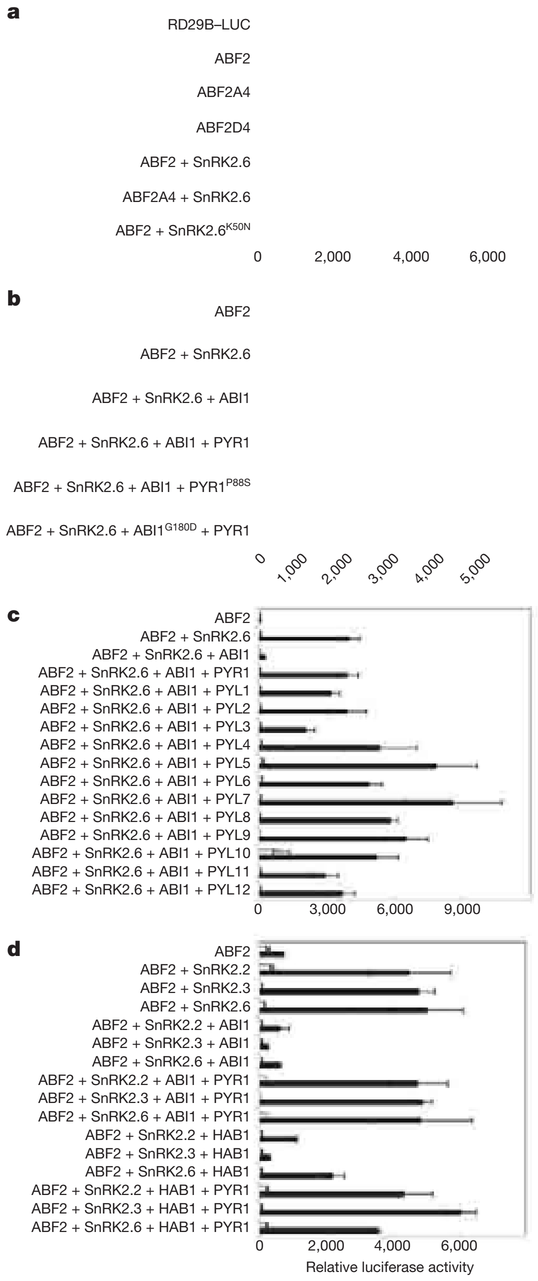


Figure 1 | Reconstitution of the ABA signalling pathway for stress-responsive gene expression in *Arabidopsis* protoplasts. a, SnRK2-mediated phosphorylation of ABF2 is sufficient for ABA-responsive gene expression. **b**, Reconstitution of ABA signalling pathway by co-expression of PYR1, ABI1, SnRK2.6 and ABF2. **c**, Reconstitution of ABA signalling pathway with different members of the PYR/PYL family. **d**, Reconstitution with different combinations of the core components. Protoplasts (2×10^4) from the *snrk2.2/3/6* triple mutant were used except in **d**, in which protoplasts from Col-0 wild-type plants were used. *RD29B*::*LUC* and *ZmUBQ*::*GUS* were used as the ABA-responsive reporter and internal control, respectively. After transfection, protoplasts were incubated for 5 h under light and in the absence of ABA (open bars) or in the presence of 5 μM ABA (filled bars). Error bars indicate s.e.m. ($n = 3$).

to enable ABA-mediated gene expression in protoplasts, providing *in vivo* evidence in favour of our previously proposed model of ABA signalling[12].

The PYR/PYL family consists of 14 members. Although genetic studies suggested redundancy in their function[12], it is not known whether all members can act as ABA receptors and transduce the ABA signal to induce gene expression. To address this matter, we reconstituted the ABA signalling pathway with different members of the PYR/PYL family. Our results show that all of the tested PYR/PYLs could antagonize the ability of ABI1 to inhibit the ABA-dependent induction of *RD29B–LUC* expression in *snrk2.2/2.3/2.6* protoplasts, although not all PYR/PYL members were equally effective (Fig. 1c). The results suggest that all of the PYR/PYLs are likely to function as ABA receptors. We also tested reconstitution of the ABA signalling pathway with different combinations of SnRK2 kinases, PP2Cs and receptors, and found that the SnRK2 kinases are inhibited by both the ABI1 and HAB1 PP2Cs, and that PYR1 or PYL2 can antagonize this inhibition. The inhibitory effect of ABI1 was stronger than that of HAB1 in the reconstituted ABA signalling system in protoplasts (Fig. 1b–d and Supplementary Fig. 1). Each of the three clade A PP2Cs (ABI1, ABI2 and HAB1) was capable of interacting with the three SnRK2 kinases (SnRK2.2, SnRK2.3 and SnRK2.6) in yeast two-hybrid (Y2H) assays, although with different intensities. For example, the ABI1 interaction was stronger than that of ABI2 or HAB1 (Supplementary Fig. 2a), which correlates with the level of inhibitory effect of ABI1 and HAB1 in the protoplast assay (Fig. 1d and Supplementary Fig. 1b). A carboxy-terminally truncated SnRK2.6 lacking amino acids 280–362 did not interact with ABI1 (Supplementary Fig. 2a), which is consistent with previous studies demonstrating that deletion of a short C-terminal domain abrogates the interaction between ABI1 and SnRK2.6 in yeast[19]. Bimolecular fluorescence complementation (BiFC) assays in tobacco show that ABI1 interacts with the SnRK2s in the nucleus as well as in the cytosol and that the C-terminal region of SnRK2.6 is required for the interaction with ABI1 (Supplementary Fig. 2b). Expression of the fusion proteins was verified by immunoblot analysis (Supplementary Fig. 2c). The interaction between ABI1 and SnRK2.6 *in vivo* was further confirmed by a co-immunoprecipitation assay using the tobacco protein extracts (Supplementary Fig. 2c).

PYR/PYLs inactivate clade A PP2Cs in an ABA-dependent manner[11–13]. In protoplast transactivation assays, we showed that PYR/PYLs can reverse the inhibitory effect of PP2Cs (Fig. 1 and Supplementary Fig. 1). We speculated that the PYR/PYLs might prevent the inhibitory effect of the PP2Cs by disrupting the interaction between the PP2Cs and the SnRK2s. We used yeast triple-hybrid assays to test whether co-expression of PYLs might disrupt the interaction between PP2Cs and SnRK2s. First, we reproduced the interaction of the ABI1, ABI2 and HAB1 PP2Cs (fused to the Gal4 activation domain (GAD)) with SnRK2.6 (fused to the Gal4 DNA-binding domain (GBD)) by using the pBridge triple-hybrid vector (Supplementary Fig. 3). Next, we cloned PYL5 and PYL8 into the SnRK2.6–pBridge construct; these proteins have been shown to act as potent inhibitors of the PP2Cs[13]. Nuclear localization of PYL5 and PYL8 in yeast is driven by fusion with a nuclear localization sequence present in the pBridge vector. Co-expression of PYL8 with GBD-SnRK2.6 led to an abrogation or decrease (depending on the dilution of the yeast culture) in the interaction with GAD–ABI1 (Supplementary Fig. 3). Similar results were obtained when GBD–SnRK2.6 and GAD–ABI2 or GAD–HAB1 were tested with either PYL8 or PYL5, respectively (Supplementary Fig. 3). These results show that co-expression of a PYL impairs the interaction of ABI1, ABI2 and HAB1 PP2Cs with SnRK2.6.

We have reconstituted the apparent entire ABA signalling pathway for stress-responsive gene expression by the co-expression of PYR/PYLs, PP2Cs, SnRK2s and ABF2 in *Arabidopsis* protoplasts (Fig. 1b–d and Supplementary Fig. 1). To verify whether these are the minimal signalling components that are both necessary and sufficient for ABA signalling in the absence of other cellular components, we attempted to reconstitute the pathway *in vitro*. We constructed recombinant maltose-binding protein (MBP)-tagged SnRK2.6 and found that it is capable of phosphorylating an ABF2 fragment as well as showing autophosphorylation (Fig. 2a, b). Incubation of glutathione *S*-transferase (GST)-tagged ABI1 but not GST with SnRK2.6 before the kinase assay substantially decreases ABF2 phosphorylation by the recombinant SnRK2.6 (Fig. 2a). SnRK2.6 pulled down from extracts of

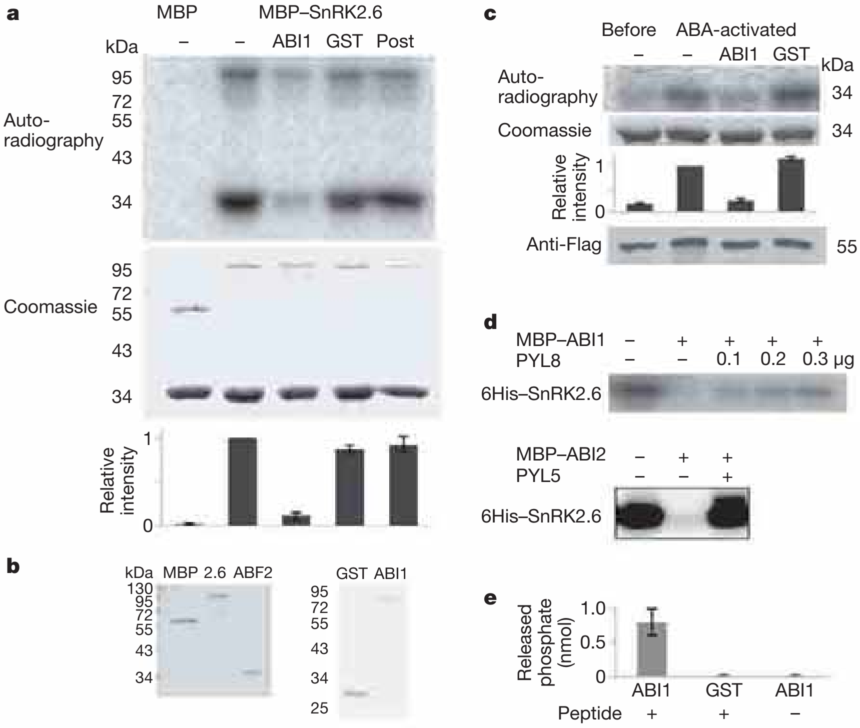


Figure 2 | ABI1 and ABI2 inhibit SnRK2.6 by dephosphorylation. **a**, SnRK2.6 is deactivated by ABI1. MBP or MBP–SnRK2.6 treated without (−) or with GST–ABI1 or GST was incubated with a GST–ABF2 fragment (Gly 73 to Gln 119) in the presence of [γ-^{32}P]ATP. In the rightmost lane (Post), GST–ABI1 was added after phosphorylation of the GST–ABF2 fragment by MBP–SnRK2.6. Bands of GST–ABF2 fragment and MBP–SnRK2.6 are indicated by an arrow and an arrowhead, respectively. Radioactivities of GST–ABF2 fragment bands were measured with a phosphoimager and were normalized, taking as unity the radioactivity of the band by MBP–SnRK2.6 without ABI1 treatment. Error bars indicate s.e.m. ($n = 5$). **b**, Coomassie staining of purified MBP, SnRK2.6, ABF2, GST and ABI1. **c**, Flag-tagged SnRK2.6 extracted from transgenic plants before and after ABA treatment was used instead of the MBP–SnRK2.6 in **a**. Coomassie staining, autoradiography and relative radioactivities (error bars indicate s.e.m.; $n = 5$) of the GST–ABF2 fragment are shown. Western blotting with anti-Flag antibody shows the amount of Flag–SnRK2.6 protein. **d**, Autoradiography of autophosphorylated SnRK2.6 showing dephosphorylation of SnRK2.6 by MBP–ABI1 and MBP–ABI2 and the effect of PYL8 and PYL5, respectively, in the presence of 1 μM ABA. **e**, Release of phosphate from the synthetic peptide HSQPKpSTVGTP, corresponding to residues 170–180 of SnRK2.6. Error bars indicate s.e.m. ($n = 3$).

ABA-treated plants is also active in phosphorylating ABF2, but SnRK2.6 from untreated plants is not. This phosphorylation is also inhibited by GST–ABI1 (Fig. 2c). ABI1 added after ABF2 phosphorylation by SnRK2.6 is not as effective in decreasing the level of phosphorylation (Fig. 2a), suggesting that ABI1 inhibits ABF2 phosphorylation by dephosphorylating SnRK2.6 (Fig. 2a). Indeed, we found that both ABI1 and ABI2 efficiently dephosphorylated SnRK2.6 (Fig. 2d). The autophosphorylated Ser 175 is essential for the kinase activity of SnRK2.6 *in vitro*[24]. We found that ABI1 can dephosphorylate a synthetic phosphopeptide corresponding to amino-acid residues His 170–Pro 180 of SnRK2.6, which is phosphorylated at Ser 175 (HSQPKpSTVGTP; Fig. 2e). These results suggest that ABI1 may deactivate SnRK2.6 by dephosphorylating Ser 175.

When His-tagged PYR1 is incubated together with GST–ABI1 and MBP–SnRK2.6, SnRK2.6-mediated phosphorylation of ABF2 is significantly recovered in the presence of 2 μM (+)-ABA (Fig. 3a, b). Without ABA, His-PYR1 cannot reverse the inhibitory effect of ABI1 on SnRK2.6-mediated phosphorylation of ABF2 (Fig. 3a). PYR1$^{\mathrm{P88S}}$, which cannot bind to and inhibit ABI1 (ref. 12), is not capable of reversing the inhibitory effect of ABI1 even in the presence of ABA (Fig. 3a). We found that in the presence of ABA, PYL8 or PYL5 can prevent the dephosphorylation of SnRK2.6 by ABI1 or ABI2 (Fig. 2d). These data are consistent with results from the protoplast transactivation assays, and they show that it is possible to reconstitute the activation of ABF2 phosphorylation by ABA *in vitro*. The ABF2 phosphorylation status in this reconstituted *in vitro* system responds to ABA in a concentration-dependent manner (Fig. 3c). The apparent concentration of ABA giving a half-maximal response is 0.8 μM, which is similar to the concentration of ABA giving a half-maximal inhibition of seed germination[11] and falls within the physiological range of ABA concentrations in plants. Similar responses to ABA were observed when ABA-activated SnRK2.6 isolated from plants instead of recombinant SnRK2.6 was used in the reconstitution assay (Supplementary Fig. 4a). Furthermore, reconstitution was also achieved when the PP2C protein HAB1 was used instead of ABI1 (Supplementary Fig. 4b). Our protoplast and *in vitro* reconstitution results support a model in which PYR1 (and PYLs) binds ABA, and then interacts with and is able to inactivate the PP2Cs. The ABA-bound receptors also disrupt the interaction between the PP2Cs and the SnRK2 kinases. These actions of the receptors prevent the dephosphorylation and thereby relieve inhibition of the SnRK2s by the PP2Cs. The relieved SnRK2s can then phosphorylate ABFs to activate ABA-responsive genes.

Consistent with our model was our previous observation[12] that the SnRK2s are substantially less activated by ABA in the *pyr1pyl1pyl2pyl4* mutant than in the wild type. The model also predicts that the SnRK2s may be constitutively activated in mutant plants deficient in the PP2Cs. Indeed, the PP2C triple mutant *abi1-2hab1-1pp2ca-1* shows a constitutive activation of 42 and 45-kDa kinases, which correspond to SnRK2.2/2.3 and SnRK2.6, respectively (Fig. 4a). This mutant displays a constitutive ABA response phenotype in germination and early seedling development (Fig. 4b, c), as reported previously[25]. In contrast, the PP2C triple mutant *abi1-2hab1-1abi2-2* does not have a constitutive ABA response as strong as that in *abi1-2hab1-1pp2ca-1* (Fig. 4b, c) and does not show a strong constitutive activation of the SnRK2s (Fig. 4a).

We have achieved the *in vitro* reconstitution of a phytohormone signal transduction pathway using recombinant proteins. The *in vitro* reconstitution results are supported by the reconstitution assays in

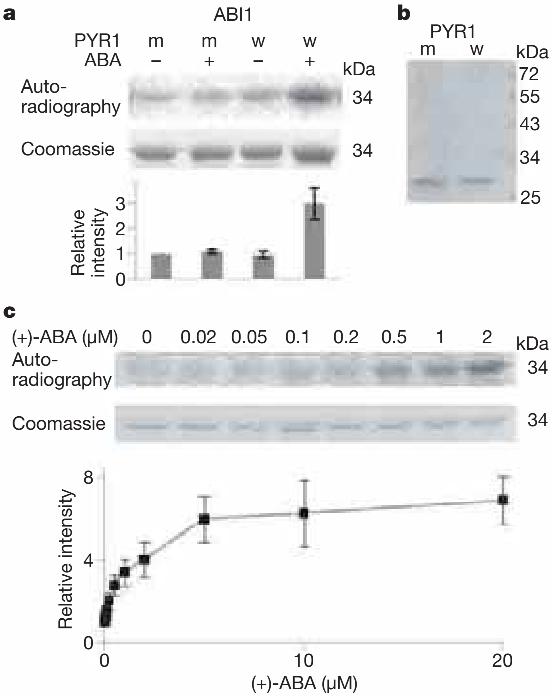


Figure 3 | The combined effect of ABA, PYR1 and ABI1 on the phosphorylation of the GST–ABF2 fragment by SnRK2.6 *in vitro*. **a**, Reconstitution of ABA regulation of ABF2 phosphorylation. MBP–SnRK2.6 treated with GST–ABI1 and His-tagged wild-type PYR1 (w) or mutated $\text{PYR1}^{\text{P88S}}$ (m) in the absence (−) or presence (+) of 2 μM (+)-ABA was incubated with a GST–ABF2 fragment (Gly 73 to Gln 119) in the presence of $[\gamma\text{-}^{32}\text{P}]$ATP. Coomassie staining, autoradiography and relative radioactivities of GST–ABF2 fragment are shown. Radioactivities of the GST–ABF2 fragment were normalized, taking as unity the radioactivity of the band with $\text{PYR1}^{\text{P88S}}$ in the absence of ABA. Error bars indicate s.e.m. ($n = 5$). **b**, Coomassie staining of PYR1 (w) and $\text{PYR1}^{\text{P88S}}$ (m). **c**, ABA dose response. MBP–SnRK2.6, GST–ABI1 and His-PYR1 were incubated with different concentrations of (+)-ABA before the kinase assay, using the GST–ABF2 fragment as substrate. Coomassie staining, autoradiography and relative radioactivities of the GST–ABF2 fragment are shown, taking as unity the radioactivity of the band in the absence of ABA. Error bars indicate s.e.m. ($n = 9$ for less than 5 μM, $n = 4$ for 5 μM or more).

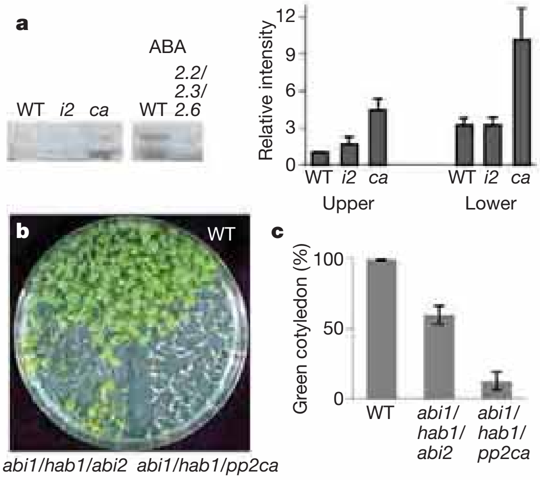


Figure 4 | Effect of PP2C mutations on ABA response phenotypes and kinase activities of SnRK2s. **a**, In-gel kinase assay showing the activities of SnRK2s in the *abi1/hab1/abi2* (*i2*) and *abi1/hab1/pp2ca* (*ca*) triple mutants. *snrk2.2/2.3/2.6* was used as a control. A GST-fused ABF2 fragment (Gly 73 to Gln 119) was used as the phosphorylation substrate. The expected positions of SnRK2.6 and SnRK2.2/2.3 are indicated by an arrow and an arrowhead, respectively. Radioactivities of the upper and lower bands were normalized, taking as unity the radioactivity of the upper band in the wild type (WT). Error bars indicate s.e.m. ($n = 3$). **b**, The PP2C triple mutants show hypersensitivity to ABA during germination and early seedling development. The photograph shows plants of the indicated genotypes grown for 14 days on MS medium containing 3% sucrose. **c**, The percentage of seedlings with green cotyledons 6 days after the end of stratification. Error bars indicate s.e.m. ($n = 3$).

the protoplasts and by genetic analysis. The protoplast reconstitution assays enabled us to test the functions of nearly all members of the PYR/PYL family. Our results indicate that all members of the family can function as ABA receptors in inducing gene expression. Although each of the proteins used in the reconstitution assays has been studied previously, it was not known how these components might connect to form a signalling pathway. Our study has revealed insights into the mechanisms of action of these components. Our results suggest that the default state of the SnRK2 protein kinases is an autophosphorylated, active state, and that the SnRK2 kinases are kept inactive by the PP2Cs through physical interaction and dephosphorylation. We found that, on binding to ABA, the PYR/PYL receptor protein can disrupt or decrease the interaction between the SnRK2s and PP2Cs and prevent the PP2C-mediated dephosphorylation of the SnRK2s, thus resulting in the activation of the SnRK2 kinases.

Successful reconstitution with the recombinant proteins implies that we have identified all essential core components of an ABA response pathway from hormone perception to phosphorylation of ABFs. Although ABA signalling in plants has been considered to be very complicated with numerous other proteins involved, our study reveals a surprising simplicity of the pathway and demonstrates that the PYR/PYLs, clade-A PP2Cs, SnRK2s and ABFs are the only core components to complete the ABA regulation of gene expression. Because there are multiple family members for each of these core components, many combinations of them are possible. The functions of the family members may overlap, but their unique spatial and temporal expression patterns may confer some distinct functions in specific tissues. Extensive genetic analysis will be necessary to determine the *in planta* importance of specific combinations of the core components.

We suggest that the other proteins previously identified as being involved in ABA responses may function to modulate the expression and/or activities of one or more of the core components defined here. Calcium and reactive oxygen signalling, RNA metabolism and protein degradation are known to be important in regulating ABA sensitivity[2–4,14,26,27]. It will be of great interest to determine how these processes may connect to one or more of the core components to affect ABA responses. It will also be interesting to determine whether other ABA response pathways such as ABA regulation of ion channels in guard cells[2,3,6] also use components of the PYR/PYLs–PP2C–SnRK2 regulatory module and whether additional receptors and core signalling components are involved.

METHODS SUMMARY

Transient activity assays were performed in *Arabidopsis* mesophyll protoplasts from Columbia wild-type or *snrk2.2/2.3/2.6* (ref. 21) plants as described previously (http://genetics.mgh.harvard.edu/sheenweb)[28]. Transfected protoplasts were incubated for 5 h in light in the absence of ABA or the presence of 5 μM ABA, and then used for the measurement of luciferase (LUC) and β-glucuronidase (GUS) activities as described previously[28]. Yeast two-hybrid and triple-hybrid assays, co-immunoprecipitation and BiFC assays were similar to those described previously[13]. Purification of GST–HAB1, His–PYR1 and His–$\text{PYR1}^{\text{P88S}}$ was performed as described previously[12]. GST, GST–ABI1, GST–ABF2 fragment, MBP and MBP–SnRK2.6 constructs were transformed into *Escherichia coli* Rosetta cells (Novagen) and the recombinant proteins were isolated by affinity purification. Purification of MBP–ABI1, MBP–ABI2, His–PYL8, His–PYL5 and His–SnRK2.6 was as described previously[13]. In-gel kinase assays were performed as described previously[20] with the modification that 300 μg of protein was loaded for samples without ABA treatment. For germination assays, seeds were plated on MS (Murashige and Skoog) nutrient medium containing 3% sucrose. In each experiment, at least 50 seeds per genotype were stratified at 4 °C for 3 days, and the presence of green cotyledons was scored after incubation for 6 days at 23 °C.

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

Received 16 September; accepted 26 October 2009.
Published online 18 November; corrected 3 December 2009 (see full-text HTML version for details).

1. Zhu, J. K. Salt and drought stress signal transduction in plants. *Annu. Rev. Plant Biol.* **53,** 247–273 (2002).
2. Hetherington, A. M. Guard cell signaling. *Cell* **107,** 711–714 (2001).
3. Schroeder, J. I., Kwak, J. M. & Allen, G. J. Guard cell abscisic acid signalling and engineering of drought hardiness in plants. *Nature* **410,** 327–330 (2001).
4. Finkelstein, R. R., Gampala, S. S. & Rock, C. D. Abscisic acid signaling in seeds and seedlings. *Plant Cell* **14** (Suppl), S15–S45 (2002).
5. Koornneef, M., Bentsink, L. & Hilhorst, H. Seed dormancy and germination. *Curr. Opin. Plant Biol.* **5,** 33–36 (2002).
6. Roelfsema, M. R. G. & Hedrich, R. In the light of stomatal opening—new insights into 'the Watergate'. *New Phytol.* **167,** 665–691 (2005).
7. Nemhauser, J. L., Hong, F. & Chory, J. Different plant hormones regulate similar processes through largely nonoverlapping transcriptional responses. *Cell* **126,** 467–475 (2006).
8. Shen, Y. Y. *et al.* The Mg-chelatase H subunit is an abscisic acid receptor. *Nature* **443,** 823–826 (2006).
9. Liu, X. *et al.* G protein-coupled receptor is a plasma membrane receptor for the plant hormone abscisic acid. *Science* **315,** 1712–1716 (2007).
10. Pandey, S., Nelson, D. C. & Assmann, S. M. Two novel GPCR-type G proteins are abscisic acid receptors in *Arabidopsis. Cell* **136,** 136–148 (2009).
11. Ma, Y. *et al.* Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324,** 1064–1068 (2009).
12. Park, S. Y. *et al.* Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324,** 1068–1071 (2009).
13. Santiago, J. *et al.* Modulation of drought resistance by the abscisic acid-receptor PYL5 through inhibition of clade A PP2Cs. *Plant J.* 10.1111/j.1365–313X.2009.03981.x (2009).
14. Li, S., Assmann, S. M. & Albert, R. Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. *PLoS Biol.* **4,** e312 (2006).
15. McCourt, P. & Creelman, R. The ABA receptors–we report you decide. *Curr. Opin. Plant Biol.* **11,** 474–478 (2008).
16. Pennisi, E. Stressed out over a stress hormone. *Science* **324,** 1012–1013 (2009).
17. Mustilli, A. C., Merlot, S., Vavasseur, A., Fenzi, F. & Giraudat, J. *Arabidopsis* OST1 protein kinase mediates the regulation of stomatal aperture by abscisic acid and acts upstream of reactive oxygen species production. *Plant Cell* **14,** 3089–3099 (2002).
18. Yoshida, R. *et al.* ABA-activated SnRK2 protein kinase is required for dehydration stress signaling in *Arabidopsis. Plant Cell Physiol.* **43,** 1473–1483 (2002).
19. Yoshida, R. *et al.* The regulatory domain of SRK2E/OST1/SnRK2.6 interacts with ABI1 and integrates abscisic acid (ABA) and osmotic stress signals controlling stomatal closure in *Arabidopsis. J. Biol. Chem.* **281,** 5310–5318 (2006).
20. Fujii, H., Verslues, P. E. & Zhu, J. K. Identification of two protein kinases required for abscisic acid regulation of seed germination, root growth, and gene expression in *Arabidopsis. Plant Cell* **19,** 485–494 (2007).
21. Fujii, H. & Zhu, J. K. *Arabidopsis* mutant deficient in three abscisic acid-activated protein kinases reveals critical roles in growth, reproduction and stress. *Proc. Natl Acad. Sci. USA* **106,** 8380–8385 (2009).
22. Johnson, R. R., Wagner, R. L., Verhey, S. D. & Walker-Simmons, M. K. The abscisic acid-responsive kinase PKABA1 interacts with a seed-specific abscisic acid response element-binding factor, TaABF, and phosphorylates TaABF peptide sequences. *Plant Physiol.* **130,** 837–846 (2002).
23. Furihata, T. *et al.* Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. *Proc. Natl Acad. Sci. USA* **103,** 1988–1993 (2006).
24. Belin, C. *et al.* Identification of features regulating OST1 kinase activity and OST1 function in guard cells. *Plant Physiol.* **141,** 1316–1327 (2006).
25. Rubio, S. *et al.* Triple loss of function of protein phosphatases type 2C leads to partial constitutive response to endogenous abscisic acid. *Plant Physiol.* **150,** 1345–1355 (2009).
26. Kuhn, J. M. & Schroeder, J. I. Impacts of altered RNA metabolism on abscisic acid signaling. *Curr. Opin. Plant Biol.* **6,** 463–469 (2003).
27. Hirayama, T. & Shinozaki, K. Perception and transduction of abscisic acid signals: keys to the function of the versatile plant hormone ABA. *Trends Plant Sci.* **12,** 343–351 (2007).
28. Yoo, S.-D., Cho, Y.-H. & Sheen, J. *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nature Protocols* **2,** 1565–1572 (2007).

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

Acknowledgements We thank R. Bressan for helpful discussions and for editing of the manuscript. This work was supported by National Institutes of Health grants (J.-K.Z.) and MICIIN and CSIC fellowships (S.R. and R.A.).

Author Contributions H.F. contributed Figs 2a–c, e, 3, 4a–c and Supplementary Fig. 4. V.C. contributed Fig. 1a–d and Supplementary Fig. 1a, b. A.R., S.R., R.A. and P.L.R. contributed Fig. 2d and Supplementary Figs 2 and 3. S.-Y.P. and S.R.C. assisted with the generation of recombinant proteins, and S.R.C. helped edit the manuscript. J.S. assisted with protoplast assays. J.-K.Z. designed the experiments, and wrote the paper together with V.C. and H.F.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.-K.Z. (jian-kang.zhu@ucr.edu).

METHODS

Plant culture. For protoplast isolation, *Arabidopsis thaliana* ecotype Columbia and *snrk2.2/2.3/2.6* triple mutant plants were grown on Jiffy7 soil (Jiffy Products) in an environment-controlled chamber (illumination of 75 μmol m^{-2} s^{-1}, photoperiod of 13 h light/11 h dark, 22 °C).

Plasmid constructs for protoplast transient assay. The RD29B promoter region[11] was amplified by PCR from Columbia genomic DNA and cloned by replacing the CBF3 promoter in the CBF3–LUC protoplast expression vector. CBF3–LUC, UQ10–GUS, ABI1 and ABI1^{G180D} protoplast expression plasmid vectors were provided by J. Sheen. ABI1 was replaced with His–PYR1/His–PYLs/His–PYR^{P88S}, HAB1–Myc, ABF2–haemagglutinin (HA)/ABF2 mutant versions, SnRK2–Flag/SnRK2.6^{K50N} sequences. All the plasmids were sequenced to confirm the sequence and avoid cloning errors. The GUS reporter plasmid was used as an internal control to normalize transfection efficiency in protoplast assays. Primers used for preparing protoplast expression vectors were as follows: PRD29B-F, 5′-CGGGATCCGTTGAATCTTGCGGAAGCA-3′; PRD29B-R, 5′-CATGCCATGGTTCAAGTGAATCAATCAAAC-3′; HAB1-F, 5′-CATGCCATGGAGGAGATGACTCCCGCAG-3′; HAB1-R1, 5′-CAGATCCTCCTCAGAAATCAGCTTTTGCTCAGGCCTGGTTCTGGTCTTGAACTTTC-3′; HAB1-R2, 5′-GCCTGCAGTCATAAGTCTTCTTCGCTTATTAATTTCTGTTCCAGATCCTCCTCAGAAA-3′; ABF2-F, 5′-CGGGATCCATGGATGGTAGTATGAATTTGG-3′; ABF2-R, 5′-GTCAAGGCCTCCAAGGTCCCGACTCTGTCCT-3′; SnRK2.2-F, 5′-CATGCCATGGATCCGGCGACTAATTCAC-3′; SnRK2.2–Flag-R, 5′-GTCCTTGTAGTCAGAAGGCCTGAGAGCATAAACTATCTC-3′; SnRK2.2–Flag–Pst-R, 5′-GCCTGCAGTCACTTGTCATCGTCGTCCTTGTAGTCAGAAG-3′.

In the SnRK2.2–Flag plasmid, SnRK2.2 was replaced with SnRK2.3 or SnRK2.6: SnRK2.3-F, 5′-CATGCCATGGATCGAGCTCCGGTGACCA-3′; SnRK2.3-R, 5′-GTCAAGGCCTGAGAGCGTAAACTATCTCTCCGCT-3′; SnRK2.6-F, 5′-CATGCCATGGATCGACCAGCAGTGAGT-3′; SnRK2.6-R, 5′-GTCAAGGCCTCATTGCGTACACAATCTCTC-3′.

His–PYR1 was constructed with following primers: His–PYR1-F1, 5′-CACCATCACCCATGGATGCCTTCGGAGTTAACACCAG-3′; His–PYR1-F2, 5′-CGGGATCCATGAAACATCACCATCACCATCACCCATGGATG-3′; PYR1-R, 5′-GTCAAGGCCTTCACGTCACCTGAGAACCACTTC-3′.

In His–PYR1 plasmid, PYR1 was replaced with PYLs with following primers: PYL1-F, 5′-CATGCCATGGATGGCGAATTCAGAGTCCTCCTC-3′; PYL1-R, 5′-GTCAAGGCCTTTACCTAACCTGAGAAGAGTTG-3′; PYL2-F, 5′-CATGCCATGGATGAGCTCATCCCCGGCCGTG-3′; PYL2-R, 5′-GTCAAGGCCTTTATTCATCATCATGCATAGGTG-3′; PYL3-F, 5′-CATGCCATGGATGAATCTTGCTCCAATCCATG-3′; PYL3-R, 5′-GTCAAGGCCTTCAGGTCGGAGAAGCCGTGGA-3′; PYL4-F, 5′-CATGCCATGGATGCTTGCCGTTCACCGTCCTTC-3′; PYL4-R, 5′-GTCAAGGCCTTCACAGAGACATCTTCTTCTTG-3′; PYL5-F, 5′-CATGCCATGGATGAGGTCACCGGTGCAACTC-3′; PYL5-R, 5′-GTCAAGGCCTTTATTGCCGGTTGGTACTTCGAG-3′; PYL6-F, 5′-CATGCCATGGATGCCAACGTCGATACAGTTTC-3′; PYL6-R, 5′-GTCAAGGCCTTTACGAGAATTTAGAAGTGTTCTC-3′; PYL7-F, 5′-CATGCCATGGATGGAGATGATCGGAGGAGAC-3′; PYL7-R, 5′-GTCAAGGCCTTCAAAGGTTGGTTTCTGTATGA-3′; PYL8-F, 5′-CATGCCATGGATGGAAGCTAACGGGATTGAGAAC-3′; PYL8-R, 5′-GTCAAGGCCTTTAGACTCTCGATTCTGTCGTG-3′; PYL9-F, 5′-CATGCCATGGATGATGGACGGCGTTGAAGG-3′; PYL9-R, 5′-GTCAAGGCCTTCACTGAGTAATGTCCTGAGAAG-3′; PYL10-F, 5′-CATGCCATGGATGAACGGTGACGAAACAAAGAAG-3′; PYL10-R, 5′-GTCAAGGCCTTCATATCTTCTTCTCCATAGATTCTG-3′; PYL11-F, 5′-CATGCCATGGATGGAAACTTCTCAAAAATATC-3′; PYL11-R, 5′-GTCAAGGCCTTTACAACTTTAGATGAGCCACC-3′; PYL12-F, 5′-CATGCCATGGATGAAAACATCTCAAGAACAGCA-3′; PYL12-R: 5′-GTCAAGGCCTTTAAGTGAGCTCCATCATCTTC-3′.

***In vitro* mutagenesis.** ABF2–HA plasmid was used for *in vitro* mutagenesis by using the GeneTailor Site-Directed Mutagenesis System (catalog no. 12397-014; Invitrogen) with the following mutagenic primers: ABF2S26-R, 5′-ACCTTGTCTAGTCAACCCTCCACCTCCACCAC-3′; ABF2S26A-F, 5′-AGGGTTGACTAGACAAGGTGCTATATACTCGTTG-3′; ABF2S26D-F, 5′-AGGGTTGACTAGACAAGGTGATATATACTCGTTG-3′; ABF2S86-R, 5′-GCCTTGCCTCTGCAGCTGCAAACCCTCTTG-3′; ABF2S86A-F, 5′-CAGCTGCAGAGGCAAGGCGCTTTGACTCTGCCTC-3′; ABF2S86D-F, 5′-CAGCTGCAGAGGCAAGGCGATTTGACTCTGCCTC-3′; ABF2S94-R, 5′-AAGCGTTCGAGGCAGAGTCAACGAGCCTTG-3′; ABF2S94A-R1, 5′-AAGCGTTCGAGGCAGAGTCAAAGCGCCTTG-3′; VC-ABF2S94D-R1, 5′-AAGCGTTCGAGGCAGAGTCAAATCGCCTTG-3′; ABF2S94A-F, 5′-ACTCTGCCTCGAACGCTTGCTCAGAAGACGGTTG-3′; ABF2S94D-F, 5′-ACTCTGCCTCGAACGCTTGATCAGAAGACGGTTG-3′; ABF2T135-R, 5′-TTGCTGCCTCTGACTCTGACTCTGACTCTG-3′; ABF2T135A-F, 5′-TCAGAGTCAGAGGCAGCAAGCTTTAGGTGAAGTA-3′; ABF2T135D-F, 5′-TCAGAGTCAGAGGCAGCAAGATTTAGGTGAAGTA-3′.

Protoplast isolation and transactivation assay. Plants about 4 weeks old were used for protoplast isolation as described previously (http://genetics.mgh.harvard.edu/sheenweb)[28]. All the chemicals used in protoplast isolation were obtained from Sigma. Leaf strips about 0.5–1 mm in width cut from the middle part of leaves were vacuum infiltrated with enzyme solution containing 20 mM MES pH 5.7, 1.5% (w/v) cellulase R10 (Yakult Pharmaceutical Industry), 0.4% (w/v) macerozyme R10 (Yakult Pharmaceutical Industry), 0.4 M mannitol, 20 mM KCl, 10 mM CaCl$_2$, 1 mM 2-mercaptoethanol (optional) and 0.1% BSA. After infiltration for 30 min, leaf strips in enzyme solution were incubated in the dark for 3 h. Protoplasts were diluted with W5 (2 mM MES pH 5.7 containing 154 mM NaCl, 125 mM CaCl$_2$ and 5 mM KCl) solution to a final concentration of 2×10^5 cells ml^{-1}. After 30 min resting, W5 solution was removed and protoplasts were resuspended in MMg solution (4 mM MES pH 5.7 containing 0.4 M mannitol and 15 mM MgCl$_2$). For transfection, 100 μl of protoplasts (2×10^4 cells) were mixed with plasmid constructs (in 10 μl) and 110 μl of PEG solution (40% w/v PEG-4000 in doubly distilled water containing 0.2 M mannitol and 100 mM CaCl$_2$). After incubation for 10 min, transfection was stopped by the addition of 440 μl of W5 solution. Protoplasts were harvested by centrifugation at 100*g* for 2 min at room temperature (22–25 °C) and resuspended in about 100 μl of W5 solution. Protoplasts were then incubated in WI solution (4 mM MES pH 5.7 containing 0.5 M mannitol and 20 mM KCl) with 0 or 5 μm (+)-ABA for 5 h. Protoplasts were harvested by centrifugation at 100*g* for 2 min, frozen in liquid N$_2$ and stored at −80 °C until further use. The frozen protoplasts were resuspended in 100 μl of protoplast lysis buffer (2.5 mM Tris-phosphate pH 7.8 containing 1 mM dithiothreitol, 2 mM DACTAA, 10% (v/v) glycerol and 1% (v/v) Triton X-100). LUC mix (Promega) (100 μl) was added to 10 μl of the lysate, and LUC activity was measured with a plate reader (Wallac VICTOR2 plate reader). Protoplast lysate (2 μl) was mixed with 10 μl of 4-methylumbelliferyl β-D-glucuronide (MUG) substrate mix (10 mM Tris-HCl pH 8 containing 1 mM MUG and 2 mM MgCl$_2$) and incubated for 30 min at 37 °C. The reaction was stopped by the addition of 100 μl of 0.2 M Na$_2$CO$_3$, and the fluorescence of 4-methylumbelliferone) was quantified with a plate reader (Wallac VICTOR2 plate reader).

The RD29B promoter fused with the LUC coding sequence was used as an ABA-responsive reporter gene (7 μg of plasmid per transfection). *ZmUBQ*::*GUS* was included in each sample as an internal control (3 μg per transfection). ABF2–HA and its mutant versions, SnRK2.6–Flag and SnRK2.6^{K50N}–Flag, His–PYR/PYLs and HAB1–Myc plasmid constructs were used at 3 μg per transfection, and ABI1 was used at 2 μg per transfection.

***In vitro* phosphatase and kinase assays.** MBP–SnRK2.6 (1 μg), MBP (1 μg) or Flag–SnRK2.6 on beads was incubated for 20 min with GST–ABI1 (1 μg), GST (1 μg), His–PYR1 (1 μg) and/or His–PYR1^{P88S} (1 μg), as indicated, in 40 μl of phosphatase buffer (50 mM Tris-HCl pH 7.0, 60 mM magnesium acetate, 0.1 mM EGTA and 0.1% 2-mercaptoethanol) containing the indicated concentration of (+)-ABA (Biosynth AG) at 30 °C. After removal of the solution, GST–ABF2 fragment (10 μg) was added. The reaction mixture (20 μl) contained 25 mM Tris pH 7.4, 12 mM MgCl$_2$, 1 μM ATP, 5 μCi of [γ-^{32}P]ATP, 1 mM dithiothreitol, 1 mM Na$_3$VO$_4$ and 5 mM NaF. The reaction mixtures were incubated for 30 min at 30 °C. The reaction was stopped by the addition of Laemmli sample buffer, and the mixtures were then subjected to SDS–PAGE. Radioactivity was detected with a Typhoon phosphoimager (GE Healthcare). SnRK2.6 kinase deactivation assays (Fig. 2d) were performed by previous incubation with protein phosphatase for 30 min at room temperature. Assays to test the recovery of SnRK2.6 activity were performed by incubation of the protein phosphatase for 10 min with the receptors in the presence of 1 μM (+)-ABA.

To measure phosphate release from synthetic phosphopeptide, the phosphopeptide (10 nmol) from Biomatik was incubated for 10 min with 1 μg of ABI1 or GST in phosphatase buffer (40 μl) at 30 °C. Released phosphate was measured as described previously[29].

Construction of plasmids for yeast two-hybrid and triple-hybrid analysis. The coding sequence of SnRK2.6, SnRK2.2 and SnRK2.3 was amplified by PCR with the following pairs of primers: F4g33950 (5′-ATGGATCGACCAGCAGTGAGT-3′) and R4g33950 (5′-TTTGTCGACTCACATTGCGTACACAATCT-3′), F3g50500 (5′-ATGGATCCGGCGACTAATTCA-3′) and R3g50500 (5′-TTTGTCGACTCAGAGAGCATAAACTATCT-3′), and F5g66880 (5′-ATGGATCGAGCTCCGGTGACC-3′) and R5g66880 (5′-TTTGTCGACTTAGAGAGCGTAAACTATCT-3′), respectively.

The PCR product was initially cloned into pCR8/GW/TOPO, double-digested with EcoRI–SalI and cloned into pGBT9, where the kinase coding sequences were fused to the GAL4 DNA-binding domain (GBD). pGBT9 constructs in which SnRK2s acted as baits were faced in two-hybrid assays with the following pGADT7 constructs: ABI1, ABI2 and HAB1, where the PP2C coding sequences were fused to the GAL4 activation domain (GAD).

To perform triple-hybrid experiments, the SnRK2.6 EcoRI–SalI fragment described above was cloned into the pBridge vector (Clontech), in which the kinase coding sequence is fused to GBD. Next, the coding sequences of PYL8 and PYL5 were cloned into the NotI site of pBridge–SnRK2.6. To reproduce the yeast two-hybrid interaction described above, yeast host AH109 was co-transformed with one of the following plasmids, pGADT7–ABI1/pGADT7–ABI2/pGADT7–HAB1, encoding GAD–PP2C fusions, and pBridge–SnRK2.6, encoding a GBD–SnRK2.6 fusion; to test the interference of ABA receptors on this interaction, pBridge–SnRK2.6 + PYL8 and pBridge–SnRK2.6 + PYL5 were employed.

BiFC. Experiments were performed as described previously[13]. Constructs were made in the pSPYNE-35S[30] as well as the Gateway vector pYFP^{C}43 (provided by A. Ferrando). The ABI1 coding sequence was excised from pCR8/GW/TOPO as a BamHI–SmaI fragment and cloned into pSPYNE-35S. The coding sequences of SnRK2.6, SnRK2.2, SnRK2.3 and ΔCSnRK2.6, previously cloned into pCR8/GW/TOPO, were recombined by LR reaction (recombination reaction between an *attL* attachment site present in the entry vector and *attR* present in the destination vector, catalysed by a host of recombination proteins) into pYFP^{C}43 destination vector.

Immunoprecipitation. Protein extracts for immunodetection experiments were prepared from *Nicotiana benthamiana* leaves infiltrated with *Agrobacterium tumefaciens* C58C1 (pGV2260) transformed with different fusion proteins containing SnRK2.6/OST1, the C-terminal deletion of SnRK2.6, the amino-terminal deletion, and full-length ABI1. Protein extracts from the transfected plants were used for immunoprecipitation at 4 °C with super-paramagnetic MACS microbeads coupled to monoclonal anti-c-Myc antibody (Miltenyi Biotec). The antigen–antibody complex was eluted, fractionated on an 8% SDS–PAGE gel and transferred to Immobilon-P membranes (Millipore), and probed either with monoclonal anti-c-Myc (clone 9E10; Roche) or anti-GFP^C (clone JL-8; Clontech) as primary antibodies and enhanced chemiluminescence anti-mouse peroxidase (GE Healthcare) as secondary antibody. Detection was performed with the enhanced chemiluminescence advance western blotting detection kit (GE Healthcare).

To immunoprecipitate SnRK2.6, 3-week-old seedlings of transgenic plants expressing Flag–SnRK2.6 treated with 100 μM ABA were used. Protein extracts were incubated for 2 h with anti-Flag antibody beads (Sigma) at 4 °C. The beads were washed six times with pre-chilled extraction buffer.

29. Van Veldhoven, P. P. & Mannaerts, G. P. Inorganic and organic phosphate measurements in the nanomolar range. *Anal. Biochem.* **161**, 45–48 (1987).
30. Walter, M. *et al.* Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. *Plant J.* **40**, 428–438 (2004).

Vol 462 | 3 December 2009 | **doi:10.1038/nature08591**

The abscisic acid receptor PYR1 in complex with abscisic acid

Julia Santiago[1]*, Florine Dupeux[2]*, Adam Round[2], Regina Antoni[1], Sang-Youl Park[3], Marc Jamin[4], Sean R. Cutler[3], Pedro Luis Rodriguez[1] & José Antonio Márquez[2]

The plant hormone abscisic acid (ABA) has a central role in coordinating the adaptive response in situations of decreased water availability as well as the regulation of plant growth and development. Recently, a 14-member family of intracellular ABA receptors, named PYR/PYL/RCAR[1–3], has been identified. These proteins inhibit in an ABA-dependent manner the activity of a family of key negative regulators of the ABA signalling pathway: the group-A protein phosphatases type 2C (PP2Cs)[4–6]. Here we present the crystal structure of *Arabidopsis thaliana* PYR1, which consists of a dimer in which one of the subunits is bound to ABA. In the ligand-bound subunit, the loops surrounding the entry to the binding cavity fold over the ABA molecule, enclosing it inside, whereas in the empty subunit they form a channel leaving an open access to the cavity, indicating that conformational changes in these loops have a critical role in the stabilization of the hormone–receptor complex. By providing structural details on the ABA-binding pocket, this work paves the way for the development of new small molecules able to activate the plant stress response.

ABA is currently recognized as the pivotal plant hormone that coordinates the complex regulatory network required to cope with water stress[7]. During drought or salt stress, ABA levels in the plant can rise up to 40-fold, triggering the closure of stomata as well as the accumulation of osmocompatible solutes, dehydrins and LEA proteins[7], all of which contribute to the avoidance and tolerance of water stress. ABA can also stimulate root growth to improve water uptake[7]. Group-A protein phosphatases 2C (PP2Cs), such as ABI1, ABI2, HAB1 and PP2CA, are key negative regulators of ABA signalling, and have a critical role in this response[4–6]. Recently, two G-protein-coupled receptors (GPCRs), GTG1 and GTG2, have been identified as plasma-membrane-localized ABA receptors[8]. However, deletion of both proteins impairs, but does not completely abolish, the ABA response, suggesting the existence of alternative ABA receptors[9]. The existence of intracellular ABA receptors was recently confirmed by the discovery of the PYR/PYL/RCAR proteins[1–3]. This protein family contains 14 members in *Arabidopsis thaliana*, some of which show dual cytosolic and nuclear localization, and are able to inhibit the activity of ABI1, ABI2 and HAB1 in an ABA-dependent manner[1–3]. In the absence of PP2Cs, RCAR1 (also known as PYL9) and PYL5 bind ABA with a dissociation constant (K_d) of 0.66 μM and 1.1 μM, respectively; however, in the presence of a PP2C, the affinity increases by more than tenfold, showing a K_d in the nanomolar range (64 nM for the RCAR1–ABI2 pair[2], and 38 nM for PYL5–ΔNHAB1[3]). This has prompted some authors to propose that PYR/PYL/RCAR and PP2Cs could act as co-receptors. The PYR/PYL/RCAR proteins are predicted to contain a Bet v I domain, a versatile scaffold for the binding of hydrophobic ligands[10,11]. Here we describe the crystal structure of the PYR1 protein in complex with ABA at 2.0 Å resolution and elucidate the interactions that stabilize the hormone–receptor complex.

Gel filtration coupled to multi-angle laser light scattering (MALLS)[12] and small angle X-ray scattering analysis show that *Arabidopsis thaliana* PYR1 is a dimer both in the presence and absence of ABA (see Supplementary Fig. 1). MALLS analysis also indicates that the interaction between PYR1 and the catalytic core of the protein phosphatase HAB1 (ΔNHAB1) occurs only in the presence of the hormone with a stoichiometry of 1:1 (see Supplementary Fig. 1). Because PYR1 does not interact with HAB1 in the absence of ABA, our MALLS data indicate that the PYR1 dimer dissociates upon ABA-promoted interaction with the phosphatase. *Arabidopsis thaliana* PYR1 crystals were obtained in the presence of 1 mM (+)-ABA and X-ray diffraction data were collected and refined to 2.0 Å resolution (see Methods for a description of the experimental procedures and Supplementary Table 1 for data collection and refinement statistics). The refined model ($R_{\mathrm{work}}/R_{\mathrm{free}} = 20.0/23.4$) contains two similar copies of the PYR1 dimer in the crystal asymmetric unit (see Supplementary Information).

The structure of the PYR1 monomer consists of a Bet v I fold[10,13] plus a 27-amino-acid amino-terminal extension that contains two short helical segments (Fig. 1). The PYR1 Bet v I domain consists of

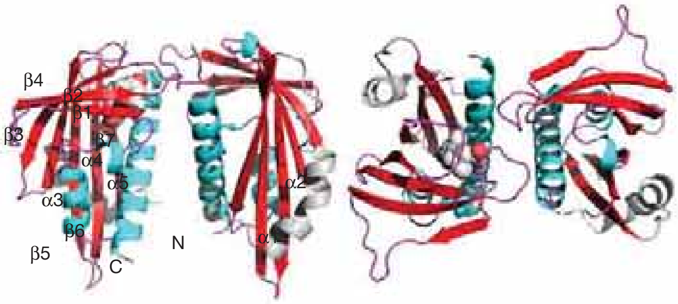


Figure 1 | The structure of the PYR1 dimer. Two different views of the PYR1 dimer are shown. The region corresponding to the Bet v I fold is coloured according to secondary structure (strands in red, helices in cyan and loops in magenta). The 27-amino-acid N-terminal extension containing two short helical segments is depicted in grey. Secondary structure elements as well as the N and C termini are indicated (as in Supplementary Fig. 3). One of the subunits in each of the dimers contains an (+)-ABA molecule (atoms represented as spheres) in the binding pocket, whereas the second subunit is empty.

[1]Instituto de Biología Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas-Universidad Politécnica de Valencia, ES-46022 Valencia, Spain. [2]European Molecular Biology Laboratory, Grenoble Outstation and Unit of Virus Host-Cell Interactions, UJF-EMBL-CNRS, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France. [3]Department of Botany and Plant Sciences. Center for Plant Cell Biology. University of California, Riverside, California 92521, USA. [4]Virus Host Cell Interactions (UVHCI) UMI 3265, Université Joseph Fourier-EMBL-CNRS, 6 rue Jules Horowitz, 38042 Grenoble Cedex 9, France.
*These authors contributed equally to this work.

seven strands forming a bent β-sheet and producing a central cavity that resembles that of a folded hand. The walls of the cavity are completed by a long carboxy-terminal α-helix (α5) juxtaposed to the β-sheet. Two small helices (α3 and α4) between strands β1 and β2 close the bottom side of the cavity. The dimerization region involves residues in loops α4–β2 and β3–β4 and the N-terminal part of helix α5 (see Fig. 1 and Supplementary Fig. 3).

A continuous electron density cloud that could not be attributed to either protein or solvent molecules was observed inside the cavity of one of the subunits of each of the dimers (see Fig. 2a). This density could be unambiguously interpreted as a molecule *S*-2-*cis*-4-*trans* abscisic acid. The structure of the PYR1-bound (+)-ABA is similar to that of crystalline ABA[14]. The carboxylic and diene groups, in the isoprenoid moiety, are found in a plane which is orthogonal to the plane formed by the cyclohexenone ring. The cyclohexenone ring is puckered, with the C6′ carbon below the plane of the ring.

ABA is almost completely buried inside the PYR1 cavity and stabilized by a series of interactions that perfectly match the polar and hydrophobic character of its different functional groups (see Fig. 2). Most of the polar interactions between the protein and ABA are mediated by water molecules. Only two direct polar contacts are found and they involve the amine group of Lys 59 and the backbone amide of Ala 89, which interact with ABA's carboxylic and ketone groups, respectively. A network of water molecules is very favourably coordinated between the hydroxyl and carboxylic groups of the ABA molecule and the side chains of Glu 94, Tyr 120, Ser 122 and Glu 141 (see Fig. 2). Adding to this, the backbone amide of Arg 116 establishes a water-mediated hydrogen bond with the oxygen in the ketone group of ABA. This set of ABA-contacting residues is invariant in the PYR/PYL/RCAR protein family with the exception of Lys 59, which seems to be replaced by glutamine in PYL13 (see Supplementary Fig. 3). The functional relevance of these contacts is supported by genetic studies which show that the E94K and E141K substitutions confer resistance to pyrabactin[1], a selective ABA agonist.

Stereospecificity in the hormone–receptor interaction seems to be contributed mostly by steric constraints imposed by hydrophobic residues that conform the upper part of the PYR1 cavity to match accurately the van der Waals surface of the (+)-ABA molecule rather than through polar interactions (Fig. 2b). The mono-methyl group at position C2' is confined in a tight space between Phe 61, Val 163 and Val 83, while the cavity is significantly wider at the opposite side, where the much larger dimethyl group at C6' finds an optimal fitting. Phe 61 is strictly conserved whereas Val 163 and Val 83 are highly conserved or replaced by bulkier isoleucine and leucine residues in other PYL proteins (see Supplementary Fig. 3). The cavity in PYR1 is significantly larger than ABA, which occupies its upper part (see Fig. 3). The side chains of Lys 59 and Arg 79 seem to limit the access of the ligand to the lower part of the cavity, which contains a number of water molecules. Ligand binding studies reported here and elsewhere indicate that PYR/PYL/RCAR proteins have a stoichiometry of ABA-binding close to 1:1 both for isolated receptors as well as in the presence of the phosphatases, suggesting that in solution both subunits are able to bind the hormone. Therefore, it is likely that the alternative configuration observed is favoured by crystal packing.

The overall structure of the two subunits in the PYR1 dimer is very similar; however, notable differences appear in residues 84–89, 113–118

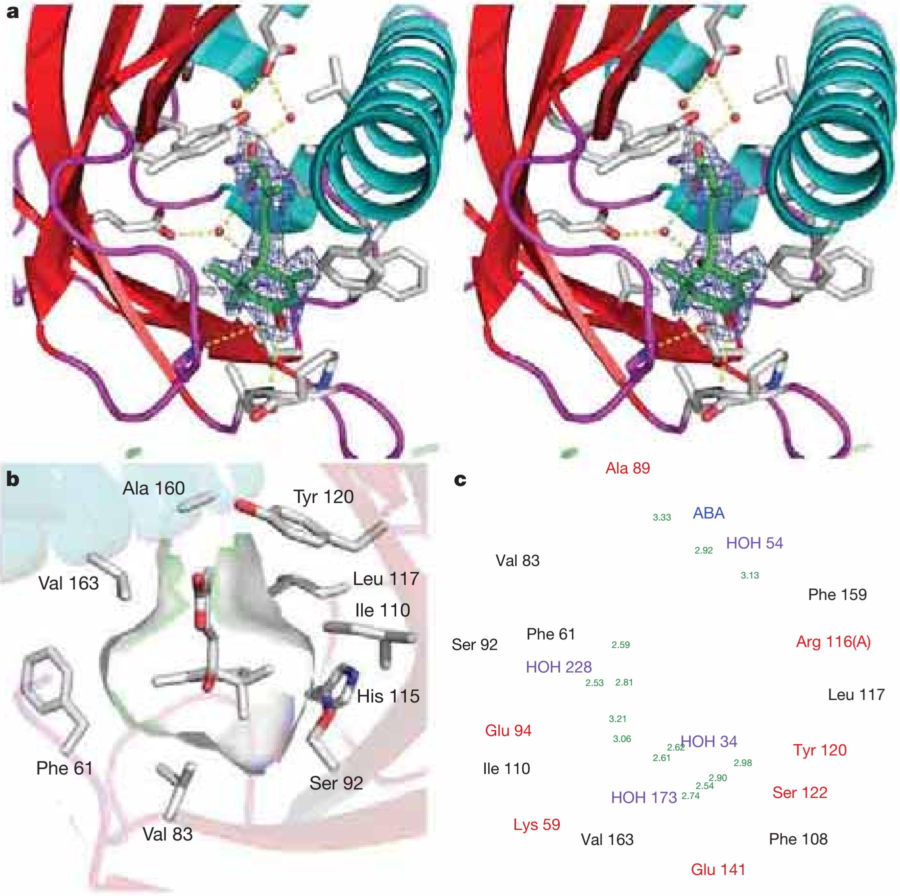


Figure 2 | Abscisic acid binding. a, Stereo representation of the ABA binding site showing omit electron density (blue mesh) around the ligand region. Abscisic acid is represented as sticks. Major polar interactions between ABA and the protein are indicated (yellow dotted lines). **b**, The protein surface in the PYR1 cavity conforms tightly to the shape of the ABA molecule (centre) favouring binding of the (+) enantiomer. **c**, Detailed description of the interactions generated with the program Ligplot. Hemispheres represent hydrophobic interactions whereas lines represent polar interactions. HOH indicates water molecules. The yellow ribbon around the ABA molecule indicates low solvent accessibility.

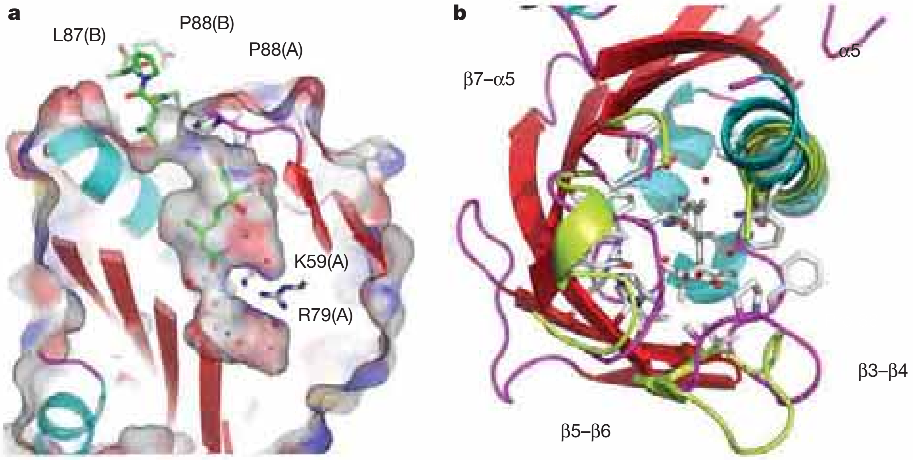


Figure 3 | Conformational changes in the loops surrounding the ABA binding cavity. a, The PYR1 ligand-binding cavity is shown. Key amino acids are labelled; the letters in parenthesis refer to subunits A or B of the dimer. Water molecules are indicated as red spheres. **b**, The ligand-bound and free forms of PYR1 have been superimposed. The hormone-bound form is coloured according to secondary structure (as in Fig. 1). The corresponding loops of the hormone-free molecule and the C-terminal part of helix α5 showing a different conformation are displayed in light green. The positions of loops β3–β4, β5–β6 and the N-terminal region of α5 are highlighted. The approximate positions of loop β7–α5 (disordered in the structure) are indicated by dotted lines. In the ABA-bound form of PYR1, the loops surrounding the entry to the cavity fold over the ABA molecule, trapping it inside. The same loops in the unliganded form adopt an open conformation that leaves the cavity accessible to the solvent. The positions of residues Pro 88 and Ser 152 involved in interactions with the PP2Cs are indicated.

and 153–159 located in loops β3–β4 and β5–β6 and the N-terminal part of helix α5, respectively. These loops surround the upper part of the PYR1 cavity and in the ligand-bound form adopt a closed conformation folding over the ABA molecule and isolating it almost completely from the solvent as shown in Fig. 3. Notably, in the ABA-free subunit, these loops adopt an open conformation that creates a passage into the cavity. The closed conformation of the loops in the ABA-bound subunit is stabilized by hydrophobic and polar interactions between residues in the loops and the cyclohexenone moiety of the ABA molecule (see Fig. 3). The equivalent residues are exposed to the solvent in the ABA-free subunit. The β3–β4 loop (amino acids 84–89) of the empty subunit points towards the ABA-bound subunit rather than towards the inside of the cavity, with Leu 87 inserted in a hydrophobic pocket formed by loops β3–β4, β5–β6 and the N-terminal part of helix α5 of the ABA-bound molecule (see Fig. 3a). This probably stabilizes both the open conformation of this loop and the closed conformation of the ABA-bound molecule. The conformational changes of the loops flanking the cavity's entry probably have a role similar to a lid stabilizing the ligand inside the cavity, as the subunit showing the open conformation remains empty, despite the fact that the residues in the inner part of the cavity adopt a configuration favourable for ABA binding.

Point mutations P88S and S152L located in the β3–β4 and β7–α5 loops, respectively, cause a reduction in the capacity of PYR1 to interact with and inhibit the protein phosphatase HAB1[1]. Isothermal titration calorimetry measurements show that the P88S mutation does not impair ABA binding, but significantly decreases the stability of the ternary complex ABA–PYR1(P88S)–PP2C as compared to the wild type (see Supplementary Information). This suggests that the loop β3–β4 is not only important for the stabilization of the ligand, as the PYR1 structure suggests, but that it is also involved in interactions with the PP2Cs. Similarly, Ser 152 is located in the loop β7–α5, which is disordered in this structure, and does not contact the ABA molecule, indicating again that the loops surrounding the entry to the ABA binding cavity are required for direct interactions with the phosphatases. On the basis of the genetic, biochemical and structural data, we propose that the interactions between the phosphatase and lid in its closed conformation stabilize the ternary ABA–receptor–PP2C complex, which could also explain the increased affinity of PYR1 for ABA in the presence of the phosphatase. Alternatively, the dissociation of the PYR1 dimer and the formation of the PYR1–phosphatase complex could lead to a reorganization of the PYR1 ABA-binding site.

The structure described here shows precise details on the specific interactions between ABA and PYR1, confirming the role of the PYR1 protein as an ABA receptor. Our observations also confirm that the major determinants of stereoselectivity are spatial constraints around positions 2 and 6 in the ring of the ABA molecule, rather than interactions involving polar groups, which as already suggested[15], may explain the striking biological activity of the unnatural (−)-ABA stereoisomer. The structural details of the ABA–receptor interactions described here pave the way for the development of agonist molecules that could be useful to cope with water stress in crops.

METHODS SUMMARY

PYR1, PYR1(P88S) and ΔNHAB1 were expressed and purified as described previously[3,16]. Crystallization conditions for PYR1 were identified at the High Throughput Crystallization Laboratory of the EMBL Grenoble Outstation (https://htxlab.embl.fr). The crystals used for data collection were obtained by the vapour diffusion method with a protein concentration of 25 mg ml^{-1} in the presence of 0.2 M sodium magnesium acetate, 0.1 M sodium cacodylate pH 6.5 and 18% PEG 8000 and 1 mM (+)-ABA and were frozen with 20% glycerol as cryo-protectant. X-ray data were collected at the ID14eh4 beamline of the ESRF. XDS[17], Coot[18] and REFMAC[19] programs were used for crystallographic analysis. Initial phases were obtained by the molecular replacement method with the model of the XoxI protein (3CNW) and the program Phaser[20]. Small angle X-ray scattering data were collected at the BioSAXS station (ID14eh3) of the ESRF. Samples were exposed using 30-μl volumes loaded into a quartz capillary mounted in vacuum and moved through the beam during exposure to minimize the effect of radiation damage. The scattering from the buffer was measured before and after each sample measurement and the average scattering used for background subtraction with PRIMUS[21]. The theoretical scattering curves from the three possible dimeric PYR1 ensembles were calculated and fitted to the merged experimental scattering curve for PYR1 using the program CRYSOL[22]. MALLS and isothermal titration calorimetry experiments were carried out as described previously[3,12]. Figures were generated with Pymol (DeLano Scientific LLC) and Ligplot[23].

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

Received 25 September; accepted 20 October 2009.

1. Park, S. Y. *et al.* Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324**, 1068–1071 (2009).
2. Ma, Y. *et al.* Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324**, 1064–1068 (2009).

3. Santiago, J. *et al.* Modulation of drought resistance by the abscisic acid-receptor PYL5 through inhibition of clade A PP2Cs. *Plant J.* doi:10.1111/j.1365-313X.2009.03981.x (16 July 2009).
4. Leung, J. *et al. Arabidopsis* ABA response gene ABI1: features of a calcium-modulated protein phosphatase. *Science* **264,** 1448–1452 (1994).
5. Meyer, K., Leube, M. P. & Grill, E. A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. *Science* **264,** 1452–1455 (1994).
6. Rubio, S. *et al.* Triple loss of function of protein phosphatases type 2C leads to partial constitutive response to endogenous abscisic acid. *Plant Physiol.* **150,** 1345–1355 (2009).
7. Verslues, P. E., Agarwal, M., Katiyar-Agarwal, S., Zhu, J. & Zhu, J. K. Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. *Plant J.* **45,** 523–539 (2006).
8. Pandey, S., Nelson, D. C. & Assmann, S. M. Two novel GPCR-type G proteins are abscisic acid receptors in *Arabidopsis. Cell* **136,** 136–148 (2009).
9. Christmann, A. & Grill, E. Are GTGs ABA's biggest fans? *Cell* **136,** 21–23 (2009).
10. Radauer, C., Lackner, P. & Breiteneder, H. The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol. Biol.* **8,** 286 (2008).
11. Iyer, L. M., Koonin, E. V. & Aravind, L. Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. *Proteins* **43,** 134–144 (2001).
12. Gerard, F. C. *et al.* Unphosphorylated rhabdoviridae phosphoproteins form elongated dimers in solution. *Biochemistry* **46,** 10328–10338 (2007).
13. Ponting, C. P. & Aravind, L. START: a lipid-binding domain in StAR, HD-ZIP and signalling proteins. *Trends Biochem. Sci.* **24,** 130–132 (1999).
14. Ueda, H. & Tanaka, J. The crystal and molecular structure of *dl*-2-*cis*-4-*trans*-abscisic acid. *Bull. Chem. Soc. Jpn* **50,** 1506–1509 (1997).
15. Milborrow, B. V. The chemistry and physiology of abscisic acid. *Annu. Rev. Plant Physiol.* **25,** 259–307 (1974).
16. Saez, A., Rodrigues, A., Santiago, J., Rubio, S. & Rodriguez, P. L. HAB1–SWI3B interaction reveals a link between abscisic acid signaling and putative SWI/SNF chromatin-remodeling complexes in *Arabidopsis. Plant Cell* **20,** 2972–2988 (2008).
17. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26,** 795–800 (1993).
18. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
19. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
20. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
21. Konarev, P. V., Volkov, V. V., Sololova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36,** 1277–1282 (2003).
22. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **28,** 768–773 (1995).
23. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8,** 127–134 (1995).

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

Acknowledgements We thank A. McArthy and S. Brockhauser for support during X-ray data collection and R. Serrano and S. Cusack for critical reading of the manuscript. We are grateful to the European Synchrotron radiation facility (ESRF) and the EMBL for access to the macromolecular crystallography and BioSAXS beamlines. This work was supported by grant BIO2008-00221 from Ministerio de Educación y Ciencia and Fondo Europeo de Desarrollo Regional and Consejo Superior de Investigaciones Científicas (fellowships to J.S. and R.A.). Access to the high throughput crystallization facility of the Partnership for Structural Biology in Grenoble (PSB) (https://htxlab.embl.fr) was supported by the European Community–Research Infrastructure Action PCUBE under the FP7 'Capacities' specific programme.

Author Contributions J.S. contributed with the cloning, protein purification, ITC, MALLS and helped with crystallization and SAXS experiments. F.D. performed protein purification, crystallization and crystal refinement experiments and helped with X-ray data collection. A.R. supervised SAXS data collection and performed data analysis. M.J. carried out MALLS experiments and analysis. R.A. carried out cloning and protein purification. S.-Y.P. and S.R.C. carried out cloning of mutant PYR1 proteins and contributed to discussions. P.L.R. contributed to discussions and writing of the manuscript. J.A.M. supervised the work and performed data collection, structure solution and refinement as well as writing of the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structure have been deposited in the Protein Data Bank under accession code 3K90. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.A.M. (marquez@embl.fr).

METHODS

Cloning, expression and purification of PYR1, PYR1(P88S) and ΔNHAB1 recombinant proteins. The ORF *Arabidopsis PYR1* was amplified by polymerase chain reaction (PCR) from a pUni clone obtained from ABRC, using FwAt4g17870 ACCATGGCTTCGGAGTTAACACC and RvAt4g17870 TCACG TCACCTGAGAACCACT primers and cloned to pCR/GW/TOPO vector. The coding sequence was excised using NcoI–EcoRI digestion and cloned into pETM11. PYR1(P88S) was produced as described previously[1]. The ΔNHAB1 construct was obtained according to ref. 3. BL21 (DE3) cells transformed with the corresponding pETM11 or pTE28 constructs were grown in 1 l of LB medium containing 50 μg ml^{-1} kanamycin to an optical density at 600 nm of 0.6–0.8. Then, 1 mM IPTG was added and induction was carried out over night at 20 °C. Cells were harvested 12 h after induction and stored at −80 °C before purification. Purification was carried out at 4 °C. A 2-l pellet was re-suspended in 100 ml of Lysis buffer (20 mM Tris pH 7.5, 300 mM NaCl, 15 mM Imidazol, 2 mM mercaptoethanol and Merk's protease cocktail inhibitor) and the cells were sonicated in a Misonix Sonifier. A cleared lysate was obtained after centrifugation at 20,000*g* for 45 min and protein was loaded to a 1-ml Ni HiTrap HP column using the AKTA Prime system. The protein extract loaded to the Ni HiTrap HP column was washed with 15 ml of wash buffer (20 mM Tris pH 7.5, 300 mM NaCl and 1 mM mercaptoethanol). The bound protein was then eluted with a gradient from 0–100% of elution buffer (20 mM Tris pH 7.5, 300 mM NaCl, 1 mM mercaptoethanol and 250 mM Imidazol). The fractions containing the protein were then dialysed for 2 h in 20 mM Tris pH 7.5, 150 mM NaCl, 1 mM $MnCl_2$ and 1 mM mercaptoethanol (dialysis buffer) at 4 °C, right after purification. The protein was then passed through a gel filtration step using a S200 column. Protein concentration was carried out with Amicon Ultra Centrifugal Filters (Millipore) and flash-frozen with liquid nitrogen. In the case of cleaved protein, the protein collected from the Ni purification was incubated with TEV protease, in a ratio of 1:100, for 2 h at 4 °C. Then, the protein was dialysed for 2 h in 20 mM Tris pH 7.5, 150 mM NaCl, 15 mM Imidazol and 1 mM mercaptoethanol at 4 °C. Second, Ni purification was carried out using a 1 ml Ni HiTrap HP column and an AKTA Purifier System. The cleaved protein was collected from the flow through. A gel filtration step, using a S200 column, was performed next with 20 mM Tris pH 7.5, 150 mM NaCl and 1 mM mercaptoethanol buffer. Concentration and flash freezing was carried out as described above. In the case of ΔNHAB1, 2 mM $MnCl_2$ was added to all the buffers.

Crystallography. Initial crystallization conditions for PYR1 were identified at the High Throughput Crystallization Laboratory of the EMBL Grenoble Outstation (https://htxlab.embl.fr). The crystals used for data collection were obtained by the vapour diffusion method with a protein concentration of 25 mg ml^{-1} in the presence of 0.2 M sodium magnesium acetate, 0.1 M sodium cacodylate pH 6.5, 18% PEG 8000 and 1 mM (+)-ABA, and were frozen with 20% glycerol as cryo-protectant. X-ray data were collected at the ID14eh4 beamline of the ESRF. XDS[17] was used for data processing. Initial phases were obtained by the molecular replacement method with the model of the XoxI protein (3CNW) deposited in the Protein Data Bank using the program Phaser. Model building and refinement was carried out using the programs Coot and REFMAC[19].

Small angle X-ray scattering experiments. Small angle X-ray scattering (SAXS) data were collected at the BioSAXS station (ID14-3) of the ESRF (http://www.esrf.fr/UsersAndScience/Experiments/MX/About_our_beamlines/ID14-3). ID14-3 is a fixed-energy (13.32 keV, $\lambda = 0.931$ Å) station optimized for solution scattering from biological macromolecules. Samples are exposed using 30-μl volumes loaded into a 2-mm quartz capillary mounted in vacuum using an automated robotic system which enables the sample to be moved through the beam during exposure to minimize the effect of radiation damage. The measurement cell is cleaned (using a commercial detergent for quartz capillaries), rinsed (with distilled water) and dried using compressed air after each measurement using the automated system to minimize cross contamination and dilution of the samples. Two-dimensional scattering images were collected using the Vantec2000 gas-filled detector from Bruker at a distance of 1.745 m from the sample. Standard data collection time of 5 min was used for all samples split into ten 30-s time frames to assess and remove effects from radiation damage to the samples. Individual time frames are processed automatically and independently by the data collection software (BsxCUBE) developed by the BLISS team at the ESRF, yielding individual radially averaged curves of normalized intensity versus scattering angle ($s = 4\pi \mathrm{Sin}\theta/\lambda$). Time frames are combined, excluding any data points which are affected by aggregation induced by radiation damage, to give the average scattering cure for each measurement. The scattering from the buffer alone was measured before and after each sample measurement and the average of the scattering before and after each sample is used for background subtraction using the program PRIMUS[21] from the ATSAS package developed by EMBL Hamburg. A range of PYR1 protein concentrations (1.3 to 7.1 mg ml^{-1}) was measured to assess and remove any concentration-dependant inter-particle effects. PYR1 was measured with 1 mM ABA and excess ABA and no significant differences in the scattering were observed with increasing ABA concentration. The theoretical scattering from the different dimer configurations was fitted to the merged scattering curve (free from effects of radiation damage and concentration-dependant inter-particle effects) for PYR1 with 1 mM ABA using the program CRYSOL[22].

Isothermal titration calorimetry (ITC) experiments were performed as described[3] using a VP-ITC calorimeter equipped with the control, data acquisition and analysis software ORIGIN 7 (Microcal).

SEC–MALL experiments. Multi-angle laser light scattering (MALLS) coupled to size exclusion chromatography was performed as follows. Size-exclusion chromatography (SEC) was carried out with an S200 Superdex column (GE Healthcare) equilibrated with 20 mM Tris-HCl, 150 mM NaCl and 1 mM β-mercaptoethanol buffer. On-line MALLS detection was performed with a DAWN-EOS detector (Wyatt Technology Corp.) using a laser emitting at 690 nm and by refractive index measurement using an RI2000 detector (Schambeck SFD). Weight-averaged molar masses (M_w) were calculated using the ASTRA software (Wyatt Technology Corp.)[12].

Vol 462 | 3 December 2009 | doi:10.1038/nature08615 nature

LETTERS

Hidden alternative structures of proline isomerase essential for catalysis

James S. Fraser[1], Michael W. Clarkson[2], Sheena C. Degnan[1], Renske Erion[1], Dorothee Kern[2] & Tom Alber[1]

A long-standing challenge is to understand at the atomic level how protein dynamics contribute to enzyme catalysis. X-ray crystallography can provide snapshots of conformational substates sampled during enzymatic reactions[1], while NMR relaxation methods reveal the rates of interconversion between substates and the corresponding relative populations[1,2]. However, these current methods cannot simultaneously reveal the detailed atomic structures of the rare states and rationalize the finding that intrinsic motions in the free enzyme occur on a timescale similar to the catalytic turnover rate. Here we introduce dual strategies of ambient-temperature X-ray crystallographic data collection and automated electron-density sampling to structurally unravel interconverting substates of the human proline isomerase, cyclophilin A (CYPA, also known as PPIA). A conservative mutation outside the active site was designed to stabilize features of the previously hidden minor conformation. This mutation not only inverts the equilibrium between the substates, but also causes large, parallel reductions in the conformational interconversion rates and the catalytic rate. These studies introduce crystallographic approaches to define functional minor protein conformations and, in combination with NMR analysis of the enzyme dynamics in solution, show how collective motions directly contribute to the catalytic power of an enzyme.

It has become widely accepted that not only the chemical steps of an enzymatic reaction, but also protein conformational rearrangements, contribute to the rate acceleration of enzymes[1,3–6]. For human CYPA, a peptidyl-prolyl *cis*/*trans* isomerase, NMR relaxation experiments revealed that millisecond motions during catalysis occur in a 'dynamic network' that could be described as a two-state interconversion between enzyme conformations bound to the *cis*- and *trans*-Pro substrates. Further, the free enzyme samples the same two conformations on a similar timescale, but with the equilibrium shifted far towards one substate[4]. Although these 'major' and 'minor' conformations are sampled during turnover, studies aimed at modelling the catalytic mechanism of CYPA[7–10] have focused on the substrate peptide and have not explained enzyme conformational changes. Moreover, superposition of 48 CYPA crystal structures, including complexes with peptides, inhibitors, and physiological substrates such as the HIV capsid protein[11,12], shows no structural heterogeneity that could rationalize the NMR results[4] (Supplementary Fig. 1).

In search of the catalytically essential, minor conformer of free CYPA detected in solution, we first extended the resolution of the crystal structure to 1.2 Å under cryogenic conditions. This high-resolution structure, however, yielded no evidence for an alternative conformation of the dynamic network. We then used a novel algorithm, Ringer[13], to systematically sample the electron density around each dihedral angle to discover additional unmodelled side-chain conformers. In a test set of 402 structures determined at 1.5 Å resolution or higher, >15% of the residues with unbranched side chains have unmodelled secondary electron-density peaks in the range of 0.3–1σ, below the normal noise threshold of 1σ. Strikingly, these peaks are significantly enriched at low-energy rotameric positions, supporting the interpretation that the peaks reflect true minor populations. While Ringer identified discrete side-chain heterogeneity for active-site residues Met 61 and Arg 55 in the 1.2-Å CYPA electron-density map (Supplementary Fig. 2), these alternative conformations were insufficient to explain the extension of the dynamic network away from the active site, particularly to Leu 98 and Ser 99, which showed the largest differences in NMR chemical shifts between the major and minor conformers[4,14].

To address this discrepancy between the X-ray and NMR analyses, we collected 1.39-Å diffraction data at ambient temperature to explore the possibility that the low-temperature data collection might alter the conformational distribution in the crystal[15,16]. Ringer plots indeed uncovered 0.3–1σ features for alternative rotamers of several residues including Leu 98 and Ser 99, in addition to the side-chain heterogeneity observed at cryogenic temperature (Fig. 1a). These results emphasize that crystal freezing can alter conformational distributions.

Modelling the minor conformation of Ser 99 produced a clash of the side-chain hydroxyl group with the Phe 113 ring. Inspection of the electron density surrounding Phe 113 revealed an alternative conformation below the 1σ threshold. $F_o - F_c$ difference density maps calculated without bias from model rebuilding confirmed the alternative conformers of Leu 98, Ser 99 and Phe 113 (Fig. 1b, c, Supplementary Fig. 3). Ringer did not identify the Phe 113 minor rotamer owing to a concomitant shift in the backbone that places Cγ within the envelope of the major conformer. This finding emphasizes that no algorithm can yet identify with complete fidelity all structural features without further scrutiny of the electron-density distribution. Refinement verified a network of alternative side-chain rotamers covering the entire active site and extending into the buried regions of the dynamic network previously detected by NMR (Fig. 1d).

The room-temperature diffraction data afford a possible structural rationale for the large NMR chemical-shift differences between the substates in free CYPA. The alternative side-chain rotamers, particularly of Phe 113, are predicted to alter the magnetic environment of the surrounding residues. Only a coupled switch of rotamers is consistent with the steric repulsion of major and minor rotamers, such as the Ser 99/Phe 113 clash in the dynamic network (Fig. 1b).

To critically test the idea that these two conformers interconvert during turnover, we designed a mutation distant from the active site to stabilize the minor CYPA substate. Ser 99, a buried residue in the dynamic network located >14 Å from the catalytic Arg 55, was replaced by Thr to fill the space occupied by both Ser 99 rotamers. This conservative change was designed to stabilize Phe 113 in the 'out' position by emulating the steric clash between the minor Ser 99 rotamer and the Phe 113 'in' position. Crystal structures of

---

[1]Department of Molecular and Cell Biology/QB3, University of California, Berkeley, California 94720-3220, USA. [2]Department of Biochemistry, Howard Hughes Medical Institute, Brandeis University, Waltham, Massachusetts 02454, USA.

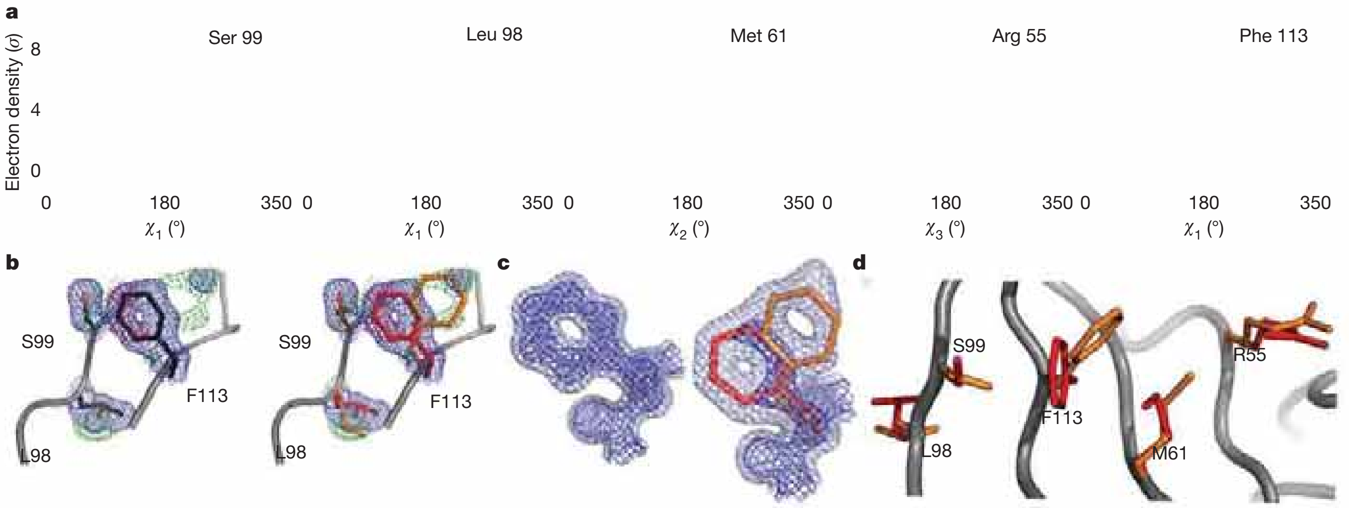


Figure 1 | Room-temperature X-ray crystallography and Ringer analysis detect conformational substates in CYPA. **a**, Local maxima above the 0.3σ threshold (yellow line) in Ringer plots[13] of electron density by dihedral angle (χ_i) reveal alternative side-chain conformations in room-temperature (red line) but not cryogenic (blue line) electron density for Ser 99, Leu 98, Met 61 and Arg 55. The peak for the minor conformation of Phe 113 occurred within the envelope of the major conformer due to a shift in the protein backbone. **b**, Electron-density maps calculated using room-temperature X-ray data define the alternative conformers of Leu 98, Ser 99 and Phe 113. Shown are $2F_o - F_c$ electron density (blue mesh; 1σ), and positive (green) and negative (red) $F_o - F_c$ difference density (3σ). **c**, $2F_o - F_c$ composite simulated-annealing omit electron density maps (1.0σ, dark blue; 0.3σ, light blue) show a unique conformation for Phe 113 in the 1.2-Å-resolution cryogenic structure (blue) and distinct major (red) and minor (orange) conformers in the 1.39-Å-resolution room-temperature structure. Electron density around the main chain and the surrounding residues was omitted for clarity. **d**, Steric collisions across the network of major (red) and minor (orange) conformers of Arg 55, Met 61, Phe 113 and Ser 99 explain how side-chain motions link the active site to remote buried residues.

the Ser99Thr mutant, solved at 1.6-Å and 2.3-Å resolution, indeed showed Thr 99 mimicking the alternative Ser 99 conformations, and Phe 113 was detected only in the exposed 'out' rotamer (Fig. 2a, b, and Supplementary Fig. 4). This change in rotamer populations was buttressed by three-bond *J*-coupling solution NMR experiments showing that the dominant Phe 113 χ_1 angle changed from +60° in wild-type CYPA to −60° in the Ser99Thr mutant (Supplementary Table 2). In the Ser99Thr variant structures, Thr 99, Phe 113 and Met 61 occupy the minor rotamers. The positions of Leu 98 and Arg 55 are consistent with either of the rotamers seen in wild-type CYPA. This pattern corroborates the conclusions that Ser 99, Phe 113, Met 61 and possibly Arg 55 are conformationally coupled (Fig. 2c) and that the Ser99Thr mutation severely reduces the population of the major conformation seen in wild-type CYPA.

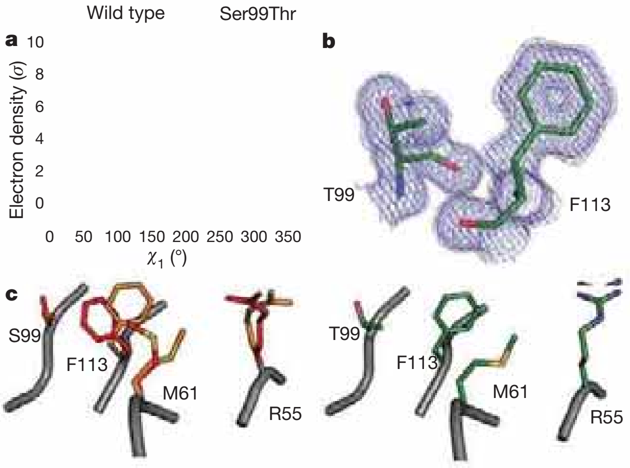


Figure 2 | The structure of the Ser99Thr mutant resembles the minor conformer of wild-type CYPA. **a**, χ_1 Ringer plot (0.3σ threshold is shown as yellow line) of the Ser99Thr mutant (dashed green) and room-temperature, wild-type Ser 99 CYPA structure (red) show that Thr99 occupies both positions populated by the Ser99-OHγ group. The angular offset between the major peaks reflects a backbone shift. **b**, The $2F_o - F_c$ simulated-annealing omit electron density map of the Ser99Thr CYPA mutant (1.0σ, dark blue; 0.3σ, light blue) shows apparently unique conformations for Thr 99 and Phe 113. The structure confirmed the prediction that rotation of Phe 113 to the 'out' position is coupled to rotation of the Ser 99 hydroxyl to the minor rotamer. **c**, Phe 113 and Met 61 in Ser99Thr CYPA (green, right) are detected exclusively in the position of the minor state of the wild-type enzyme (orange, left).

Connecting this interpretation of the crystal structures to the solution behaviour of the enzyme, NMR detected amide chemical-shift differences between wild-type and Ser99Thr CYPA in most active-site and core residues of the dynamic network (Fig. 3a,b). Although it is tempting to speculate that the Ser99Thr mutation 'traps' the minor state of wild-type CYPA based on our crystallographic data (Fig. 2), NMR relaxation-dispersion analysis[17] of the mutant enzyme revealed conformational exchange in both regions that showed collective motions in the wild-type enzyme (Fig. 3c). These regions include residues coupled to the active site (group I) and residues in the 65–85 loops (group II)[4]. In addition, for most of the group I residues, the peaks in Ser99Thr CYPA shift relative to wild type in the same direction as peaks in the previously characterized Lys82Ala variant, which displays a small increase in the population of the minor state[4]. The Ser99Thr mutation, however, causes much larger shifts (Fig. 3b, Supplementary Figs 5, 6), indicating that the populations are inverted. Together with our crystallographic and NMR *J*-coupling data, these results show that the Ser99Thr mutation shifts the structural equilibrium strongly towards a conformation that recapitulates key features of the previously undefined minor state of wild-type CYPA.

In addition to this dramatic population shift, the Ser99Thr mutation also specifically slows the motions of only group I residues by at least two orders of magnitude, pushing them into the slow NMR time regime. The slow dynamics of group I are easily identified by the characteristic increase in dispersion amplitude with increasing temperature, while the fast dynamics in the group II loop region, unaffected by the mutation, are characterized by the opposite trend (Fig. 3d, Supplementary Figs 7,8). This result is reinforced by the static magnetic-field (B_0) dependence of relaxation due to chemical exchange (R_{EX}) as determined by the α value[18]: at 25 °C the R_{EX} of group I residues is independent of B_0 ($<\alpha> = 0.16$, Supplementary

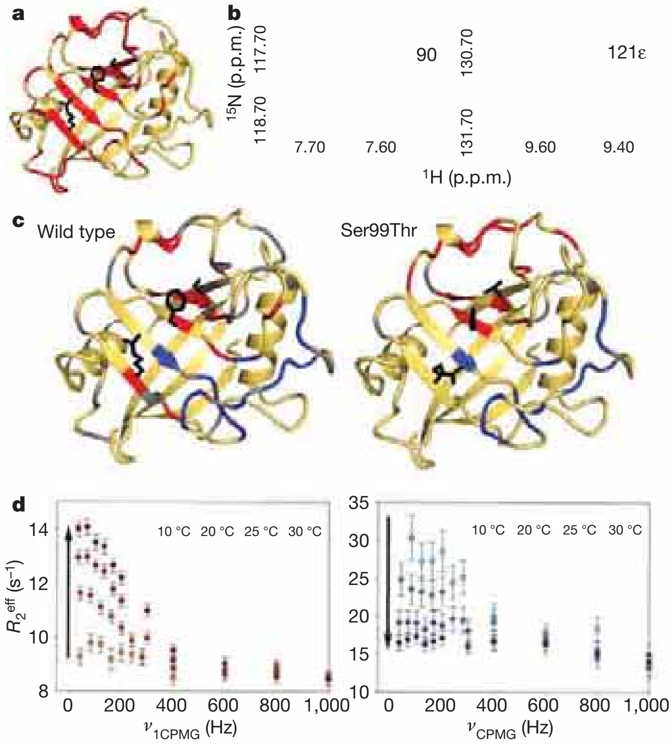


Figure 3 | The Ser99Thr mutation shifts the equilibrium towards the minor wild-type conformation and slows motions in the dynamic network in free CYPA. **a**, Significant ^{1}H–^{15}N chemical-shift differences between Ser99Thr and wild-type CYPA (red) propagate through group I residues (Arg 55, Phe 113 and Ser 99 shown as black sticks). **b**, Linear amide chemical shift changes (arrows) between wild-type (black), Lys82Ala (red) and Ser99Thr (blue) CYPA reflect the inversion of the major/minor equilibrium due to the Ser99Thr mutation. **c**, Residues undergoing slow (red) or fast (blue) motions on the NMR timescale in Ser99Thr (right) coincide with previously identified group I (red) and group II (blue) residues in wild-type (left) CYPA (amides in grey are prolines or overlapped peaks). **d**, Temperature dependence of the apparent ^{15}N transverse relaxation rate (R_2^{eff}) at increasing refocusing field strength (ν_{CPMG}) for group I (left) and group II (right) in Ser99Thr CYPA reveal that the mutation impedes group I conformational dynamics ($R_{\mathrm{EX}} \approx k_1$ and R_{EX} increases with temperature). In contrast, group II residues are unaffected by the mutation and display the opposite temperature dependence characteristic of fast motions on the NMR timescale. Dispersion curves were normalized to the intrinsic transverse relaxation rate (R_2^0) at 30 °C. Error bars indicate ± the root-mean-square deviation of R_2^{eff} calculated from the variance of peak intensities (see Methods).

Table 3), characteristic of slow exchange, while for group II residues the field-dependence is quadratic ($\langle\alpha\rangle = 2.0$), characteristic of fast exchange. Quantitative analysis of the group I residues yields a rate constant of conversion from the major to minor state (k_1) for Ser99Thr CYPA of $1.0 \pm 0.3\ \mathrm{s}^{-1}$ at 10 °C, in contrast to $\sim 60\ \mathrm{s}^{-1}$ for wild-type CYPA[4]. In the slow exchange regime of Ser99Thr, k_1 is well-determined and the major peak represents the true chemical shift of the major state (Fig. 3b).

Given the slower conformational transitions in Ser99Thr CYPA, what is the effect of this mutation on catalytic turnover? In a protease-coupled enzymatic assay[19], the Ser99Thr mutant showed a 300-fold reduction in the catalytic efficiency of the *cis*-to-*trans* isomerization of the peptide succinyl-AlaAlaProPhe-p-nitroanilide (AAPF) (Fig. 4a). This large reduction in catalytic efficiency resembles the effects of the Arg55Lys mutation, which removes the active-site residue thought to promote the chemical step of the reaction but does not perturb the enzyme dynamics[4] or global structure (Supplementary Fig. 9). To separate the energetic contributions to the binding and isomerization steps, we measured the dissociation constants of the peptide substrate for the CYPA variants using NMR titration experiments. Peptide affinity (K_{D}) was weakened

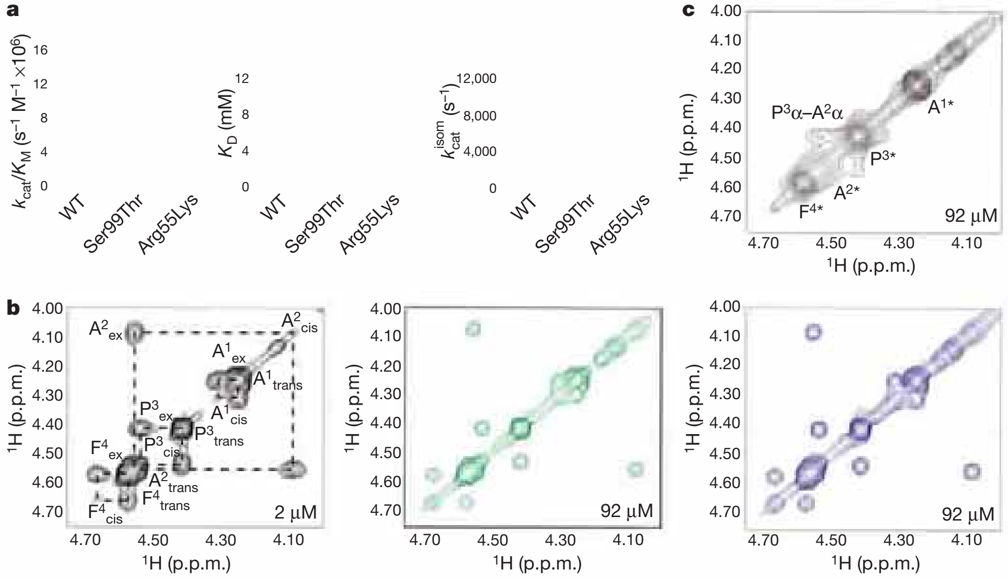


Figure 4 | Impeded motions in the dynamic network severely reduce the catalytic power of a chemically competent enzyme. **a**, Mutations affecting the enzyme dynamics (Ser99Thr) or the chemical step (Arg55Lys) each drastically reduce $k_{\mathrm{cat}}/K_{\mathrm{M}}$ (ref. 19) by reducing the bidirectional isomerization step on the enzyme ($k_{\mathrm{cat}}^{\mathrm{isom}}$) and not substrate affinity (K_{D}) of CYPA. **b**, ^{1}H–^{1}H NOE-exchange spectra at 0.2 s mixing time showing isomerization of the peptide AAPF (1 mM) by catalytic amounts of wild-type (black), Ser99Thr (green) and Arg55Lys (blue) CYPA (concentrations indicated in the spectra). Assignments and dashed lines connecting exchange (off-diagonal) and auto peaks are included for wild type. Much higher concentrations of the Ser99Thr and Arg55Lys variants are needed relative to wild-type CYPA to obtain similar exchange peaks, reflecting severely reduced catalytic activity. **c**, The *cis*- and *trans*- peaks coalesce (asterisk) for wild-type CYPA at the same enzyme concentration as the mutant forms due to its much greater activity. The only remaining off-diagonal peak is a $P^3\alpha$-$A^2\alpha$ NOE characteristic of a *cis*-prolyl peptide bond.

only 3–6 fold by the Ser99Thr (6.7 ± 0.8 mM) and Arg55Lys (11.3 ± 2.5 mM) mutations (Fig. 4a, Supplementary Fig. 10), suggesting that the major effect of these mutations is on the catalytic turnover number (k_{cat}) and not on the dissociation constant of all enzyme-substrate forms (K_M).

To independently quantify the mutational effects on the rate constant of the isomerization step (k_{cat}^{isom}), we measured catalysis of the *cis*/*trans* isomerization directly by ZZ exchange spectroscopy[20] (Fig. 4b). These experiments detect the overall rate of converting the entire substrate pool from the *cis*- to the *trans*-Pro conformation and vice versa. The Ser99Thr and Arg55Lys mutations reduce overall turnover by similar amounts compared to wild-type CYPA (Fig. 4, Supplementary Fig. 11). Quantitative fitting of these exchange data[20] indicates that the Ser99Thr mutation severely reduces the rate of the bidirectional isomerization step on the enzyme ($190 \pm 20\,s^{-1}$ for Ser99Thr compared to $13{,}000 \pm 800\,s^{-1}$ for wild-type CYPA). Strikingly, this reduction (~70-fold) parallels the upper bound for the reduction in the rate constant for the major-to-minor conformational change for group I residues (~60-fold) in the free enzyme (Figs 3c, d).

The similar rate reductions in the free enzyme and of substrate turnover strongly suggest that the slowed fluctuations due to the Ser99Thr mutation underlie the reduction of the catalytic rate. The effects of this 'molecular traffic-jam' mutation remote from the active site support the idea that the coupled motions of the dynamic network tied to the rotation of Phe 113 play a key role in turnover. The conformational transition of Phe 113 observed here suggests that the side-chain rotation directly and distinctly impacts *cis*- and *trans*-forms of the proline-substrate during catalysis. Indeed, smaller[21] or larger[4] replacements for Phe 113 substantially decrease CYPA activity, suggesting the wild-type enzyme occupies a narrow structural optimum for catalysis. Both the interaction of the Arg 55 side chain with the substrate[8,9], which facilitates the chemistry, and side-chain motions throughout the dynamic network, which enable facile interconversion of conformational states, are necessary for catalysis (Fig. 4). Therefore, neither the dynamics nor active-site chemistry that lowers the transition-state energy are sufficient to promote efficient turnover.

Only by collecting X-ray diffraction data at ambient temperature did we observe agreement between the conformational substates detected using NMR and X-ray crystallography. This finding underscores previous studies demonstrating that the conformational distributions in crystals of myoglobin[22,23] and RNase[16,24] are restricted at cryogenic temperatures. Unlike these classic studies, which accessed the manifold of crystal structures indirectly through reductions in protein volume, inhibition of ligand binding and reductions in crystallographic *B*-values at cryogenic temperatures, here we use Ringer[13] to directly sample the electron density. This automated, systematic approach reveals alternative conformational substates. Our results emphasize that the common practices of crystal freezing and building models at the 1σ threshold, rather than capturing the conformational distribution in the crystal, can eliminate information about functionally critical conformations.

Our complementary results from NMR and X-ray crystallography for CYPA demonstrate a general strategy to discover structural ensembles of interconverting substates, to identify which regions of enzymes couple to active sites, and to evaluate the roles of hidden, higher energy conformations in catalysis by other enzymes[3,25]. Particularly by identifying structurally remote residues that couple to the active site, defining the multiple conformations that contribute to function can afford a deeper understanding of the effects of sequence variation in disease and evolution[26]. Such knowledge may enable progress not only in understanding and manipulating the mechanisms of numerous macromolecular systems, but also in defining the manifold of conformations accessible to inhibitors and therapeutics[27].

METHODS SUMMARY

CYPA was purified as described[4]. Crystals were grown at pH 7.5 using PEG3350 or DL-malate as precipitants. All X-ray data sets derive from single crystals. For room-temperature data collection, we used the Rigaku free mounting device at ALS Beamline 12.3.1[28]. Short exposures and crystal translation every 10 frames limited radiation damage. Ringer analysis was performed by real-space sampling of electron density around side-chain dihedral angles[13]. Inspection of plots of electron density versus dihedral angle revealed Ringer peaks above 0.3σ that guided placement of alternative conformers. This modelling process was performed iteratively using both conventional and simulated-annealing composite omit maps.

Constant-time ^{15}N backbone CPMG TROSY relaxation-dispersion data[17] were collected in an interleaved manner on Varian INOVA 500 and 600 spectrometers and a Bruker AVANCE-800 equipped with a cryoprobe. Dispersion curves were fitted to the full Carver-Richards equation[29], confirming the results for wild-type CYPA where the exchange is in the intermediate to fast time regime[4]. In contrast, for Ser99Thr CYPA, global fits of dispersion of group-I residues at all temperatures, assuming temperature-independent $\Delta\omega$, unambiguously revealed exchange in the slow regime ($R_{EX} \approx k_1$, rate constant for efflux from the major state). Numerical fitting and estimation of the dispersion amplitude gave the same rate constant k_1 at 10 °C within experimental error.

Three-bond, aromatic *J* couplings were obtained at 25 °C as described[30]. Dissociation constants for AAPF were obtained by titrating the peptide into a solution of 0.2 mM CYPA at 6 °C and fitting the resulting chemical shifts to single-site exchange. The enzyme-catalysed *cis/trans* isomerization rates at 6 °C were measured using ^1H–^1H NOESY spectra of AAPF and fitting the data to equations for ZZ exchange[20]. The bidirectional k_{cat}^{isom} was calculated from the fitted exchange rate using the K_D measured in titrations (see Online Methods).

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

Received 13 August; accepted 26 October 2009.

1. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450,** 964–972 (2007).
2. Mittermaier, A. & Kay, L. E. Review—new tools provide new insights in NMR studies of protein dynamics. *Science* **312,** 224–228 (2006).
3. Boehr, D. D., McElheny, D., Dyson, H. J. & Wright, P. E. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science* **313,** 1638–1642 (2006).
4. Eisenmesser, E. Z. *et al.* Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438,** 117–121 (2005).
5. Hammes-Schiffer, S. & Benkovic, S. J. Relating protein motion to catalysis. *Annu. Rev. Biochem.* **75,** 519–541 (2006).
6. Schramm, V. L. & Shi, W. Atomic motion in enzymatic reaction coordinates. *Curr. Opin. Struct. Biol.* **11,** 657–665 (2001).
7. Agarwal, P. K. *Cis/trans* isomerization in HIV-1 capsid protein catalyzed by cyclophilin A: insights from computational and theoretical studies. *Proteins* **56,** 449–463 (2004).
8. Hamelberg, D. & McCammon, A. Mechanistic insight into the role of transition-state stabilization in cyclophilin A. *J. Am. Chem. Soc.* **131,** 147–152 (2009).
9. Li, G. H. & Cui, Q. What is so special about Arg 55 in the catalysis of cyclophilin A? Insights from hybrid QM/MM simulations. *J. Am. Chem. Soc.* **125,** 15028–15038 (2003).
10. Trzesniak, D. & Van Gunsteren, W. F. Catalytic mechanism of cyclophilin as observed in molecular dynamics simulations: pathway prediction and reconciliation of X-ray crystallographic and NMR solution data. *Protein Sci.* **15,** 2544–2551 (2006).
11. Howard, B. R., Vajdos, F. F., Li, S., Sundquist, W. I. & Hill, C. P. Structural insights into the catalytic mechanism of cyclophilin A. *Nature Struct. Biol.* **10,** 475–481 (2003).
12. Ke, H. M. & Huai, Q. Crystal structures of cyclophilin and its partners. *Front. Biosci.* **9,** 2285–2296 (2004).
13. Lang, P. T. *et al.* Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Sci.* (submitted).
14. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science* **295,** 1520–1523 (2002).
15. Halle, B. Biomolecular cryocrystallography: structural changes during flash-cooling. *Proc. Natl Acad. Sci. USA* **101,** 4793–4798 (2004).
16. Rasmussen, B. F., Stock, A. M., Ringe, D. & Petsko, G. A. Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature* **357,** 423–424 (1992).
17. Loria, J. P., Rance, M. & Palmer, A. G. A TROSY CPMG sequence for characterizing chemical exchange in large proteins. *J. Biomol. NMR* **15,** 151–155 (1999).
18. Millet, O., Loria, J. P., Kroenke, C. D., Pons, M. & Palmer, A. G. The static magnetic field dependence of chemical exchange linebroadening defines the NMR chemical shift time scale. *J. Am. Chem. Soc.* **122,** 2867–2877 (2000).

19. Kofron, J. L., Kuzmic, P., Kishore, V., Colonbonilla, E. & Rich, D. H. Determination of kinetic constants for peptidyl prolyl *cis-trans* isomerases by an improved spectrophotometric assay. *Biochemistry* **30,** 6127–6134 (1991).
20. Farrow, N. A., Zhang, O. W., Forman-Kay, J. D. & Kay, L. E. A Heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. *J. Biomol. NMR* **4,** 727–734 (1994).
21. Zydowsky, L. D. *et al.* Active site mutants of human cyclophilin A separate peptidyl-prolyl isomerase activity from cyclosporine A binding and calcineurin inhibition. *Protein Sci.* **1,** 1092–1099 (1992).
22. Frauenfelder, H. *et al.* Thermal expansion of a protein. *Biochemistry* **26,** 254–261 (1987).
23. Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* **280,** 558–563 (1979).
24. Tilton, R. F., Dewan, J. C. & Petsko, G. A. Effects of temperature on protein structure and dynamics: X-ray crystallographic studies of the protein ribonuclease-A at nine different temperatures from 98 to 320K. *Biochemistry* **31,** 2469–2481 (1992).
25. Beach, H., Cole, R., Gill, M. L. & Loria, J. P. Conservation of μs-ms enzyme motions in the apo- and substrate-mimicked state. *J. Am. Chem. Soc.* **127,** 9167–9176 (2005).
26. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324,** 203–207 (2009).
27. Lee, G. M. & Craik, C. S. Trapping moving targets with small molecules. *Science* **324,** 213–215 (2009).
28. Kiefersauer, R. *et al.* A novel free-mounting system for protein crystals: transformation and improvement of diffraction power by accurately controlled humidity changes. *J. Appl. Crystallogr.* **33,** 1223–1230 (2000).
29. Davis, D. G., Perlman, M. E. & London, R. E. Direct measurements of the dissociation-rate constant for inhibitor-enzyme complexes via the $T_{1\rho}$ and T_2(CPMG) methods. *J. Magn. Reson. Ser. B* **104,** 266–275 (1994).
30. Hu, J. S., Grzesiek, S. & Bax, A. Two-dimensional NMR methods for determining χ_1 angles of aromatic residues in proteins from three-bond $J_{C'C\gamma}$ and $J_{NC\gamma}$ couplings. *J. Am. Chem. Soc.* **119,** 1803–1804 (1997).

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

Acknowledgements We thank S. Marqusee and B. Krantz for discussions; S. Classen, G. Meigs, J. Holton, A. Samelson, N. Echols, P. Afonine, and the Phenix team for technical support; J. Tainer for access to Rigaku free-mounting device at ALS Beamline 12.3.1; J. Pelton and D. Wemmer for providing essential help and access to NMR facilities. J.S.F. was supported by US NSF and Canadian NSERC fellowships. This work was funded by the US National Institutes of Health (to T.A.) and the US National Institutes of Health, the US Department of Energy Office of Basic Energy Sciences, and the Howard Hughes Medical Institute (to D.K.).

Author Contributions J.S.F., S.C.D. and R.E. performed the X-ray experiments, M.W.C. performed the NMR experiments, and M.W.C. and J.S.F. performed the activity and binding assays. J.S.F., M.W.C., D.K. and T.A. analysed data and wrote the paper. All authors contributed to data interpretation and commented on the manuscript.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited in the PDB under accession codes 3K0M, 3K0N, 3K0O, 3K0P, 3K0Q and 3K0R. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.A. (tom@ucxray.berkeley.edu) or D.K. (dkern@brandeis.edu).

METHODS

Crystallography. The standard purification[4] was augmented with a Superdex S75 column (GE Healthcare) equilibrated in 20 mM HEPES pH 7.5, 100 mM NaCl and 0.5 mM TCEP. Crystals (wild-type and Ser99Thr) with $P2_12_12_1$ symmetry were grown by hanging-drop vapour diffusion by mixing an equal volume of well solution (100 mM HEPES pH 7.5, 22% PEG 3350, 5 mM TCEP) and protein (wild-type CYPA at 60 mg ml^{-1} and mutants at 40 mg ml^{-1}). Trigonal ($P3_221$) crystals of Ser99Thr and Arg55Lys CYPA were grown similarly from 1.8 M DL-malic acid. For cryogenic data sets (collected at 100 K), crystals were flash frozen in liquid nitrogen with well solution plus 15% xylitol as a cryoprotectant for the orthorhombic crystals and 10% glycerol for the trigonal crystals. Wild-type CYPA cryogenic X-ray data were collected at Beamline 9-1 at the Stanford Synchrotron Radiation Laboratory. Ser99Thr and Arg55Lys cryogenic data sets were collected at Beamline 8.3.1 at the Lawrence Berkeley National Laboratory Advanced Light Source (ALS).

Room-temperature X-ray data were collected at 15 °C with 96% humidity using a temperature- and humidity-controlled goniometer head[28] (the Rigaku free-mounting device) at ALS Beamline 12.3.1. To limit the effects of radiation damage while maintaining high signal-to-noise, we collected short exposures (0.1 s) for 180° with 1° oscillations. A large crystal (1 mm × 0.5 mm × 0.3 mm) was translated halfway through this rapid collection protocol. Following this initial low-exposure pass, an additional complete data set of 90° with longer exposures (1 s) was collected while translating the crystal every 10 frames. During data processing, we ensured that radiation damage had not degraded the data quality by subdividing the data into subsets of 45 frames and confirming that unit cell parameters, scale factors and χ^2 statistics were consistent throughout the data set[31]. Owing to the short total exposure time (108 s), the limited exposure of each spot on the crystal, the size of the crystal and the finite rate of crystal damage, the data were not significantly influenced by radiation damage.

Reflections were processed using HKL2000[32]. Structures were solved using PHASER[33] through phenix.automr[34] with 2CPL chain A as the search model. To corroborate the features of residues 98, 99 and 113 observed in the initial electron density, reflections were processed using MOSFLM, through ELVES[35], and XDS, through XIA2[36], and the structures were additionally solved using 1RMH chain A as the search model. The features described in the text were consistent across electron-density maps calculated from data processed with all three programs and both molecular-replacement search models.

Coordinates, atomic displacement parameters, and occupancy (where appropriate) were refined using phenix.refine[34] with manual rebuilding using COOT[37]. Validation was performed using MOLPROBITY[38] and PROCHECK[39]. Data collection, processing, refinement, and validation statistics are listed in Supplementary Table 1. Structural figures were prepared using PyMol[40].

Ringer. Ringer[13] analyses electron density in real space around the dihedral angles of the modelled side chains to discover unmodelled alternative conformers. For each residue, each side-chain dihedral angle is rotated in 10° increments, and the electron density value calculated using a cubic spline interpolation is tabulated at the position of the rotated atom. In sp^3 geometry, for example, the sampled position is 1.53 Å away and oriented 111.1° from the preceding side-chain carbon atom. In regions of electron density that are not occupied by modelled atoms, Ringer peaks $>0.3\sigma$ are significantly enriched at rotameric positions, and χ_1 peaks are strongly correlated with rotameric χ_2 peaks in long side chains, providing evidence that these features correspond to alternative conformers[13]. The contour level of 0.3σ provides an initial working cut-off to detect alternative rotamers in high-resolution electron density maps. Alternative conformers are modelled after inspection of the electron density maps for continuous or correlated electron density for the side chain and subsequently refined.

Comparison of crystal structures. Forty-eight CYPA structures with 100% sequence identity (Supplementary Fig. 1) were superimposed using Theseus[41] and analysed for backbone root-mean-square-deviation (r.m.s.d.) to establish regions of conformational heterogeneity or flexibility.

NMR methods. Unless otherwise specified, NMR samples consisted of 1 mM wild-type or mutant CYPA in 50 mM Na_2HPO_4 pH 6.5, 0.02% NaN_3, 1 mM DTT, 10% D_2O. Constant-time relaxation dispersion spectra[17,42] were collected in an interleaved manner, processed using NMRPipe[43] software, and analysed using NMRView[44] and custom scripts. Errors in peak intensities were estimated using the variance for non-exchanging peaks, signal to noise, and duplicate points. Errors for the fits were estimated using 200 Monte Carlo simulations. CPMG relaxation dispersion data collected at 25 °C at 500, 600 and 800 MHz not only constrained the fitted parameters but also provided α values for all residues[18].

We note that in the slow exchange regime of Ser99Thr, CPMG relaxation data and fitting them to the Carver–Richards equation is not robust to determine chemical shift differences or populations. The absence of the minor peak, in light of the estimated line-broadening, suggests an upper population limit for this species of about 10%. However, k_1 is well determined in the slow time regime by the value of R_{ex}.

Chemical shift differences between mutant and wild-type backbone amides or side-chain indoles (Trp 121) were considered significant when $\Delta\delta$ was greater than 0.05 p.p.m. in the ^{1}H dimension or 0.25 p.p.m. in the ^{15}N dimension. Three-bond J couplings were assessed using difference spectra[30].

NOESY experiments were performed at 6 °C in a buffer containing 860 μM AAPF and 100% D_2O using a standard gradient-filtered experiment[45]. Intensity curves were fitted to previously described equations[20] in Origin 6.0 using populations for the *cis* and *trans* isomer determined from the intensities of the peaks in a NOESY spectrum with a mixing time of zero. The fits of the data for all three enzyme forms yielded the corresponding exchange rates (k_{ex}), but the same R_1 values for the auto-peaks within experimental error. The bidirectional $k_{cat}{}^{isom}$ was calculated from this k_{ex} using the following equations:

$$k_{ex} = \frac{v^{isom}}{[S]}$$

$$v^{isom} = \frac{k_{cat}^{isom}[E][S]}{K_M + [S]}$$

where v^{isom} is the reaction velocity, K_M is the Michaelis-Menten constant, and [S] and [E] are substrate and enzyme concentrations, respectively. For fast dissociation relative to the rate of the isomerization, $K_M \approx K_D$, so:

$$k_{cat}^{isom} = \frac{k_{ex}(K_D + [S])}{[E]}$$

This yields a bidirectional $k_{cat}{}^{isom}$ when one uses the K_D averaged for the *cis* and *trans* isomers obtained from the titration experiments.

Coupled enzymatic assay. k_{cat}/K_M for the enzyme catalysed *cis*-to-*trans* isomerization of AAPF was measured at 10 °C using the standard protease coupled assay[19].

31. Southworth-Davies, R. J., Medina, M. A., Carmichael, I. & Garman, E. F. Observation of decreased radiation damage at higher dose rates in room temperature protein crystallography. *Structure* **15,** 1531–1541 (2007).
32. Otwinowski, Z. & Minor, W. in *Macromolecular Crystallography* Part A, 307–326 (Methods in Enzymology, Vol. 276, Academic, 1997).
33. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40,** 658–674 (2007).
34. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
35. Holton, J. & Alber, T. Automated protein crystal structure determination using ELVES. *Proc. Natl Acad. Sci. USA* **101,** 1537–1542 (2004).
36. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26,** 795–800 (1993).
37. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
38. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–W383 (2007).
39. Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. Procheck—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26,** 283–291 (1993).
40. Delano, W. L. The PyMOL Molecular Graphics System (DeLano Scientific, Palo Alto, 2008); 〈http://www.pymol.org〉.
41. Theobald, D. L. & Wuttke, D. S. THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics* **22,** 2171–2172 (2006).
42. Mulder, F. A. A., Mittermaier, A., Hon, B., Dahlquist, F. W. & Kay, L. E. Studying excited states of proteins by NMR spectroscopy. *Nature Struct. Biol.* **8,** 932–935 (2001).
43. Delaglio, F. *et al.* NMRPipe—a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
44. Johnson, B. A. & Blevins, R. A. NMR View—a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4,** 603–614 (1994).
45. Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. Investigation of exchange processes by 2-dimensional NMR-spectroscopy. *J. Chem. Phys.* **71,** 4546–4553 (1979).

ERRATUM

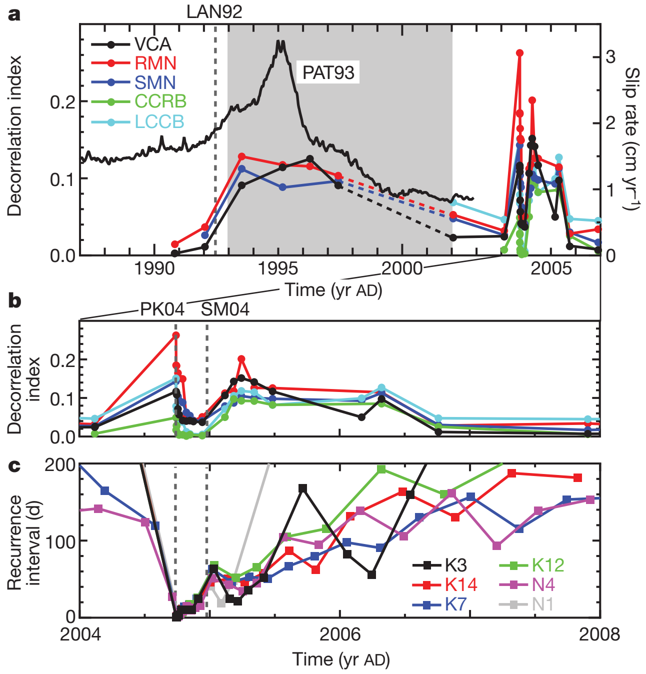
doi:10.1038/nature08545

Remote triggering of fault-strength changes on the San Andreas fault at Parkfield

Taka'aki Taira, Paul G. Silver, Fenglin Niu & Robert M. Nadeau

Nature 461, 636–639 (2009)

In Figure 2a of this Letter, in the key, the green line was inadvertently labelled 'SMN' instead of 'CCRB'. Also, in Figure 2c, the *y*-axis label should be 'Recurrence interval (d)' and not 'Recurrence index (d)'. The correct figure is shown below.



ERRATUM

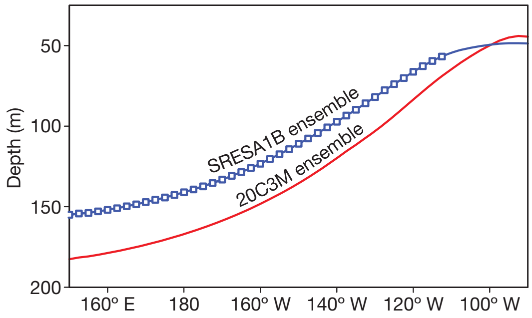
doi:10.1038/nature08546

El Niño in a changing climate

Sang-Wook Yeh, Jong-Seong Kug, Boris Dewitte, Min-Ho Kwon,
Ben P. Kirtman & Fei-Fei Jin

Nature **461**, 511–514 (2009)

In Figure 4 of this letter, the key for the 20C3M ensemble (red line) and the SRESA1B ensemble (blue line) were inadvertently mislabelled. The correct figure is shown below.



ERRATUM

doi:10.1038/nature08547

Integration of neuronal clones in the radial cortical columns by EphA and ephrin-A signalling

Masaaki Torii, Kazue Hashimoto-Torii, Pat Levitt & Pasko Rakic

Nature 461, 524–528 (2009)

On page 1 of this Letter, *Efna2* was inadvertently listed as *Efna1* when describing the triple knockout mouse (knockout for *Efna2*, *Efna3* and *Efna5*). Also, in the online-only Methods section of this Letter, the triple knockout mouse (*Efna2/a3/a5*) was inadvertently written as *Epha2/a3/a5*.

Ever-increasing resolution

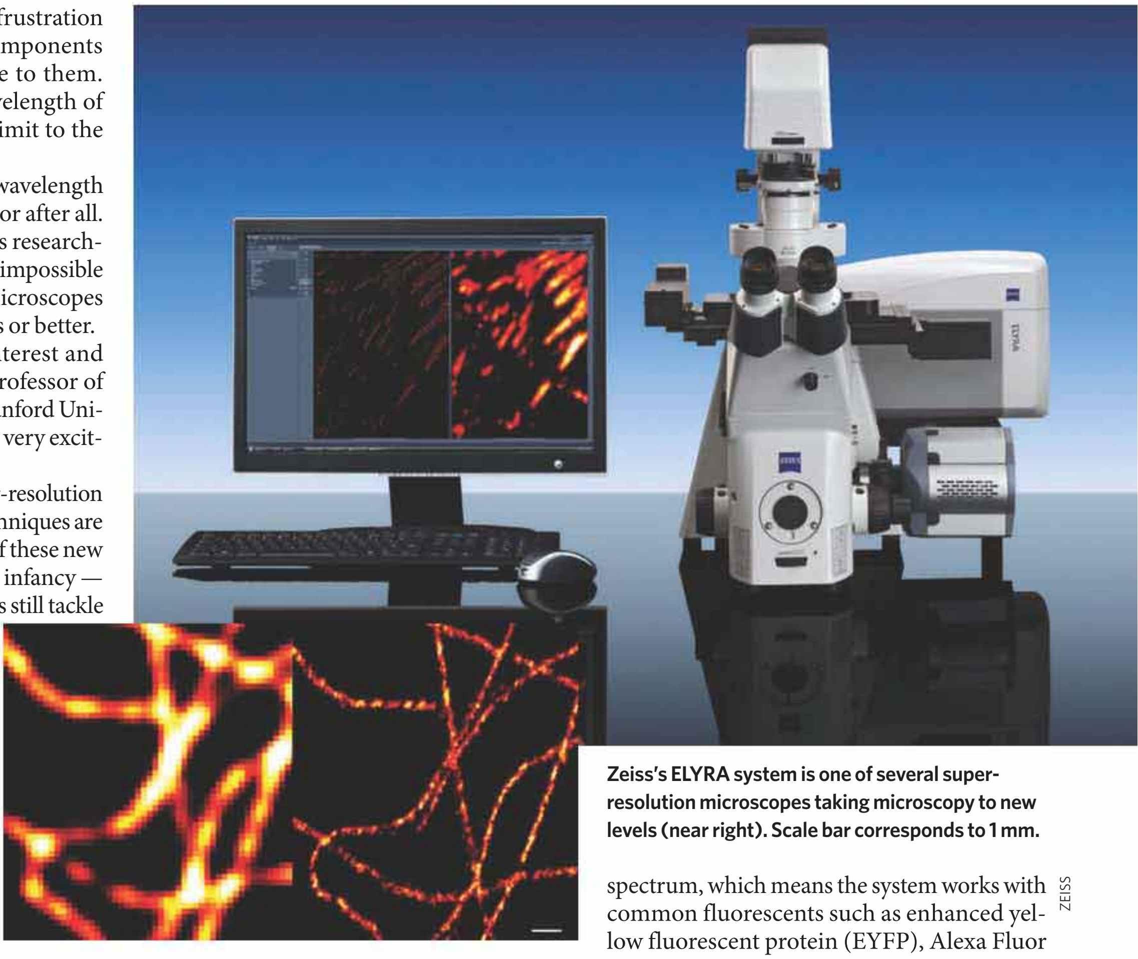
Overcoming the limitations of spatial and temporal resolution to image within a cell is no easy feat. Kelly Rae Chi examines the latest diffraction-busting technologies.

For many years it was a source of frustration for biologists that the internal components of a cell were practically invisible to them. Researchers believed that the wavelength of light determined a fundamental limit to the resolution of optical microscopes.

However, it now seems that the wavelength of light was not such a limiting factor after all. Super-resolution technology allows researchers to see details that are difficult or impossible to image with conventional light microscopes — at resolutions of 100 nanometres or better.

"There's a huge explosion of interest and progress," says W. E. Moerner, a professor of chemistry and applied physics at Stanford University in California. "That makes it very exciting to watch and to participate in."

Although the theory behind super-resolution is advancing, and many different techniques are now validated, commercialization of these new microscopes is to some extent in its infancy — so much so that many research teams still tackle super-resolution using their own lenses, lasers and algorithms.



Zeiss's ELYRA system is one of several super-resolution microscopes taking microscopy to new levels (near right). Scale bar corresponds to 1 mm.

N. SHIGUI, N. HIROKAWA/TOKYO UNIV.

ZEISS

Upping commercialization

The past year or so has been particularly active in terms of commercialization. Throughout 2008, Applied Precision of Issaquah, Washington, tested prototypes of its DeltaVision OMX system. The final version is available this year and is already installed in 11 laboratories. It uses three-dimensional structured-illumination microscopy, or 3D-SIM, which illuminates a sample with a series of light patterns that look like bar codes. The low-resolution light patterns reflect off the fine structure of the sample to create moiré fringes. By applying bar codes in different orientations and processing the reflections using computer algorithms, the microscope generates a high-resolution image of the underlying structure.

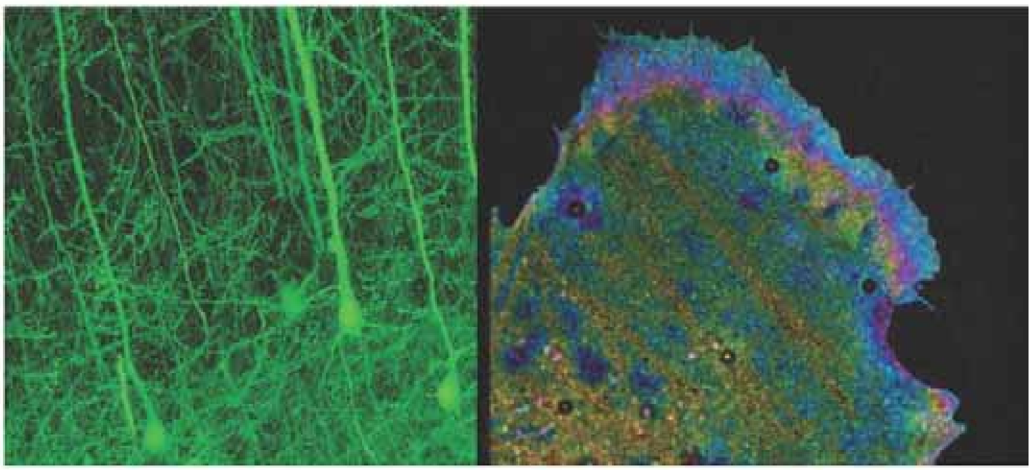
Carl Zeiss, headquartered in Oberkochen, Germany, is also using the SIM method in its system, ELYRA S.1, to be launched in January. And the company is launching ELYRA P.1, which is based on photoactivated localization microscopy (PALM). This uses photoactivatable proteins that are scattered throughout the sample. Taking multiple frames of the photoactivated dots, and combining them into a high-resolution image, the ELYRA P.1 produces a resolution of around 20 nm. Zeiss also offers a single system, the ELYRA PS.1, that combines laser scanning microscopy, PALM and SIM.

Leica Microsystems in Wetzlar, Germany, is using a variation of stimulated emission depletion (STED) microscopy in its newest super-resolution system. The technique uses a laser beam to excite fluorescent dyes, or fluorophores, within the sample. As with a normal laser scanning microscope, the size of the excited spot determines the resolution of the microscope. In STED microscopy, to improve resolution and narrow the focus of the beam, an excitatory laser pulse is immediately followed by a ring-shaped depletion laser pulse to leave molecules in a smaller region in the centre of the ring in an excited state. This allows resolution of structural details in the 100 nm range. In Leica's new system, TCS STED CW, the excitatory and depletion laser beams are continuous and overlayed rather than pulsed. Whereas the TCS STED microscope was priced at around US$1 million, the cost of the TCS STED CW microscope will be closer to the price range of high-end confocal microscopes.

The lasers in TCS STED CW microscopy use the intermediate range of the electromagnetic spectrum, which means the system works with common fluorescents such as enhanced yellow fluorescent protein (EYFP), Alexa Fluor 488 and fluorescein isothiocyanate (FITC). "As long as it's a good confocal sample, it's a good STED sample as well," says Tanjef Szellas, product manager for super-resolution at Leica in Mannheim in Germany.

"Building or buying a confocal without a STED option doesn't make much sense," says Stefan Hell, a physicist at the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany, who pioneered the super-resolution movement by creating STED microscopy and other forms of imaging that bypass the limits of diffraction — about 200 nm (see 'Breaking the light barrier', overleaf). "The commercial hurdles, in terms of costs and system complexity for basic STED microscopy, are gone."

Nikon, headquartered in Tokyo, is using a super-resolution approach called stochastic optical reconstruction microscopy (STORM) — a similar system to PALM — developed in the lab of Xiaowei Zhuang, a physicist at Harvard University in Cambridge, Massachusetts. STORM uses photoswitchable fluorescent dyes that are activated and deactivated by different coloured light. By generating overlapping images in different colours and compiling them, the system can produce three-dimensional images that have a spatial resolution ten times better than confocal images in each dimension, Zhuang says. Nikon's instrument, called N-STORM, should be available by May 2010.



New microscopy techniques are showing structures (pyramidal neurons, left and a U2OS bone cell, right) in unprecedented detail.

N. WEILER, K. MICHEVA, ST. SMITH/ STANFORD UNIV.

H. HESS

Technology tweaks

The pioneers of super-resolution microscopy are continuing to improve their methods with better sample preparation, a few strategically placed pieces of hardware and more sophisticated algorithms.

One way to achieve greater resolution in fixed samples is to cut thinner slices of tissue. Some groups, including Zhuang's, are collaborating to combine STORM with array tomography, which involves embedding fixed tissue in resin and cutting ultrathin (50–200 nm) slices. "Array tomography is a potential adjunct to all the super-resolution methods," says Stephen Smith of Stanford University in Palo Alto, California, who first described the method. Using this combination, his group has achieved a 14-fold improvement in axial resolution, and twofold improvements in the each of the two dimensions, compared with confocal imaging.

Other combination strategies are possible. For example, two-photon excitation microscopy, an established method that allows researchers to see further into tissues, can be combined with STED microscopy. Earlier this year, Hell's group used the technique on mammalian cell nuclei tagged with fluorescent nanoparticles, obtaining a resolution of less than 70 nm in the focal plane[1].

Bernardo Sabatini, a neurobiologist at Harvard Medical School in Boston, Massachusetts, used the same combination to image slender neuronal projections and dendritic spines in brain slices, which are not visible using conventional methods[2]. Sabatini and his team built the system themselves. They created the ring shape in the depletion laser using a phase plate to split the laser light so that one part served as the central spot and the other part as the ring. The result was a three-fold improvement in spatial resolution over normal two-photon techniques.

Advances in resolution have not been restricted to two-dimensional samples. Harald Hess and his colleagues at Howard Hughes Medical Institute's Janelia Farm Research Campus in Virginia have modified PALM to do 3D imaging by adding an interferometer and a second objective lens[3]. Called iPALM, this method can pinpoint fluorescent labels to within 10–20 nm in three dimensions. However, the samples must be around 225–250 nm thick, and ensuring that the extra optics needed are perfecty aligned is a time-consuming task. As a result, iPALM might prove challenging to commercialize, Hess says.

Groups led by Moerner and Rafael Piestun at the University of Colorado, Boulder, have

BREAKING THE LIGHT BARRIER

In 1873, German physicist Ernst Abbe proposed that diffraction fundamentally limited the resolution that any microscope could achieve to around half the wavelength of light. And, despite many advances, microscopes didn't threaten to challenge this law of physics for more than a century. Stefan Hell, now a director of the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany, was the first to show that the diffraction limit could be beaten.

Hell, while a postdoc at the University of Turku in Finland in the 1990s, thought that, with the right lasers, he could activate a fluorescent spot and then shrink it by superimposing a larger, hollow beam of light to deplete all the light emission except for that at the centre of the spot. He called the technique stimulated emission depletion (STED) microscopy. Although many physicists were initially sceptical of Hell's ideas, by 2000 he had used STED to produce the first nanoscale fluorescence images[10].

Super-resolution microscopy has blossomed since, allowing researchers to see cellular processes unfolding at nanometre scales. "This is something that the field has desired since people first started looking through light microscopes," says Jan Liphardt, a biophysicist at the Lawrence Berkeley National Laboratory in California.

Since Hell's work, the field has been boosted by other groups, including those of Eric Betzig, a physicist at the Howard Hughes Medical Institute's Janelia Farm Research Campus in Virginia and Jennifer Lippincott-Schwartz, a cell biologist at the National Institutes of Health in Bethesda, Maryland. In 2006, the groups reported that they had increased resolution by harnessing single-molecule photoactivatable fluorescent proteins and compiling images of thousands to millions of them[11]. They called the approach photo-activated localization microscopy (PALM).

At Harvard University in Cambridge, Massachusetts, physicist Xiaowei Zhuang has developed three-dimensional (3D) stochastic optical reconstruction microscopy (STORM), which uses photoswitchable probes to temporally separate the overlapping images of individual molecules and so boost resolution to ten times better than the diffraction limit.

Yet another approach — fluorescence PALM (FPALM) — was developed in 2006 by Samuel Hess, a physicist at the University of Maine in Orono. His group's technique involves looking at thousands of fluorophores at once, and localizing on small numbers at a time. These methods have already begun to demostrate their utility. For example, in 2007, Hess's group showed that FPALM could be used to detect proteins clustering in lipid rafts in living cells[12]. In 2008, Lippincott-Schwartz's group combined PALM with single-particle tracking to detect the movement of membrane proteins in live cells[13]. And Zhuang's group used 3D STORM to image microtubules and other molecular structures within monkey kidney cells[14], later extending the method to multicolour 3D imaging of whole cells[15]. Hell's group, early in 2008, used the STED method to show the movement of synaptic vesicles inside living neurons at video rate[16].

But the field is just warming up. "To people like me who were trained in physics or optics in the 1990s, it's just unbelievable that one can image below the resolution of light," says Bernardo Sabatini, a neurobiologist at Harvard Medical School in Boston, Massachusetts. "The major revolution for the next 5 or 10 years is getting these advances to answer biological questions."

K.R.C.

come up with another 3D technique, called double-helix PALM. This adds two lenses and a spatial light modulator to a standard wide-field microscope. The modulator splits the light into a double-helix shape that, when shined on to a fluorescent tagged molecule, generates two spots of light. The angular orientation of the spots can be used to calculate the depth of the molecule. “The double helix has a much deeper depth of field over which the objects remain in focus,” says Moerner. “In current experiments we use a two-micrometre range, but that can be changed depending on how the phase modulator is designed.”

In collaboration with Robert Twieg at Kent State University in Ohio, Moerner's team has localized molecules with 20 nm precision in depth and 12 nm precision in length and width[4], and has used the system to image bacterial proteins whose structures have never before been seen in cells.

Going live

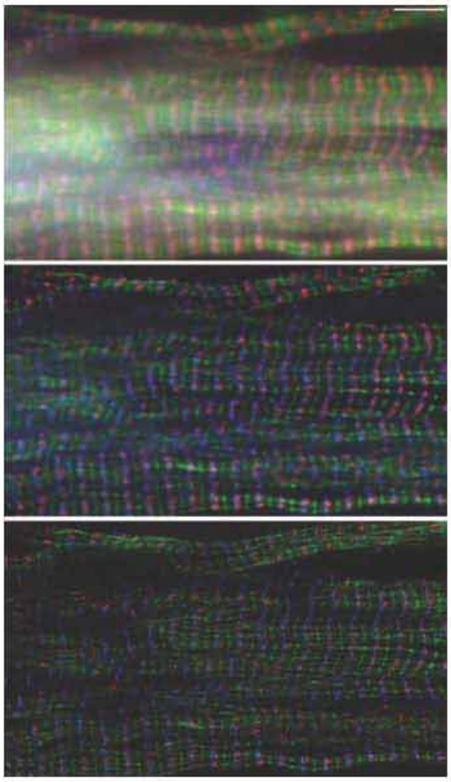
SIM is also finding application in 3D systems — through imaging live cells. Mats Gustafsson and his colleagues at Janelia Farm sped up the bar-code patterns used in two-dimensional SIM to allow it to image live cells, which are constantly moving. They did this with a liquid-crystal spatial light modulator, a mirror-like device, 1.5 centimetres long, that generates patterns of light using thousands of pixels that can be controlled individually. The modulator allows the microscope to generate new patterns of light about 1,000 times faster than the normal SIM equipment. The technique has allowed Gustafsson's team to see proteins moving along individual microtubules within living cells of the fruitfly *Drosophila melanogaster* at 100 nm resolution[5].

SIM will soon become even faster because of advances in camera technology. “You can already buy cameras that go up to 1,000 frames per second, and that speed can more or less be directly translated into speed increase in structured illumination,” says Rainer Heintzmann, a molecular biologist at King's College London, who has used the system to see the inner folds of a cell's mitochondria.

Bernardo Sabatini has modified microscopy software.

R. E. KLEIN, HHMI/AP

One of the main hurdles to widespread use of super-resolution microscopy in biology is the lack of appropriate probe molecules. Developing new probes is a trial-and-error process of seeing whether each protein variant will work in a variety of biological samples. “This is where the long, hard work is,” Hess says. “Right now there are few examples of single colours, but making them work in pairs will be the part that will bring a lot of value to the field.” Having two or more colours would help biologists by allowing them to track more than one protein in a cell. Two-colour experiments can work but, at present, the second colour has usually faded by the time the first colour has been completed, Hess says.



Increasing detail: skeletal muscle cells seen by wide-field fluorescence image (top); standard deconvolution image; and by Applied Precision's DeltaVision OMX system (bottom).

C. T. PAPPAS, D.A. ELLIOTT, C.C. GREGORIO

A group led by Vladislav Verkhusha, an associate professor of anatomy and structural biology at Albert Einstein College of Medicine at Yeshiva University in New York, has developed a new red probe[6] that works using PALM methods. Called photoactivatable monomeric cherry, or PAmCherry1, the protein becomes red when activated by violet light. It seems to have better pH stability, faster photoactivation and better photostability than other red photoactivatable probes, the group says.

Several groups are working on small organic probes that can be used with PALM or STORM. Earlier this year, for example, researchers at Bielefeld University in Germany imaged the inner membranes of mitochondria[7] using commercially available fluorophores. The researchers triggered the fluorophores to light up on cue by altering the chemical environment.

And Hell is backing an unlikely material as a potential probe: diamond. His team found that STED microscopy can pick up subtle structural defects in diamonds with 10–20 nm resolution[8]. So if the diamonds can be ground into 35-nanometre-sized pieces and attached to antibodies — which researchers have shown is possible — they could be used as tags that will never bleach under STED microscopy's lasers. There is a “clear prospect” of using diamonds for cell-biology applications, Hell says. “It is not a pipe dream. It's farther down the road but is coming.”

Applying software

An increasingly important aspect of microscopy — particularly super-resolution systems — is the software that constructs the final images. Scientists who want to modify the output or improve the accessibility of their super-resolution systems can develop algorithms and write programs for image processing. Many scientists have developed their own software and made it available to the community. “The design of the microscope and its modification is available to anybody who knows how to program,” says Sabatini, and once you know how the microscope works, it is possible to determine where the photons should be and, from that, how the image can be constructed.

Sabatini has helped at least ten other labs to implement super-resolution methods. “Biologists feel comfortable taking the software we wrote and modifying it for their application,” he says. “In our own lab, a lot of the advances we have week to week are software driven. We get an idea for an experiment, and we just reprogram the code.” That makes experiments much less expensive, and the weekly progress wouldn't be possible using off-the-shelf systems, he adds.

Zhuang's group has created software that allows image processing to happen as the data are acquired. Unlike a conventional fluorescence image, a single STORM image is obtained by stacking many frames. “It used to be that we could only do post-image processing,” she says. “Now we have analysis software that you can use to see the build-up of the high-resolution image as it is being recorded.” This will speed up experiments, she says, because if there is something about the sample you dislike, you can change it before the image construction is finished.

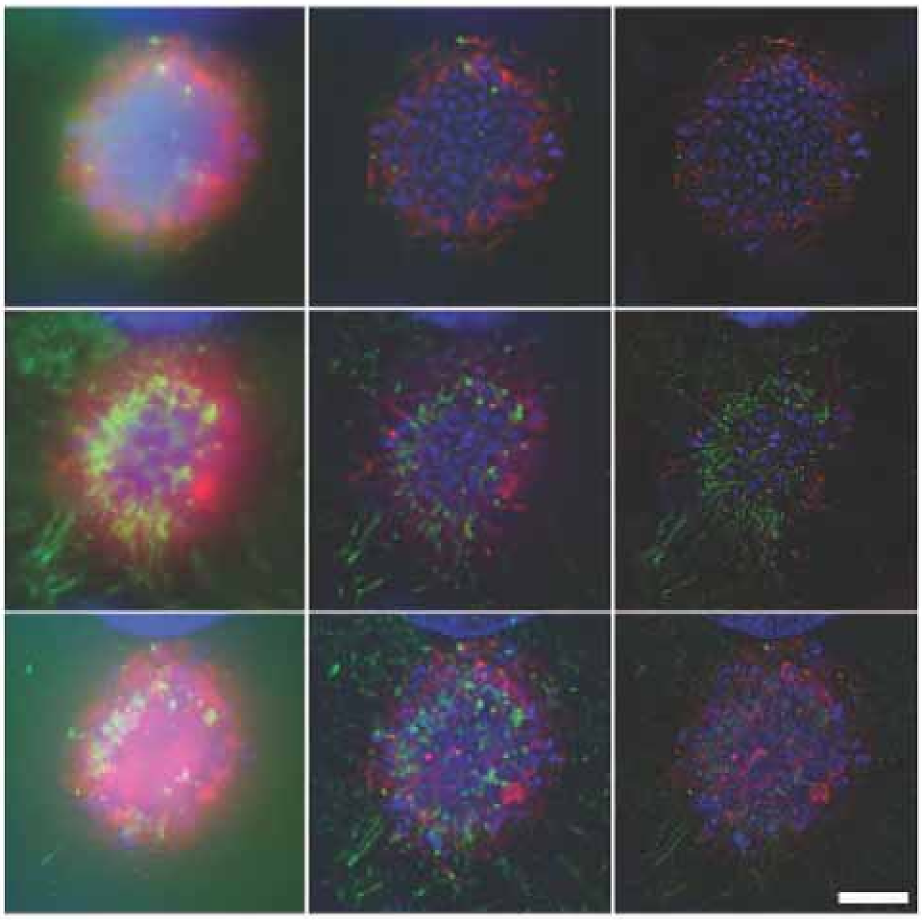
One-stop system?

Biologists still have a long wish list for super-resolution imaging technology, probes and software. “We're at least one or two years away from the technology being fully established,” says Jan Liphardt, a biophysicist at the University of California, Berkeley. “My own preference is to keep building my own hardware.”

Although improvements are happening fast, and developers are quickly adding to the pool of 3D and live-cell imaging data, there is still no one system that can readily do everything a biologist wants, says M. Cristina Cardoso, a cell biologist at the Technical University of Darmstadt, Germany.

Cardoso and her colleagues used 3D-SIM

and spatially modulated illumination to compare the sizes of DNA replication spots — called replicons — on chromosomes within a cell nucleus, and how they change over time. Replicons are, on average, 125 nm in diameter. Although these super-resolution techniques allowed her to identify three- to fivefold more replicons than the numbers previously reported[9], the researchers would have preferred to conduct the experiment in live cells to get better temporal resolution. However, with current systems this would have meant giving up some spatial resolution, so instead they stopped replication and fixed the cells at various points in their cycle.



D. HIGASHI, D. ELLIOTT, M. SO

Epithelial cells seen by (left to right) wide field, deconvolution and OMX microscopy. Scale bar: 4 μm. Top row, midplane through microcolony; centre, bottom of colony; bottom, maximum intensity projections.

Rivalling Hubble

Live-cell approaches are improving, but there are still challenges. In STORM, PALM and SIM, multiple pictures must be taken and then blended together to reveal all of the detail. “Does that represent structure or motion of some sort?” Moerner says, adding that it is an ongoing challenge to distinguish the two.

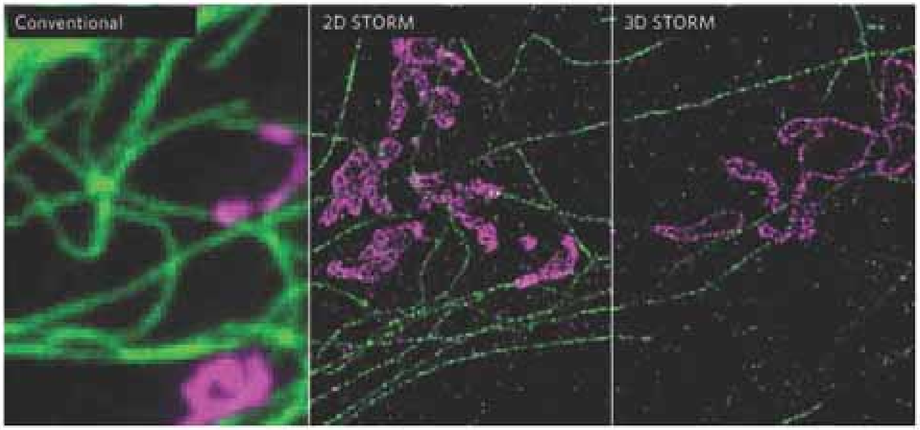
The challenges have not stopped large collaborations from harnessing super-resolution microscopes to answer big questions. Liphardt says that the University of California, Berkeley, has seven to ten PALM/STORM systems, built by researchers there and at other labs, that will be used for research into tumour-cell signalling as part of a US$15.7-million grant awarded by the US National Cancer Institute.

About a year ago, a consortium formed, which includes Zhuang and Smith, to construct a detailed map of neuronal connections in the brain. This ‘connectome’ project will combine super-resolution microscopy, including STORM, with array tomography and electron microscopy to see neuronal details at a resolution of 100 nm or better over the entire brain of a mouse, and eventually a human. To get a complete wiring diagram of the brain, says Zhuang, one needs to resolve the long neuronal projections — axons — that are invisible under a traditional light microscope.

“We’re starting to get pictures out of the mouse brain that rival anything from the Hubble Space Telescope, and we’re just getting started,” Smith says. But to make use of that information will require a lot more than looking at pretty pictures: the new imaging data will also push forwards development of software used to reconstruct and analyse connections in three dimensions. “The tools for comparing and understanding the elaborate circuit diagrams are beyond anything that anyone has thought of today,” he says.

The promise of super-resolution microscopy — thought for so long to be little more than a dream — is starting to become a reality. Researchers have taken different approaches and are using tools and techniques borrowed from physics, chemistry and computing technology to bring the nanoscopic world to our macroscopic eyes. Although commercialization is progressing, there is still plenty of room for the do-it-yourself biologist to modify and improve their systems, and produce images of stunning complexity that will rival anything else in science. ■

Kelly Rae Chi is a freelance journalist based in Cary, North Carolina, USA.



ZHUANG LAB, HARVARD UNIV./HHMI; LICHTMAN LAB,HARVARD UNIV.

Composite two-colour image of microtubules (green) and mitochondria (magenta) in a cell.

1. Moneron, G. & Hell, S. W. *Opt. Express* **17,** 14567–14573 (2009).
2. Ding, J. B., Takasaki, K. T. & Sabatini, B. L. *Neuron* **63,** 429–437 (2009).
3. Shtengel, G. *et al. Proc. Natl Acad. Sci. USA* **106,** 3125–3130 (2009).
4. Pavani, S. R. *et al. Proc. Natl Acad. Sci. USA* **106,** 2995–2999 (2009).
5. Kner, P., Chhun, B. B., Griffis, E. R., Winoto, L. & Gustafsson, M. G. *Nature Meth.* **6,** 339–342 (2009).
6. Subach, F. V. *et al. Nature Meth.* **6,** 153–159 (2009).
7. van de Linde, S. *et al. Photochem. Photobiol. Sci.* **8,** 465–469 (2009).
8. Han, K. Y. *et al. Nano Lett.* **9,** 3323–3329 (2009).
9. Baddeley, D. *et al. Nucleic Acids Res.* doi:10.1093/nar/gkp901 (2009).
10. Klar, T. A., Jakobs, S., Dyba, M., Egner, A. & Hell, S. W. *Proc. Natl Acad. Sci. USA* **97,** 8206–8210 (2000).
11. Betzig, E. *et al.* Science **313,** 1642–1645 (2006)
12. Hess, S. T. *et al. Proc. Natl Acad. Sci. USA* **104,** 17370–17375 (2007).
13. Manley, S. *et al. Nature Meth.* **5,** 155–157 (2008).
14. Huang, B., Wang, W., Bates, M. & Zhuang, X. *Science* **319,** 810–813 (2008).
15. Huang, B., Jones, S. A., Brandenburg, B. & Zhuang, X. *Nature Meth.* **5,** 1047–1052 (2008).
16. Westphal, V. *et al. Science* **320,** 246–249 (2008).

by different coloured light. By generating overlapping images in different colours and compiling them, the system can produce three-dimensional images that have a spatial resolution ten times better than confocal images in each dimension, Zhuang says. Nikon's instrument, called N-STORM, should be available by May 2010.

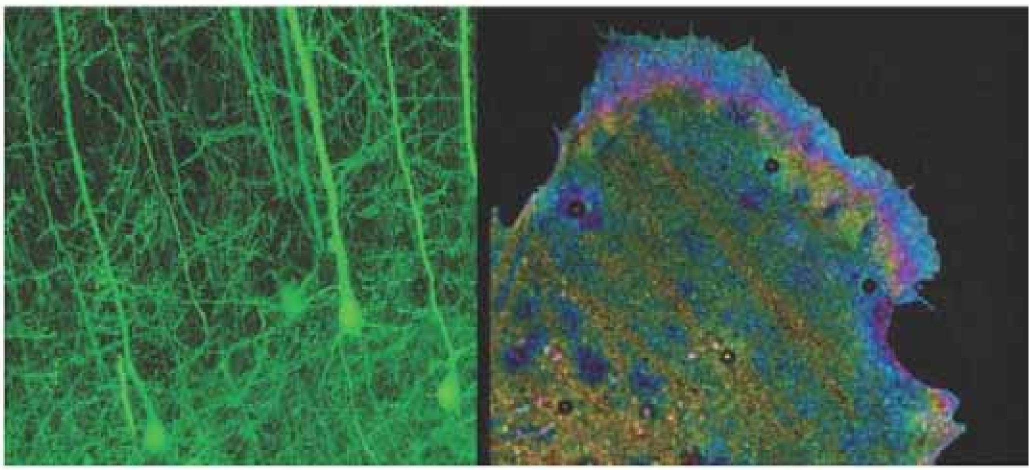
Technology tweaks

The pioneers of super-resolution microscopy are continuing to improve their methods with better sample preparation, a few strategically placed pieces of hardware and more sophisticated algorithms.

One way to achieve greater resolution in fixed samples is to cut thinner slices of tissue. Some groups, including Zhuang's, are collaborating to combine STORM with array tomography, which involves embedding fixed tissue in resin and cutting ultrathin (50–200 nm) slices. "Array tomography is a potential adjunct to all the super-resolution methods," says Stephen Smith of Stanford University in Palo Alto, California, who first described the method. Using this combination, his group has achieved a 14-fold improvement in axial resolution, and twofold improvements in the each of the two dimensions, compared with confocal imaging.

Other combination strategies are possible. For example, two-photon excitation microscopy, an established method that allows researchers to see further into tissues, can be combined with STED microscopy. Earlier this year, Hell's group used the technique on mammalian cell nuclei tagged with fluorescent nanoparticles, obtaining a resolution of less than 70 nm in the focal plane[1].

Bernardo Sabatini, a neurobiologist at Harvard Medical School in Boston, Massachusetts, used the same combination to image slender neuronal projections and dendritic spines in brain slices, which are not visible using conventional methods[2]. Sabatini and his team built the system themselves. They created the ring shape in the depletion laser using a phase plate to split the laser light so that one part served as the central spot and the other part as the ring. The result was a three-fold improvement in spatial resolution over normal two-photon techniques.



New microscopy techniques are showing structures (pyramidal neurons, left and a U2OS bone cell, right) in unprecedented detail.

N. WEILER, K. MICHEVA, ST. SMITH/ STANFORD UNIV.

H. HESS

Advances in resolution have not been restricted to two-dimensional samples. Harald Hess and his colleagues at Howard Hughes Medical Institute's Janelia Farm Research Campus in Virginia have modified PALM to do 3D imaging by adding an interferometer and a second objective lens[3]. Called iPALM, this method can pinpoint fluorescent labels to within 10–20 nm in three dimensions. However, the samples must be around 225–250 nm thick, and ensuring that the extra optics needed are perfecty aligned is a time-consuming task. As a result, iPALM might prove challenging to commercialize, Hess says.

Groups led by Moerner and Rafael Piestun at the University of Colorado, Boulder, have

BREAKING THE LIGHT BARRIER

In 1873, German physicist Ernst Abbe proposed that diffraction fundamentally limited the resolution that any microscope could achieve to around half the wavelength of light. And, despite many advances, microscopes didn't threaten to challenge this law of physics for more than a century. Stefan Hell, now a director of the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany, was the first to show that the diffraction limit could be beaten.

Hell, while a postdoc at the University of Turku in Finland in the 1990s, thought that, with the right lasers, he could activate a fluorescent spot and then shrink it by superimposing a larger, hollow beam of light to deplete all the light emission except for that at the centre of the spot. He called the technique stimulated emission depletion (STED) microscopy. Although many physicists were initially sceptical of Hell's ideas, by 2000 he had used STED to produce the first nanoscale fluorescence images[10].

Super-resolution microscopy has blossomed since, allowing researchers to see cellular processes unfolding at nanometre scales. "This is something that the field has desired since people first started looking through light microscopes," says Jan Liphardt, a biophysicist at the Lawrence Berkeley National Laboratory in California.

Since Hell's work, the field has been boosted by other groups, including those of Eric Betzig, a physicist at the Howard Hughes Medical Institute's Janelia Farm Research Campus in Virginia and Jennifer Lippincott-Schwartz, a cell biologist at the National Institutes of Health in Bethesda, Maryland. In 2006, the groups reported that they had increased resolution by harnessing single-molecule photoactivatable fluorescent proteins and compiling images of thousands to millions of them[11]. They called the approach photo-activated localization microscopy (PALM).

At Harvard University in Cambridge, Massachusetts, physicist Xiaowei Zhuang has developed three-dimensional (3D) stochastic optical reconstruction microscopy (STORM), which uses photoswitchable probes to temporally separate the overlapping images of individual molecules and so boost resolution to ten times better than the diffraction limit.

Yet another approach — fluorescence PALM (FPALM) — was developed in 2006 by Samuel Hess, a physicist at the University of Maine in Orono. His group's technique involves looking at thousands of fluorophores at once, and localizing on small numbers at a time. These methods have already begun to demostrate their utility. For example, in 2007, Hess's group showed that FPALM could be used to detect proteins clustering in lipid rafts in living cells[12]. In 2008, Lippincott-Schwartz's group combined PALM with single-particle tracking to detect the movement of membrane proteins in live cells[13]. And Zhuang's group used 3D STORM to image microtubules and other molecular structures within monkey kidney cells[14], later extending the method to multicolour 3D imaging of whole cells[15]. Hell's group, early in 2008, used the STED method to show the movement of synaptic vesicles inside living neurons at video rate[16].

But the field is just warming up. "To people like me who were trained in physics or optics in the 1990s, it's just unbelievable that one can image below the resolution of light," says Bernardo Sabatini, a neurobiologist at Harvard Medical School in Boston, Massachusetts. "The major revolution for the next 5 or 10 years is getting these advances to answer biological questions."

K.R.C.

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL
Microscopy and imaging equipment			
Andor Technologies	Cameras, imaging software, confocal microscopes	Belfast, Northern Ireland	**www.andor.com**
Applied Precision	Super-resolution microscope system and software	Issaquah, Washington	**www.api.com**
Attocube Systems	Ultra-high-precision spatial positioning systems and scanning-probe microscopes.	Munich, Germany	**www.attocube.com**
Carl Zeiss	Super-resolution microscopes, image-analysis software	Jena, Germany	**www.zeiss.com** ●
Chroma Technology	Optical filters	Rockingham, Vermont	**www.chroma.com**
Coherent	Lasers and optical components	Santa Clara, California	**www.coherent.com**
Conix Research	Motorized stages, remote focus accessories, computerized Z-motor, epi filter changers, filters	Springfield, Oregon	**www.conixresearch.com**
Del Mar Photonics	Laser systems for multiphoton imaging, scanning-probe microscopy	San Diego, California	**www.dmphotonics.com**
Electro-Optical Products	Standard, custom, OEM components and sub-systems with drive electronics	Fresh Meadows, New York	**www.eopc.com**
FEI	Electron and ion-beam microscopes and tools for life sciences	Hillsboro, Oregon	**www.fei.com**
Holoeye	Diffractive optical elements and spatial light modulators	Berlin, Germany	**www.holoeye.com**
Image Optic	Wavefront sensing hardware and software; adaptive optics technologies	Orsay, France	**www.imagine-optic.com**
JENOPTIK Laser, Optik, Systeme	Lasers, optics, micro-optics and sensors; optical and micro-optical components	Jena, Germany	**www.progres-camera.com**
JPK Instruments	Nano-analytical instruments, particularly based on atomic-force microscope and optical tweezers systems, for life-sciences and soft-matter applications	Berlin, Germany	**www.jpk.com**
Lab Essentials	Microscopes, cameras and laboratory supplies	Monroe, Georgia	**www.labessentials.com**
Leica Microsystems	Super-resolution systems and image-analysis software	Wetzlar, Germany	**www.leica-microsystems.com**
Mad City Labs	Flexure-based nanopositioning systems	Madison, Wisconsin	**www.madcitylabs.com**
MathWorks	Image acquisition, process and analysis; images and video for algorithm development and system design	Natick, Massachusetts	**www.mathworks.com**
McCrone Microscopes and Accessories	Polarized light microscopes, Linkam heating/cooling stages, variety of digital-imaging systems	Westmont, Illinois	**www.mccronemicroscopes.com**
MicroImage Video Systems	Video cameras for life-science applications	Boyertown, Pennsylvania	**www.mivs.com**
Micron Technology	Displaytech liquid-crystal spatial light modulators	Boise, Idaho	**www.displaytech.com**
Micro-optik	Hardware and software for image analysis	Deursen, the Netherlands	**www.micro-optik.com**
Mir Enterprises	Consultancy, design, simulation, fabrication and testing of micro- and nanotechnology systems	London, UK	**www.mir-enterprises.com**
NanoAndMore	Equipment and consumables for nanotechnology research; imaging software; vibration-isolation equipment	Wetzlar, Germany	**www.nanoandmore.com**
Newport	Lasers and light sources, opto-mechanical components and mounts, optical filters and gratings	Irvine, California	**www.newport.com**
Nikon	Microscopes, objectives, digital cameras, imaging software and analysis tools	Melville, New York	**www.nikoninstruments.com**
NKT Photonics	Fibre-based lasers and light sources	Birkerød, Denmark	**www.nktphotonics.com**
Olympus America	Microscopes and imaging systems; digital imaging; *in vivo* imaging solutions	Center Valley, Pennsylvania	**www.olympusamerica.com**
Omega Optical	Optical filters and coatings for imaging, custom filters	Brattleboro, Vermont	**www.omegafilters.com**
OZ Optics	Lasers and fibre-optic components	Ontario, Canada	**www.ozoptics.com**
Park Systems	Systems, software and expertise for the nanoscale measurements required in life sciences	Suwon, South Korea	**www.parkafm.com**
Physick Instrumente	Stages, linear motors and precision motion-control equipment for microscopy	Karlsruhe, Germany	**www.physikinstrumente.com**
Pixon	Image-enhancement solutions for medical imaging and microscopy applications	Stony Brook, New York	**www.pixon.com**

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL	
QImaging	CCD and CMOS digital cameras for quantitative image analysis and high-resolution images	Surrey, British Columbia, Canada	**www.qimaging.com**	
RPC Photonics	Vortex phase plates and diffractive optical elements	Rochester, New York	**www.rpcphotonics.com**	
SYNCROSCOPY	Image-analysis and acquisition software	Cambridge, UK	**www.syncroscopy.com**	
Thorlabs	Prisms, lenses and optical components	Newton, New Jersey	**www.thorlabs.com**	
TOPTICA Photonics	Diode and ultrafast fibre lasers	Munich, Germany	**www.toptica.com**	
Uniblitz	Electronic shutters for microscopy, CCD, video and low-level laser switching	Rochester, New York	**www.uniblitz.com**	

Fluorescent probes

Biotium	Fluorescent dyes and indicators; cell-biology and apoptosis assay kits	Hayward, California	**www.biotium.com**	
BioVision	Fluorescent protein vectors, antibodies, tools and kits for cell-biology research	Mountain View, California	**www.biovision.com**	
Clontech	Fluorescent proteins and reagents, fluorescent reporter systems	Mountain View, California	**www.clontech.com**	
Evrogen	Fluorescent proteins and molecular-biology systems	Moscow, Russia	**www.evrogen.com**	
GE Healthcare	Fluorescent proteins for cell analysis and drug-screening applications	Chalfont St Giles, UK	**www4.gelifesciences.com**	
Invitrogen	Fluorescent proteins, dyes and quantum dots	Carlsbad, California	**www.invitrogen.com**	
Mo Bi Tec	Fluorescent dyes and conjugates; fluorescent calcium indicators	Göttingen, Germany	**www.mobitec.de**	

General

Biomol	Services for chemical synthesis, cell culture and antibody production	Hamburg, Germany	**www.biomol.de**	
Bio-Rad	Products, instruments and software for life-sciences research	Hercules, California	**www.bio-rad.com**	●
Eppendorf	Consumables for molecular biology; instrumentation	Hamburg, Germany	**www.eppendorf.com**	●
LOT-Oriel	Scientific instrumentation for imaging and spectroscopy	Darmstadt, Germany	**www.lot-oriel.com**	
MatTek	*In vitro* human-tissue equivalents (human reconstructed tissue)	Ashland, Massachusetts	**www.mattek.com**	
MDS Analytical Technologies	Liquid-handling and microplate-processing equipment; imaging instruments	Sunnyvale, California	**www.moleculardevices.com**	
Merck	Chemicals, kits and reagents for molecular- and cell-biology-related research	Darmstadt, Germany	**www.merck.de**	●
Metrohm USA	Laboratory instruments, consumables	Riverview, Florida	**www.metrohmusa.com**	
PerkinElmer	Instruments, reagents and kits for life-sciences research	Waltham, Massachusetts	**las.perkinelmer.com**	
QIAGEN	Gene-amplification kits; sample and assay technologies	Venlo, the Netherlands	**www1.qiagen.com**	
Roche Diagnostics	Reagents and kits for molecular biology; genomics instrumentation and software	Basel, Switzerland	**www.roche-applied-science.com**	●
SPI Supplies	Sample preparation equipment; consumable supplies; accessories for microscopy	West Chester, Pennsylvania	**www.2spi.com**	
Takara Bio	Reagents, kits and consumables for molecular biology	Shiga, Japan	**www.takara-bio.com**	●
TEMwindows.com	Support films featuring pure silicon TEM windows with 5–15 nm thick windows as amorphous films or with 10–50 nm pores for high-resolution imaging	West Henrietta, New York	**www.temwindows.com**	
Thermo Fisher Scientific	Chemicals and reagents for life-sciences research; instruments, robotics and automation	Waltham, Masachusetts	**www.thermofisher.com**	

● **see advertisement**

Q&A

Jorge Gardea-Torresdey of the University of Texas–El Paso received the 2009 Distinguished Scientist Award from the Society for the Advancement of Chicanos and Native Americans.

Did you have a career-defining moment?

My paternal grandfather owned silver and gold mines in northern Mexico. I started going with him to the mines when I was five, six, seven years old and I fell in love with the mining process.

I asked why the different minerals were different colours and why they were shiny. That got me interested in chemistry.

What challenges did you encounter in your career?

Growing up as a Mexican American I faced much discrimination. My classmates used to call me 'dumb Mexican' even though I got straight As. My professors used to say, "Why do you think you're so smart?"

But the United States is changing, and in Texas, where there is a large proportion of Hispanics, I see the new future. We are accepted now more than we were in the past. I just received this beautiful award, and I am really pleased about it.

Did your family accept your career choice?

I grew up in a very wealthy business family that was involved in many types of business — mining, furniture stores, real estate. I am the oldest of ten kids and it was expected I would go into the businesses. But I used to hide out and read books on chemistry when I was supposed to be watching the employees in one of our furniture stores. Eventually, I told my father and grandfather that I wasn't going to be a businessman, and they were very upset.

Have you ever thought about leaving Texas?

I was considered for head of the Institute of the Environment at the University of California, Los Angeles; other opportunities were for federal posts. But I stayed here in Texas because I want to be valuable and helpful to Hispanics.

My goal is to see the completion of our new chemistry and computer-sciences building on campus in about a year.

What was your most important scientific discovery?

In 2002 my group discovered that alfalfa plants will take up gold from gold-enriched soil, forming gold nanoparticles inside their tissues.

Our finding is important because gold nanoparticles can be heated to 427 °C and burn cancer cells. Harvesting these particles from plants could help us develop a cure for cancer.

What's your biggest career satisfaction?

Mentoring students. They are the future of the United States, and it is more satisfying to me than research. I produced some of the first Hispanic scientists in my field, and several now work for the US Environmental Protection Agency — one is an environmental scientist in Dallas and another is an environmental engineer along the US–Mexico border. They are doing good for the environment. ■

Interview by Karen Kaplan.

POSTDOC JOURNAL

Inspiration and satisfaction

When I was little, I wanted to build a sled. I envisioned an enclosed sled that glided along on three runners; I spent many evenings drawing a detailed steering mechanism, sketching a plan for the passenger capsule and carefully planning the sled's dimensions. I imagined how it would fly over the snow. I was sure I would build it, and I trembled with impatience to start. But it never came to be.

I had several boyhood projects that burned intensely in my imagination, but then faded in the light of reality. At the beginning of my graduate career, I was terrified that my research would have the same fate as those projects: burning inspirations that ultimately amounted to nothing as I moved on to my next idea.

Three years ago, when I learned that my first manuscript had been accepted, the excitement was electric. I had finally taken a moment of inspiration and followed it to completion. With each subsequent manuscript acceptance, I become more confident that I can in fact finish the projects that inspire me.

Today, looking at a recently published paper I co-authored, I think of the excited moments of inspiration that preceded it, as well as my year and a half of work, some of it tedious. My satisfaction on seeing the paper in print is worth much more than those transient inspirations.

As a boy, I avoided work in favour of the joy of inspiration; as a scientist, I've at last started to fully appreciate the final published product. ■

Sam Walcott is a postdoc in theoretical biophysics at Johns Hopkins University in Baltimore, Maryland.

IN BRIEF

More PhDs for women

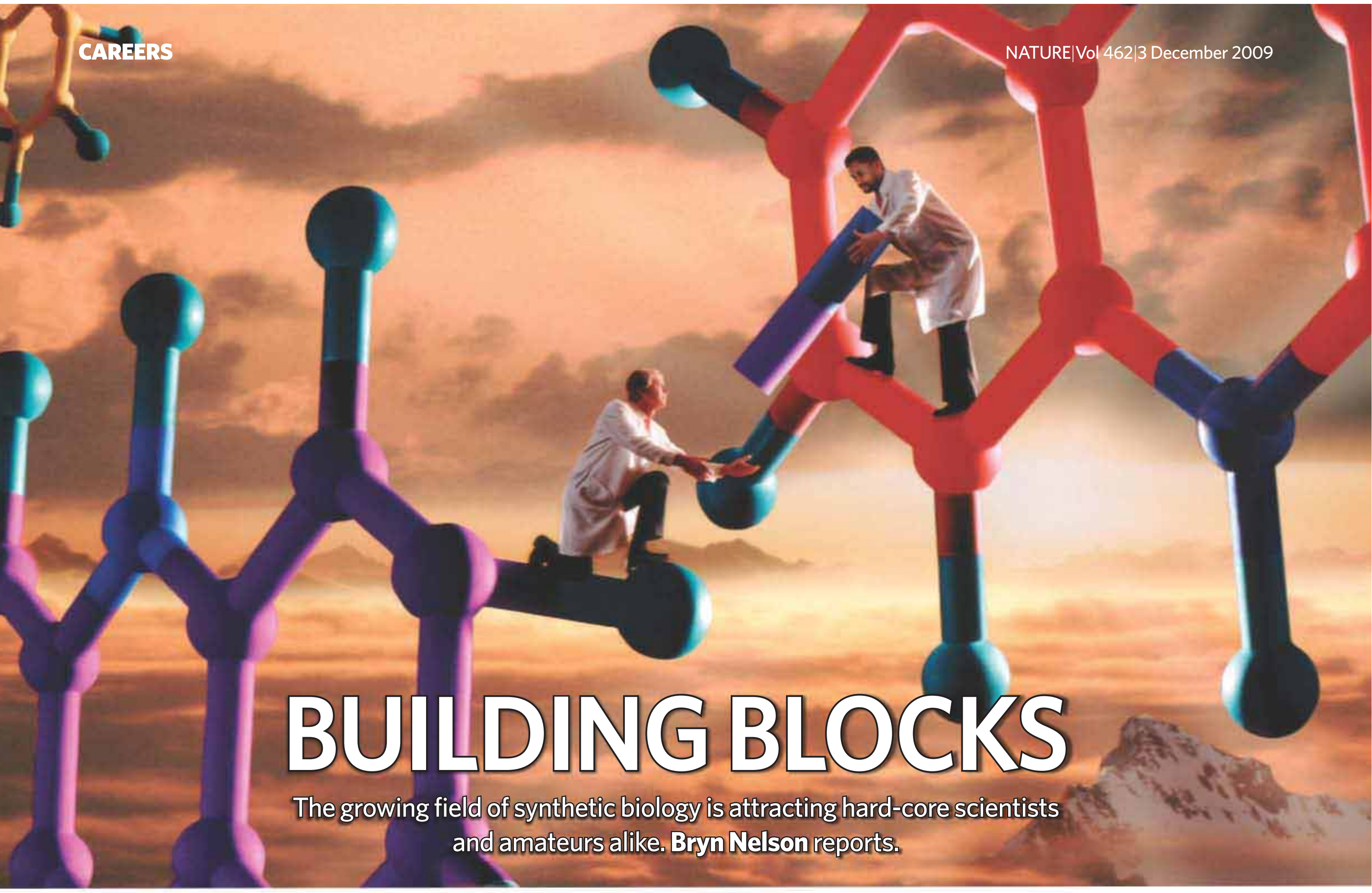
Women accounted for most of the growth in the number of science and engineering doctorates awarded in the United States between 2007 and 2008, according to the US National Science Foundation's 2008 Survey of Earned Doctorates (see go.nature.com/8HRei6). The report, released on 19 November, says that 32,827 science and engineering doctorates were awarded in 2008, up 3.2% from 2007. Of those, 12,959 — or 39.5% — were awarded to women, an increase of 5.7% or 699 over 2007 numbers. The number awarded to men rose to 19,845, an increase of 1.7%. The proportion awarded to women rose from 37.7% in 2003 to 39.5% in 2008; men's share dropped from 62.3% to 60.4%.

Grants follow conference

At least ten scientists, engineers and medical researchers will receive seed grants of up to US$100,000 each from the US National Academy of Sciences to explore issues surrounding the emerging field of synthetic biology (see page 684). Only those who attended a 19–22 November synthetic-biology conference run by the academy in Irvine, California, are eligible to apply. The meeting was organized in conjunction with the W. M. Keck Foundation, which is funding the grants. Recipients will be announced in April 2010. Conference participants worked on developing a common understanding of goals, exploring tools to facilitate engineering biology, and identifying applications for the field.

This time it's personalized

A new 'personalized medicine' institute hopes to seed a biomedical cluster and attract scientists. The private non-profit Ignite Institute will be based in Virginia's Fairfax County at a 27,000-square-metre facility to be built in part with $150 million of county bond money that Ignite will begin repaying after 5 years. The state of Virginia and Inova Health System will each provide $25 million. Ignite plans to employ 415 staff within 5 years, including 350 researchers in areas such as cardiovascular and metabolic disease. Graduate students will attend through partnerships with neighbouring universities. Ignite chief executive Dietrich Stephan says that affiliation with a health-care system will make personalized-medicine advances swiftly available to patients.



B. ALLEN/CORBIS

BUILDING BLOCKS

The growing field of synthetic biology is attracting hard-core scientists and amateurs alike. **Bryn Nelson** reports.

Those charting the meteoric rise of the field of synthetic biology need look no further than an international competition that boasts an impressive assortment of entries. One team of students introduced a rainbow of pigments into *Escherichia coli* to point the way towards inexpensive environmental biosensors. Another team of art-and-design students built their own gel boxes, fashioned microscopes out of webcams, and valiantly tried to re-engineer *E. coli* to smell like the earthy, moist scent of rain after a dry spell. A third team used Lego bricks to build a US$700 automated DNA-assembly robot.

The International Genetically Engineered Machine competition, more popularly known as iGEM, began in 2004 with modest aims and five undergraduate teams from the United States. It has grown to encompass increasingly sophisticated proposals from more than 100 teams hailing from 25 countries on five continents. Each spring, teams of 6–12 students and two or more academic advisers begin charting out a new project — accomplished mainly over the summer — that will be judged in the autumn. Most teams design biologically based systems, and others compete to build the best software tools. Under the competition's broad definition of an undergraduate, students from community colleges and even high schools are getting in on the act. Each team is separately backed by its own sponsors; winners earn medals and bragging rights. And many team members aspire to make some aspect of synthetic biology their life's pursuit.

Generally speaking, synthetic biology refers to research aimed at designing, constructing and evolving new genomes or biologically inspired systems. Its ranks are difficult to quantify because researchers often fall under separate bioengineering, synthetic-chemistry or systems-biology headings. Adherents cannot even agree on whether 'synthetic biology' actually now denotes a distinct scientific field or whether it's a wise label, from a public-relations perspective. But semantics aside, the science is being driven by a concept dating back to the Industrial Revolution: the ability to construct interchangeable parts more precisely, cheaply and quickly. The difference is that these parts happen to be living and exquisitely small.

Synthetic biology goes beyond the basic strategy of recombinant DNA by not only adding new genes but also engineering new metabolic pathways — an approach that could lead to other breakthroughs. "There is enormous job potential, not just in academia but also in industry for this field," says Virginia Cornish, a chemist at Columbia University in New York, who considers herself a synthetic biologist.

Several forecasters agree. Market-research company BCC Research in Wellesley, Massachusetts, pegged the global market for synthetic biology at US$234 million in 2008, divided nearly equally among three segments: chemicals and energy, biotechnology and pharmaceuticals, and research and development. The report suggests that the world market could expand tenfold to $2.4 billion by 2013, with the chemicals and energy segment dominating.

Antimalarial breakthrough

Major foundations and pharmaceutical companies are taking notice of the potential for drug development. For example, seminal research by Jay Keasling at Lawrence Berkeley National Laboratory and the University of California, Berkeley, led to a $42.6-million grant from the Bill & Melinda Gates Foundation in 2004 to further his work on a partially synthetic artemisinin antimalarial drug. It was then licensed by California start-up Amyris Biotechnologies and transferred to French drug firm Sanofi-Aventis in 2008 for large-scale production.

Smaller ventures such as GENEART in Regensburg, Germany, have also thrived. The 180-employee company bills itself as the "worldwide leading supplier for synthetic genes". A specialist in made-to-order DNA and genes, GENEART aspires to design new

J. HASELOFF

The iGEM team and others from the Royal College of Art, London School of Economics and the University of Cambridge at the Haseloff lab, Cambridge.

functions based on standardized components. Chief executive Ralf Wagner says the company sees iGEM students not only as potential employees but as potential customers.

Meanwhile, the emphasis on open-source information and readily accessible tools has attracted the new do-it-yourself biology movement, whose enthusiasts are linking up through sites such as diybio.org. Once compared to ham-radio hobbyists or basement inventors, these budding researchers are now proposing scientific projects and angling for iGEM invitations.

One branch of the field focuses on the use of biological entities for computation. Another seeks to construct life from the bottom up, exemplified by the plans of the J. Craig Venter Institute in Rockville, Maryland, which wants to use the bacterium *Mycoplasma genitalium* as the template for a stripped-down, largely synthetic version known as *Mycoplasma laboratorium*. "A lot of people think in terms of engineering and how cells are like computers," says graduate student Christina Agapakis, who began her Harvard graduate career in an interdisciplinary programme before gravitating towards synthetic biology. "I think about it more like 'How can we put together all the stuff I learned about in college and what can we make with that?'"

Although many of Agapakis's experiments and tools would be familiar to most scientists, she says it's the bottom-up approach to biology that sets the field apart. A willingness to take risks is also key. "Nothing that I proposed in my thesis panned out in the end," Agapakis concedes, noting that her thesis committee helped her to redirect her research.

Q. WANG

In 2003, Lawrence Berkeley National Laboratory opened the world's first synthetic-biology department at a major research institution. It now has 24 research scientists and 135 support staff, visiting researchers and students. In 2006, the department's leaders helped secure a $16-million grant from the National Science Foundation to create the separate Synthetic Biology Engineering Research Center (SynBERC). Similar centres are concentrated in a few US and European hotspots, notably Boston, California's Bay Area, Britain, Germany and France.

But the field is quickly gaining ground in other countries, including China and Japan. Asian universities fielded 14 iGEM teams last year. This year they had 24.

As interest builds, undergraduate, graduate and even summer courses in synthetic biology are popping up around the world. So are master's and PhD programmes. Given the field's rapid evolution, however, textbooks are all but useless — which is where the competition comes in.

Newer centres, such as Imperial College London's Institute of Systems and Synthetic Biology, integrate iGEM into undergraduate coursework to help direct students towards a graduate career in synthetic biology. "It's certainly been the idea behind how we've structured our centre," says Richard Kitney, the institute's chairman. "We've actually now got a pipeline." The institute won $12 million in government funding late last year and is hiring three new faculty members.

The Wellcome Trust, the largest UK non-governmental granting agency, is pursuing a similar strategy by funding multiple iGEM teams for next year's competition. "The more successful they become, the more attractive the subject area will be for other scientists," says Ruth Jamieson, the trust's science-portfolio adviser. "The students are key to this. They bring a lot of different disciplines into the competition." The trust also hopes to foster ethical discussions by allowing teams to include ethicists and philosophy students. "We're very keen that the public engagement and ethical consideration of synthetic biology is developed alongside the actual science," Jamieson says.

David Rejeski, the director of the Synthetic Biology Project at the Woodrow Wilson International Center for Scholars in Washington DC, says that nanotechnology has provided some early parallels for how synthetic biology might fare economically, although he believes the newer field probably faces a rockier ethical road, given public attitudes towards 'engineering' life.

"If it goes the way of nanotechnology, which I think is possible, you could have a fairly large explosion in demand, which would occur across all kinds of disciplines," Rejeski says. The attraction for biologists, engineers, chemists and computer scientists, he says, could make it a powerful job driver. Ancillary people are likely to take part as well, he says, including risk-assessment specialists, ethicists and those focused on public outreach.

Riding a new-tech wave

To help bridge the knowledge gap, Rejeski foresees a big demand for educators who are savvy about social-media applications and can reach academics, citizen biologists and the public alike. Twitter is being used as a way to point out new tools, results and ideas. Explanatory videos by iGEM teams and tutorials by the field's leaders are attracting new YouTube audiences. Google Wave, an online communication platform, isn't even available to the general public yet, yet an iGEM team from the University of Freiburg in Germany has already used it to construct a collaborative synthetic-biology application called SynBioWave.

The field attracts the unconventional too. High-school dropout Zack Booth Simpson is now a software engineer, artist and research fellow at the Center for Systems and Synthetic Biology at the University of Texas in Austin. He helped advise the university's first iGEM team in 2004. At a brainstorming session, the group came up with the idea of re-engineering bacterial cells so that a biofilm of *E. coli* could compute the edge of an image, with cells marking the boundary between light and dark by producing a black sugar. Their first image was "Hello World". Earlier this year, the group's bacterial-edge detector was published in *Cell* (J. Tabor *et al.* *Cell* **137,** 1272–1281; 2009) in collaboration with a group from the University of California, San Francisco.

Despite this success, Simpson concluded that cells are terrible as engineering devices and has turned to the more reliable computing power of DNA and RNA. Next on his to-do list is a DNA-based system that can do the microscopic equivalent of the crowd's stadium wave at football games. The field, he says, is "most amenable to people who are excited by the crazy possibilities". ■

Bryn Nelson is a freelance writer based in Seattle, Washington.

Press ‘1’ to begin

The art of conversation.

Nye Joell Hardy

JACEY

Andrew? Andrew? Are you there?

I know this is an online class, but the reason you have an audio interface is so that you can talk. Andrew? Can you hear me? Hello?

If you’re frightened, I want you to know you’re not alone. ‘Social Interaction 101’ is designed for people just like you. It’s not an uncommon difficulty these days, not since everyone started telecommuting. Even people who come from the largest families lose their verbal skills if they spend just a few years communicating only by e-mail, blog and tweet. And you haven’t been outside your house since — let me look at your profile — oh, my. Well, never mind, Andrew. Can you just say “Hi?” Andrew? No?

Okay. I’ll just talk and you can listen. No … don’t type your response. Just listen. Over the duration of the course we will cover History of Social Interaction; Breathing and Talking at the Same Time; Paying Attention to What People Say; Common Greetings; Uncomfortable Silences; What is Rudeness?; Communicative Emotions to Avoid — Silliness, Salaciousness and Sarcasm; The Ten Commandments of Manners; Weather and Sports; and Eye Contact.

Eye Contact used to be in the advanced course, but it really is an important component of social interaction, so we decided it belonged in 101. Facial Expressions are still in the advanced session, of course, and Being A Shoulder to Cry On, but I’m getting off topic here. Still with me?

You could burp or cough, if you want, even though it’s not good manners. We’ll cover all that in a later class, but I think you just need to get used to making sounds in front of others, Andrew. How about if you just clap your hands? … No?

Don’t worry then. Some people need to go through the first few classes before they’re comfortable with me reacting to the sounds they make. But think about this, Andrew: wouldn’t you like to be able to put ‘Conversationalist’ on your résumé? Let’s see what you do for a living — Human Resources Director — well. I think conversing could be a powerful tool in your management arsenal, don’t you? Andrew?

Never mind. Would you mind if I start our first topic? Would you?

All right. The very first humans were probably non-verbal, but they needed a code to understand their relationships to one another — who was boss and who had to do all the nasty chores. For this code, they created words, as they were illiterate and didn’t have computers. Different tribes of people quickly developed different codes: Chinese, Latin, Farsi, Rap … and all the other languages, you know, that were in the world. But, they had no technology. Then, someone invented Semaphore and Morse Code and DOS, and people quickly realized that talking wasn’t as efficient at communicating data.

This is true, Andrew, we all know it. And I know you might find this hard to believe, but as primitive as it is, verbal communication is still the best form for social interaction. When I think about those early human beings, I like to imagine how they communicated with one another. And do you know what I think was the earliest form of conversation?

The knock-knock joke.

A knock-knock joke is easy. It doesn’t require any emotional commitment. It’s non-threatening. This is how it works. I say “Knock-knock”, and you say, “Who’s there?” When I say a word, you repeat the word and say “Who?”

Would you like to try one? Andrew?

I’ll start by saying “Knock-knock”, just as if I were knocking on your front door, just like those delivery people who come to your house all the time. As an aside, delivery people are a remarkable holdover from pre-E-volution days: their social interactions have been studied at length, especially as they seem to recognize people they haven’t seen for months, and sometimes even years. But I digress.

Ready? Knock-knock.

Well, if you had said “Who”, I would have said “Cow.” Then you would have said “Cow who?” and I would have said, “No! Cow moo!”, and I would have laughed.

That’s just a fabulous knock-knock joke.

The ironic thing, of course, is that our cultural loss of social interaction was pretty much predicted. I’m going to read you a quote from a book. It was written way back in 1974. It’s called *The Private Future* by Martin Pawley. I’ll also send you the quote for your homework: it may be a little difficult for you to understand, here in its spoken form. Here goes:

“Alone in a centrally heated, air-conditioned capsule, drugged, fed with music and erotic imagery, the parts of his consciousness separated into components that reach everywhere and nowhere, the private citizen of the future will have become one with the end of effort and the triumph of sensation divorced from action.”

Pretty deep, huh?

First of all, Andrew, I told you not to type responses. You can only speak in this class. Second: yes. I do think this quote is talking about you, Andrew. I think this quote is talking about all of us.

Never mind. Our class time is drawing to an end, Andrew. Like I said, you don’t have to talk in these first classes if it’s too stressful. But as I can see from your class application, you said you desperately wanted to learn about social interaction. That could be motivational for you, too, I think. Don’t you think that would make it so much easier to do what you want to do? Let’s see what you wanted to do; I know you wrote it down here. Ah, yes.

Don’t you think, Andrew, that the ability to make conversation would make it so much easier to go out on a real date?

Andrew? Are you there? ■

Nye Joell Hardy is the food safety manager for an international produce company. She has rarely spoken to and has never met the people who buy her stories.

Join the discussion of Futures in *Nature* at go.nature.com/QMAm2a