THIS WEEK

Mismeasure for mismeasure

A critique of the work of Stephen Jay Gould should serve as encouragement to scrutinize the celebrated while they are still alive.

It is impossible to libel the dead, but equally impossible for them to defend themselves. That alone is reason for caution when it comes to questioning the work of scientists who are no longer with us.

Such questions have grown into a fascinating cottage industry, with reports and papers taking issue with historical research, sometimes centuries after the fact. Notable examples include the 1978 critique by Gerald Holton, a physicist and historian at Harvard University in Cambridge, Massachusetts, of data selection in the reporting of the electric charge on oil droplets by Nobel-prizewinning physicist Robert Millikan in 1913; and historian Richard Westfall's 1973 exposure of mathematical fudging by Isaac Newton in the seventeenth century.

Sometimes, such critiques are themselves questioned, such as in 2007, when Harvard biologist Daniel Hartl and Daniel Fairbanks, a biologist at Utah Valley University in Orem, came to the defence of Austrian monk Gregor Mendel, who was criticized by British statistician Ronald Fisher in 1936 over data that demonstrated genetic inheritance patterns in pea plants just a little too neatly.

This month sees the latest episode: an assault on the work of US evolutionary biologist and celebrated author Stephen Jay Gould, who died in 2002. Although the critique leaves the majority of Gould's work unscathed, it carries a special sting because it deconstructs a posthumous attack that Gould launched on nineteenth-century physician Samuel Morton. In a 1978 paper (S. J. Gould *Science* **200,** 503–509; 1978) and in his 1981 book *The Mismeasure of Man*, Gould argued that Morton's measurements of the cranial capacity of hundreds of skulls from worldwide populations, reported in works published between 1839 and 1849, were unconsciously biased, by what he claimed was the physician's prejudice that caucasians were more intelligent, and therefore would have larger skulls. As Gould was canny enough to realize, a charge of unconscious bias sticks faster in science than one of fraud.

BLIND MEASUREMENT

Now, in a paper published on 7 June, Jason Lewis, an anthropologist at Stanford University in California, and his colleagues test Gould's assertions in detail (J. Lewis *et al. PLoS Biol.* doi:10.1371/journal.pbio.1001071; 2011). They remeasured the volume of some 300 skulls in Morton's collection, which survives at the University of Pennsylvania's Museum of Archaeology and Anthropology in Philadelphia, while taking care to blind themselves to knowledge of the population that each skull came from. Comparing their measurements to Morton's, they find no evidence that his were distorted by bias. Still, because they couldn't measure all the skulls, they do not know whether the average cranial capacities that Morton reported represent his sample accurately. (Cranial capacity varies mostly as a function of overall body size and stature, which is related to climate and nutrition, and there is no clear evidence of a link between cranial capacity and intelligence.)

Lewis and his colleagues also claim to find errors in Gould's statements about Morton's data. For example, Gould claimed that Morton manipulated his grouping of samples to give the results he wanted, arbitrarily amalgamating Native American populations, while breaking down those of people of European origin into subgroups. Yet, Lewis and his colleagues say that Morton reported average cranial capacities for subgroups of both populations, sometimes on the same page or on pages near to figures that Gould quotes and therefore must have seen. Furthermore, they say that Gould misdefined the Native American samples, falsely inflating the average he calculated for that population, which Gould had used to show that Morton's average was erroneously low.

Although the new paper does not accuse Gould of intentionally misrepresenting Morton, some of its authors have raised this possibility in interviews, noting that Gould's oversights would be less troubling were he known to be a less meticulous scholar. At a minimum, Gould's staunch opposition to racism, and desire to make an example of Morton, may have biased his interpretation of Morton's data, opening Gould to charges of hypocrisy.

"Gould's staunch opposition to racism may have biased his interpretation of Morton's data."

Of course, Lewis and his colleagues have their own motivations. Several in the group have an association with the University of Pennsylvania, and have an interest in seeing the valuable but understudied skull collection freed from the stigma of bias (although, as for many nineteenth-century museum collections, its ethically dubious assembly will remain an issue). Second, their paper makes clear that they oppose the view, espoused by Gould and trumpeted by some social scientists, that the scientific method is inevitably tainted by bias. Third, in contrast to others who may have taken Gould's politically correct message at face value, at least two authors have expressed the view that scientists must be free to establish the scientific facts even when the message may be misinterpreted by those with repugnant social goals.

But these motivations are not a reason to discount the group's critique. By documenting their methods and data, as they argue Morton did, the paper's authors have made it possible for others to scrutinize their claims. Transparent documentation should allow science as a whole to be objective, even if individual authors are not.

Just as important is the readiness of the scientific community to undertake such studies, and to see them through the sometimes difficult publication process. The criticism of Gould was rejected by the journal *Current Anthropology*, and spent eight months in the review process at *PLoS Biology*. And although an undergraduate did publish a more modest study scrutinizing Gould in 1988, it is remarkable that it has taken more than 30 years for a research group to check Gould's claims thoroughly. Did Gould's compelling writing and admirable anti-racist motivations help to delay scrutiny of his facts? Quite possibly, and this is regrettable. Although future historians will be happy to scrutinize our most persuasive and celebrated luminaries, today's scientists should not leave the job to them. ■

Damned if they do

An industry approach to greener hydropower is far from perfect, but it does offer a way forwards.

The mighty Iguaçu Falls in Brazil are an excellent illustration of the power of water, so what better place for the hydropower industry to promote what it says is a fresh approach to its sustainability?

There is ample room for scepticism about the effort — known as the Hydropower Sustainability Assessment Protocol (see page 430). It is an industry-led endeavour that requires next to nothing from the industry. It grades hydropower projects, but makes no judgement on what should happen to projects that rank poorly. And it is geared towards assessment of individual dams, independent of broader questions about energy-resource development. So far, so bad. Yet, if deployed properly, it could also be an invaluable tool to inject much-needed science and reason into a planning process that has operated with little of either for much too long.

Developers and governments have historically assessed dam projects mainly on the basis of cost and power. Engineers simply survey the landscape to identify the easiest places to block channels, set up turbines and run power lines. Sediments, endemic species and the consequences of severing communication between headwaters and estuaries are very much secondary issues. Even people get short shrift, leading indigenous groups to mount the kind of intense protests that last week saw the Peruvian government shelve plans for a massive dam in the Amazon.

This standard approach has caused numerous environmental problems — such as siltation and blockages to migrating fish — in industrialized countries, which exploited their best hydropower resources long ago and are now trying to repair the damage. In some cases, the costs of improvement outweigh the benefits, and old dams are being decommissioned. But, in the developing world, hydropower projects continue to stack up. Countries in Southeast Asia and Latin America, in particular, are pursing hydropower with gusto, hoping to alleviate energy poverty and feed burgeoning economies. By one optimistic industry estimate, cumulative hydropower capacity could nearly double by 2030. Without a more coordinated approach, these countries are doomed to make the same mistakes.

The new hydropower protocol comes courtesy of the International Hydropower Association, which consulted with environmental and human-rights groups, as well as representatives from finance and government, in an effort to set out some basic principles of sustainable hydropower.

After three years of work, the result is a way to assess dam projects on a range of criteria — from planning, governance and public engagement to ecology and hydrology. It is voluntary, however, and there are no minimum standards. The protocol asks all the right questions but fails to provide any answers.

This has driven a wedge into the community of environmental and social activists that work in this arena. Critics argue that the protocol represents little more than a public-relations exercise that will allow bad developers to appear green while pursuing business as normal — often on projects that pre-date current environmental thinking. This may be true, but, unfortunately, in the political and corporate world such 'greenwash' is common. The new effort would at least create a common language with which to raise concerns, evaluate the best available science and negotiate improvements.

"The hydropower assessment protocol asks all the right questions but fails to provide any answers."

The biggest shortcoming lies in the assessment of individual dams that have already been proposed for specific locations. Much better would be an approach to analyse entire river basins in an effort to identify the most suitable locations, as well as areas where special precautions should be taken. Indeed, it might well be that some rivers should be left to flow freely to preserve ecological integrity.

The protocol does touch on these issues, raising questions about a dam's role in the broader energy mix and about wider impacts from hydroelectric development. And it could yet offer a foundation to set minimum standards in these and other areas, so that companies would need to build and operate better dams, as well as integrate them into a more comprehensive energy strategy. For all of its faults, the protocol opens another bridge to a better future. Now it's up to governments, banks and companies to make the journey across. ■

Poetry in motion

A quantitative approach to the humanities enriches research.

The Oscar-winning 1989 film *Dead Poets Society* is an unabashedly exuberant story that appeals to the Lord Byron in each of us. Robin Williams plays a charismatic English teacher at a conservative US prep school in the 1950s and, in one scene, gets his impressionable students to read a lesson from a poetry textbook aloud. The worth of a poem, they read, should be measured on two axes: its artistic perfection and its importance. As the schoolboys start to map out graphs in their notebooks, Williams cuts them off. "Excrement," he announces — that's what he thinks of the mathematical approach. A poem must be felt, not figured. He orders the boys to tear the page out of the textbook. "We're not laying pipe," he says. "We're talking about poetry."

It is hard to disagree with the spirit of that moment. We should all be passionate about our academic interests, and daring enough to rip up hidebound rules that govern them. But the scene's explicit disdain for quantitative analysis of text is as out of date as it is wrong. These days, it is the humanities scholars who equip themselves with quantitative skills who are most able to sound their 'barbaric yawps' over the roofs of the world, as Williams urged his students to do.

As the News Feature on page 436 shows, the field of digital humanities is flourishing, led by scientists such as those behind the innovative Google *n*-grams viewer, which can be used to track the frequencies of words and phrases as they appear in 4% of the books ever published. Whether mapping the transmission of Voltaire's letters across Europe, finding structural patterns in music across cultures or tracking the evolution of irregular verbs through time, these digital humanists have plenty to say. And they have the data to back it up.

That is not to say that traditional approaches in the humanities will be disappearing any time soon, or that careful, interpretive readings by experienced scholars are as arbitrary as the learning-by-feeling espoused by *Dead Poets Society*. But digitization is marching on, and in all subjects, researchers who have their ears to the ground, rather than their heads in the sand, can hear the approaching drums. Every day, more and more of the media that make up both historical and contemporary culture are being converted to electronica.

It seems just a matter of time before the humanities, like the social sciences before them, wholeheartedly embrace scientific methodology. And that should be reason to rejoice, not remonstrate. As Williams implored his young charges: *carpe diem*. Seize the day. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

Damned if they do

An industry approach to greener hydropower is far from perfect, but it does offer a way forwards.

The mighty Iguaçu Falls in Brazil are an excellent illustration of the power of water, so what better place for the hydropower industry to promote what it says is a fresh approach to its sustainability?

There is ample room for scepticism about the effort — known as the Hydropower Sustainability Assessment Protocol (see page 430). It is an industry-led endeavour that requires next to nothing from the industry. It grades hydropower projects, but makes no judgement on what should happen to projects that rank poorly. And it is geared towards assessment of individual dams, independent of broader questions about energy-resource development. So far, so bad. Yet, if deployed properly, it could also be an invaluable tool to inject much-needed science and reason into a planning process that has operated with little of either for much too long.

Developers and governments have historically assessed dam projects mainly on the basis of cost and power. Engineers simply survey the landscape to identify the easiest places to block channels, set up turbines and run power lines. Sediments, endemic species and the consequences of severing communication between headwaters and estuaries are very much secondary issues. Even people get short shrift, leading indigenous groups to mount the kind of intense protests that last week saw the Peruvian government shelve plans for a massive dam in the Amazon.

This standard approach has caused numerous environmental problems — such as siltation and blockages to migrating fish — in industrialized countries, which exploited their best hydropower resources long ago and are now trying to repair the damage. In some cases, the costs of improvement outweigh the benefits, and old dams are being decommissioned. But, in the developing world, hydropower projects continue to stack up. Countries in Southeast Asia and Latin America, in particular, are pursing hydropower with gusto, hoping to alleviate energy poverty and feed burgeoning economies. By one optimistic industry estimate, cumulative hydropower capacity could nearly double by 2030. Without a more coordinated approach, these countries are doomed to make the same mistakes.

The new hydropower protocol comes courtesy of the International Hydropower Association, which consulted with environmental and human-rights groups, as well as representatives from finance and government, in an effort to set out some basic principles of sustainable hydropower.

After three years of work, the result is a way to assess dam projects on a range of criteria — from planning, governance and public engagement to ecology and hydrology. It is voluntary, however, and there are no minimum standards. The protocol asks all the right questions but fails to provide any answers.

This has driven a wedge into the community of environmental and social activists that work in this arena. Critics argue that the protocol represents little more than a public-relations exercise that will allow bad developers to appear green while pursuing business as normal — often on projects that pre-date current environmental thinking. This may be true, but, unfortunately, in the political and corporate world such 'greenwash' is common. The new effort would at least create a common language with which to raise concerns, evaluate the best available science and negotiate improvements.

"The hydropower assessment protocol asks all the right questions but fails to provide any answers."

The biggest shortcoming lies in the assessment of individual dams that have already been proposed for specific locations. Much better would be an approach to analyse entire river basins in an effort to identify the most suitable locations, as well as areas where special precautions should be taken. Indeed, it might well be that some rivers should be left to flow freely to preserve ecological integrity.

The protocol does touch on these issues, raising questions about a dam's role in the broader energy mix and about wider impacts from hydroelectric development. And it could yet offer a foundation to set minimum standards in these and other areas, so that companies would need to build and operate better dams, as well as integrate them into a more comprehensive energy strategy. For all of its faults, the protocol opens another bridge to a better future. Now it's up to governments, banks and companies to make the journey across. ■

Poetry in motion

A quantitative approach to the humanities enriches research.

The Oscar-winning 1989 film *Dead Poets Society* is an unabashedly exuberant story that appeals to the Lord Byron in each of us. Robin Williams plays a charismatic English teacher at a conservative US prep school in the 1950s and, in one scene, gets his impressionable students to read a lesson from a poetry textbook aloud. The worth of a poem, they read, should be measured on two axes: its artistic perfection and its importance. As the schoolboys start to map out graphs in their notebooks, Williams cuts them off. "Excrement," he announces — that's what he thinks of the mathematical approach. A poem must be felt, not figured. He orders the boys to tear the page out of the textbook. "We're not laying pipe," he says. "We're talking about poetry."

It is hard to disagree with the spirit of that moment. We should all be passionate about our academic interests, and daring enough to rip up hidebound rules that govern them. But the scene's explicit disdain for quantitative analysis of text is as out of date as it is wrong. These days, it is the humanities scholars who equip themselves with quantitative skills who are most able to sound their 'barbaric yawps' over the roofs of the world, as Williams urged his students to do.

As the News Feature on page 436 shows, the field of digital humanities is flourishing, led by scientists such as those behind the innovative Google *n*-grams viewer, which can be used to track the frequencies of words and phrases as they appear in 4% of the books ever published. Whether mapping the transmission of Voltaire's letters across Europe, finding structural patterns in music across cultures or tracking the evolution of irregular verbs through time, these digital humanists have plenty to say. And they have the data to back it up.

That is not to say that traditional approaches in the humanities will be disappearing any time soon, or that careful, interpretive readings by experienced scholars are as arbitrary as the learning-by-feeling espoused by *Dead Poets Society*. But digitization is marching on, and in all subjects, researchers who have their ears to the ground, rather than their heads in the sand, can hear the approaching drums. Every day, more and more of the media that make up both historical and contemporary culture are being converted to electronica.

It seems just a matter of time before the humanities, like the social sciences before them, wholeheartedly embrace scientific methodology. And that should be reason to rejoice, not remonstrate. As Williams implored his young charges: *carpe diem*. Seize the day. ■

➲ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

Supreme Court ruling is good, bad and ugly

Monday's key US legal decision on emissions regulation was influenced by the unjustified attacks on climate science, says **Douglas Kysar.**

The US Supreme Court this week gave its opinion on *American Electric Power* v. *Connecticut*, a closely watched lawsuit that seeks to force some of the nation's largest electricity generators to cut their greenhouse-gas emissions because they contribute to climate change, which is a public nuisance.

In its second major encounter with climate change, the court held that federal judges have no authority to order emissions reductions using nuisance law, because Congress has delegated this authority to the Environmental Protection Agency (EPA) under the Clean Air Act.

The court's ruling has some good, bad and downright ugly implications for US climate policy.

Here are the good. The opinion solidified the court's landmark 2007 conclusion that the EPA has the power to regulate carbon dioxide as a pollutant. This is the basis of the EPA's beleaguered efforts to use the Clean Air Act to address emissions.

In addition, the court turned down the opportunity to block climate-change nuisance suits on broad grounds, involving the legal doctrines of standing and political question. Such a move would have had serious negative effects for any lawsuit that challenges environmental, health and safety threats, whether related to climate change or not.

Next, because the court used the EPA's authority under the Clean Air Act as the basis to block federal nuisance lawsuits, any effort by Congress to repeal that authority could see the suits reinstated. Heavy emitters emphatically do not want to face such cases, so they may reduce their attempts to get Congress to neuter the EPA in this way. (They could instead push for legislation that would strip the EPA and the courts of authority, but that would constitute such an obvious plea for climate anarchy that, one hopes, it would have to fail.)

Finally, this week's decision left open the possibility that greenhouse-gas emissions might be challenged as nuisances under state common law. During oral argument in the case, Justice Antonin Scalia rather brazenly sought out legal theories that would block climate-change lawsuits in both federal and state courts. Such a result would have been premature — none of the parties was looking for such a broad ruling. Thus, the court reserved judgment on the issue. Plaintiffs in climate-change nuisance cases will still face obstacles under state common law, including the possibility that the Clean Air Act will be interpreted to block their actions, but for now their suits may proceed.

On to the bad. The court went out of its way to emphasize that federal common-law actions would be barred, even if the EPA decides not to regulate greenhouse-gas emissions. In other words, the fact that the agency has authority under the Clean Air Act — even if it chooses not to exercise it — was enough, in the court's view, to cut the judiciary out of the equation, stating, "We see no room for a parallel track."

The problem with this is that the US system of limited and divided government is a web of interconnected nodes, not a row of parallel tracks. The courts should understand that part of judges' role is to prod and plea with other government branches, which may be better placed to address an area of societal need, but are less disposed to try.

Federal judges are not well positioned to devise rules for greenhouse-gas emissions, given the complexity of the problem and its deep interrelation with other policy issues. But unless and until a comprehensive regulatory scheme is put into operation, the threat of common-law actions should remain part of the balance of powers that will shape whatever regime does eventually emerge. That is why it was essential for the court to leave open the possibility of state common-law claims. The threat of such suits adds legal, financial and public-relations pressure to the mixture of forces that drives policy outcomes.

PART OF JUDGES' ROLE IS TO **PROD AND PLEA** WITH OTHER **GOVERNMENT BRANCHES.**

I have saved the ugly for last. It is hard not to conclude that the judges were influenced by climate-science controversies of the past few years, however contrived and overstated they have been.

Although the Supreme Court's 2007 opinion referenced what "respected scientists believe" about climate change and relied on the findings of the Intergovernmental Panel on Climate Change (IPCC), the latest opinion stated pointedly, "The court, we caution, endorses no particular view of the complicated issues related to carbon-dioxide emissions and climate change." Worse, the court suggested that readers explore "views opposing the EPA's" by consulting "Dawidoff, The Civil Heretic, N. Y. Times Magazine 32 (March 29, 2009)".

Climate cognoscenti will recognize this reference as a profile of Freeman Dyson, the theoretical physicist whose controversial views on climate change have been widely promoted by the climate-sceptic community. The court also repeated a prominent sceptical refrain about the ubiquity and supposed banality of greenhouse-gas emissions — "after all, we each emit carbon dioxide merely by breathing" — that serves only to downplay the severity and significance of industrial emissions.

That the nation's highest court would repeat this misleading refrain, and seemingly endorse Dyson's views as equal to those of the IPCC and the EPA, simply takes the breath away. ■

➲ NATURE.COM
Discuss this article online at:
go.nature.com/1b1fmp

Douglas Kysar *is Joseph M. Field '55 Professor of Law at Yale Law School in New Haven, Connecticut, and the author of* Regulating from Nowhere: Environmental Law and the Search for Objectivity *(Yale University Press, 2010).*
e-mail: douglas.kysar@yale.edu

RESEARCH HIGHLIGHTS

Selections from the scientific literature

MOLECULAR PSYCHOLOGY

Cellular cues to compulsive eating

Insulin may play a key part in the development of obesity by acting directly on the brain's reward circuitry.

Jens Brüning at the Max Planck Institute for Neurological Research in Cologne, Germany, and his group genetically engineered mice in which they could inactivate insulin signalling in parts of the mid-brain that drive feelings of motivation and reward. The engineered mice ate more than control mice and became obese. Along the way, they developed mild insulin-resistant diabetes, although their metabolism was otherwise unaffected.

If similar mechanisms occur in obese humans, the authors say, it could help to explain why they tend to eat compulsively.
Cell Metab. 13, **720–728 (2011)**

ASTRONOMY

Blaze marks star's violent death

A singularly intense beam of light was spotted shining from the centre of a distant galaxy on 28 March after a massive black hole ripped a star apart, astronomers say.

The beam, detected by NASA's Swift satellite, lasted for more than two weeks, much longer than any known gamma-ray burst — explosions associated with supernovae that decline in minutes. According to a team led by Andrew Levan of the University of Warwick in Coventry, UK, the blast of light was also around 100 times brighter than light emitted by massive black holes lurking at the centre of galaxies, which are powered as they lap up nearby gas.

So what was it? A second team, led by Joshua Bloom at the University of California, Berkeley, suggests the engine was still a huge black hole, with the mass of at least one million Suns. But the fuel may have been a single star that strayed too close (artist's impression **pictured**).
Science **10.1126/science.1207143, 10.1126/science.1207150 (2011)**

M. GARLICK/UNIV. WARWICK

J. HAGER/ROBERT HARDING WORLD IMAGERY/CORBIS

PHYLOGENETICS

Picking out the pika's origins

The collared pika (*Ochotona collaris;* **pictured**) of Alaska and Canada is descended from its more southerly relative the American pika (*Ochotona princeps*), and not vice versa as previously thought.

Data on the two North American pikas have not allowed researchers to adjudicate between competing hypotheses on the animals' origins. So Kurt Galbreath of Cornell University in Ithaca, New York, and Eric Hoberg of the Animal Parasitic Diseases Laboratory in Beltsville, Maryland, studied parasites common to both pikas to determine the mammals' ancestry and colonization route.

The duo used genetic sequences to create family trees, which show that the parasites dispersed from south to north. Because they must have done so in their hosts, this means that the pikas moved likewise.
Proc. R. Soc. B **doi:10.1098/rspb.2011.0482 (2011)**

CANCER BIOLOGY

Senescence not so harmless

Skin-cancer cells that enter senescence, a form of growth arrest that some have proposed using to combat cancer, can rally neighbouring cells to become more invasive.

Previous studies have hinted that, in addition to hindering cancer cells' growth, senescence might promote the spread of nearby cells. Corine Bertolotto at the French National Institute of Health and Medical Research in Nice and her team found that melanoma cells treated with chemotherapy drugs did senesce, but also secreted proteins associated with inflammation.

Exposing untreated melanoma cells to CCL2, one of these proteins, or to the medium in which treated cells had been grown prompted the cells to metastasize. An inflammatory pathway regulated by the protein NF-κB controlled the composition of

secreted proteins, and when NF-κB was inhibited, the cellular secretions no longer promoted metastasis.
Genes Dev. **doi:10.1101/gad.625811 (2011)**

EVOLUTION

Male harassment can doom species

The males of many species demonstrate behaviours that are harmful to females. Some, for example, can physically damage females during courting or mating. Daniel Rankin at the University of Zurich in Switzerland and his team now demonstrate that such sexual conflict can lead to a 'tragedy of the commons'.

Their models suggest that the evolution of male harassment of females during mate competition can lead to a downward spiral of fewer surviving females; these females are themselves less likely to survive because they receive even more damaging attention from males. Conversely, if female resistance to damaging behaviour evolves, making harassment too costly for males, this can prevent a species from dying out.
Am. Nat. 177, **780–791 (2011)**

CELLULAR BIOLOGY

A microscopist's litmus test

Intracellular pH levels can now be measured with a protein that glows red.

The red fluorescent protein has been dubbed pHRed by its designers, Gary Yellen and his colleagues at Harvard Medical School in Boston, Massachusetts. Light-detecting molecules of pHRed are preferentially excited by light at a wavelength of either 440 nanometres in basic conditions or 585 nanometres in acidic conditions, and the excitation ratio of the two gives a read-out of pH.

pHRed's fluorescence lifetime — the time between excitation and emission — also responds to pH. The authors used this feature in two-photon microscopy to track second-by-second pH changes in living cells.

A red pH sensor leaves the often-used green fluorescence wavelength available for simultaneous study of other cellular properties, such as energy metabolism, during multicolour imaging.
J. Am. Chem. Soc. **doi:10.1021/ja202902d (2011)**

CANCER BIOLOGY

When push comes to shove in cancer

Ovarian cancer cells spread away from the site of a primary tumour by moving healthy cells out of the way, according to work by Joan Brugge of Harvard Medical School in Boston, Massachusetts, and her colleagues.

Brugge's team used time-lapse video microscopy to spy on ovarian cancer cells as they interacted with healthy cells in culture in a model of cancer metastasis. When the two cell types came into contact, the extracellular matrix surrounding the healthy cells disassembled, and they moved away from the point of contact with the tumour cells. This displacement, Brugge's group showed, depends on three proteins: $\alpha_5\beta_1$ integrin, talin 1 and myosin. In addition, the interaction with myosin and talin 1 is required to generate the force that physically pushes healthy cells out of the way.
Cancer Discov. **doi:10.1158/2159-8274.CD-11-0010 (2011)**

OCEAN CHEMISTRY

Cushion against acidification

When sulphur and nitrogen gases from ship fuel and other local sources of combustion end up in coastal waters, they strongly diminish ocean acidification by carbon dioxide.

Keith Hunter at the University of Otago in Dunedin, New Zealand, and his colleagues used a simple chemical model to compare the effects of sulphur and nitrogen oxides and ammonia on seawater chemistry with that of CO_2. Results from three case studies — in the North Sea, the Baltic Sea and the South China Sea — suggest that, after chemical re-equilibration, the net effect of the gases on ocean chemistry is negligible.

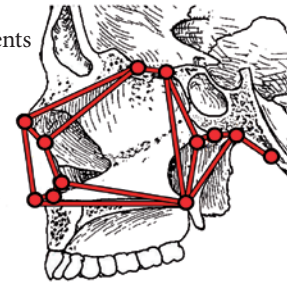
The authors conclude that in coastal regions with intense shipping activity, air–sea fluxes of acidic nitrogen and sulphur compounds effectively cancel out expected CO_2-induced ocean acidification by buffering reactions involving carbonate and bicarbonate ions in seawater. However, they suspect that such chemical buffering comes at the expense of shutting down much of the ocean uptake of excess atmospheric CO_2.
Geophys. Res. Lett. **doi:10.1029/2011GL047720 (2011)**

PHYSIOLOGY AND CLIMATE

Air conditioning all in the nose

Researchers have long speculated on the benefits offered by the variety of human nose shapes, but few have looked inside the facial feature — or the larger nasal cavity of the skull — for an answer.

Marlijn Noback of Eberhard Karls University in Tübingen, Germany, and her colleagues took computer-aided measurements of the sizes and shapes of nasal cavities (**pictured**) in a total of 100 skulls from 10 human groups living in 5 different climates. They found that individuals from cold, dry climates had higher and narrower cavities than those from hot, humid climates. The authors suggest that the high, narrow cavities may help to mix the air and increase its moisture content and temperature as it is inhaled.
Am. J. Phys. Anthropol. **doi:10.1002/ajpa.21523 (2011)**



WILEY-LISS

COMMUNITY CHOICE

The most viewed papers in science

BRAIN AND BEHAVIOUR

The voter's grey matter

HIGHLY READ on www.cell.com 15 May–14 June

The size of certain brain regions offer clues to a person's political leaning, according to Ryota Kanai at University College London and his colleagues.

In brain scans of more than 100 university students, self-reported conservatives had above-average volumes of grey matter in the right amygdala, whereas liberals had more of the grey stuff in the anterior cingulate cortex (ACC). This structural pattern fits with previous work that linked liberalism to high activity in the ACC, which is thought to help the brain sort out conflicting stimuli, as well as research correlating conservatism with higher levels of aggression under threat, an emotion probably processed by the amygdala.

Results aside, perhaps the reason that this paper was so often downloaded is that one of the co-authors is actor Colin Firth. Firth prompted the team to do the study for a BBC Radio 4 programme.
Curr. Biol. 21, **677–680 (2011)**

NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS *The news in brief*

POLICY

Pathogen dangers

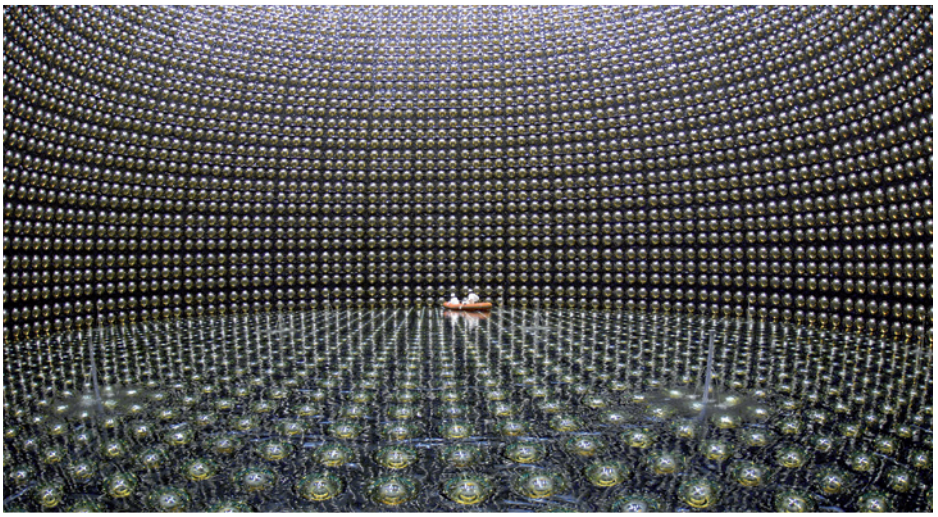
A US panel has named the most dangerous pathogens with potential for misuse. In a report released on 14 June, the Federal Experts Security Advisory Panel identified 11 biological agents that it said needed more rigorous monitoring, and recommended background checks for researchers working with them. The pathogens include bacteria and viruses that cause smallpox, the plague, anthrax, Ebola and foot-and-mouth disease. The panel also recommended that 19 agents be dropped from an existing list of 82 'select agents and toxins', including camel, goat and sheep pox viruses. The report was a response to an executive order from US President Barack Obama last year.

Asylum DNA tests

The UK Border Agency has ended a plan to use DNA samples to determine the nationalities of asylum seekers. The 'Human Provenance Pilot Project' was scientifically unworkable and was widely protested by geneticists and migration organizations (see *Nature* **461,** 697; 2009). *The Times* newspaper reported on 17 June that the project had been abandoned; the border agency says that it will not publish any evaluation of the scheme, nor the data collected.

Cut black carbon

Curbing ground-level ozone pollution and emissions of black carbon (soot particles) could cut half a degree from the global warming that is projected by 2030, save tens of millions of lives and protect agricultural crops, according to a scientific assessment released on 14 June. Sponsored by the United Nations Environment Programme and the World Meteorological Organization, the report says that black-carbon emissions from vehicles, cooking stoves and other sources could be reduced, and that emissions of methane — a precursor to ozone and itself a powerful greenhouse gas — could be cut from agricultural waste and oil and gas operations.



KAMIOKA OBS./ICRR/UNIV. TOKYO

Watching neutrinos change flavour

Some muon neutrinos that had been fired as a beam across the width of Japan changed into electron neutrinos on the way, physicists with the T2K (Tokai to Kamioka) multinational collaboration reported on 15 June (T2K Collaboration http://arxiv.org/abs/1106.2822; 2011). Previous experiments had found that neutrinos — almost massless fundamental particles — of one type (or 'flavour') seem to disappear as they travel. But the observations at the Super-Kamiokande detector (pictured) near Hida, Japan, are the strongest evidence yet that they reappear as neutrinos of a different flavour. With data-taking cut short prematurely by the earthquake in Japan on 11 March, the result falls short of the level of statistical significance needed to claim a discovery. But it raises the prospect that future experiments will be able to test the symmetry between matter and antimatter, by comparing neutrino oscillations with those of antineutrinos. See go.nature.com/ty8t3i for more.

Ethanol subsidies

The US Senate has voted to end costly federal subsidies for producing ethanol from maize (corn). The measure would eliminate a tax credit of US$0.12 per litre of ethanol and a separate tariff of $0.14 per litre on imported ethanol, and so would raise an extra $6 billion a year. The vote, on 16 June, does not spell an immediate end for subsidies, as it came in the form of an amendment attached to an economic-development bill with uncertain prospects of becoming law. But the decision suggests that lawmakers are prepared to ditch support for biofuels to help close the budget deficit. See go.nature.com/egbwop for more.

Horizon 2020

Europe's research-funding system after 2013 will be called 'Horizon 2020 — the Framework Programme for Research and Innovation', the European Commission revealed on 21 June. The name for the successor to the current €50-billion (US$72-billion) Seventh Framework Programme was chosen by an online public vote.

Global drug safety

Pronouncing itself at the centre of a "global bazaar", the US Food and Drug Administration (FDA) has announced its intention to assemble an international coalition of regulatory agencies to strengthen product safety worldwide. In a report released on 20 June, the FDA says that it will build a global data network that will allow regulators to proactively share real-time information. Margaret Hamburg, commissioner

of the FDA, says that the distinction between domestic and imported products is "obsolete"; the report notes, for example, that 80% of the active ingredients in medications sold in the United States come from elsewhere.

Climate lawsuit

The US Supreme Court has dismissed attempts by six states and other parties to force utility companies to reduce their greenhouse-gas emissions on the grounds that the emissions are a "public nuisance". The judgment, on 20 June, rules out the use of such lawsuits at the federal level to limit greenhouse gases. Instead, the Supreme Court pointed out that the Environmental Protection Agency already has the powers to impose such limits, through the Clean Air Act. See page 421 for more.

FUNDING

Plant grants

Fifteen plant-biology researchers have won a total of US$75 million from two private US foundations for their work. The Howard Hughes Medical Institute, based in Chevy Chase, Maryland, joined with the Gordon and Betty Moore Foundation in Palo Alto, California, to provide the money for what it called an "underfunded field". The

UCR STRATEGIC COMMUNICATIONS

winners, who will receive their grants over five years, include Philip Benfey at Duke University in Durham, North Carolina, who studies the development of plant roots from stem cells, and Xuemei Chen (**pictured**) at the University of California, Riverside, who looks at the formation of flowers. See go.nature.com/r4yblx for more.

EVENTS

Primate peril

Employees at a major US primate-research centre, and the animals it houses, all survived unscathed a massive chemical explosion at a nearby plant on 14 June. The New Iberia Research Center in Louisiana holds 6,500 macaques and 360 chimpanzees. Some 1,900 rhesus and pigtail macaques were housed within 350 metres of a fire and multiple explosions at the Multi-Chem facility in New Iberia, but none seemed to be harmed; they are now being monitored for stress and other adverse affects. Employees were safely evacuated from the centre. See go.nature.com/xow5dw for more.

Iran in orbit again

Iran has placed its second satellite into orbit, according to state media. The 15.3-kilogram 'Rasad' (Observation) satellite was launched on 15 June into an orbit 260 kilometres above Earth, and is transmitting images and telemetry data to tracking stations. Iran's first successful satellite launch was in February 2009. The country hopes to launch more satellites in the coming years, and to achieve human space flight by the end of the decade.

RESEARCH

Smashing data

The Large Hadron Collider has passed a key data milestone as it increases its rate of particle-smashing. On 17 June, the proton collider, located at CERN, Europe's particle-physics lab near Geneva, Switzerland, had delivered exactly one inverse femtobarn of collisions to its detectors. An inverse femtobarn is equal to about 70 trillion collisions. The Tevatron, a rival collider at Fermilab in Batavia, Illinois, has already gathered 11 inverse femtobarns of data, although at lower energies. Physicists hope to eke out 12 before the 26-year-old machine shuts down at the end of September.

Drug-approval race

Since 2003, the US Food and Drug Administration (FDA) has approved more cancer drugs than the European Medicines Agency (EMA), and approved those drugs more quickly, according to a report in *Health Affairs* on 16 June. Up to the end of March 2010, the US agency had approved 32 new anticancer drugs, whereas the EMA had approved 26. The median time from submission to approval was 182 days in the United States and 350 days in Europe. Janet Woodcock, the FDA's top drug-approval official, says that the difference extends to all categories of drug. See go.nature.com/ouxsup for more.

UK health research

The UK government has promised that medical research will receive greater attention in its revised proposals for reforming the country's public health service, published on 14 June. Britain's health minister would be given a new duty to "promote research" in the National Health Service, the plans say. See go.nature.com/x4qsbb for more.

COMING UP

25–30 JUNE
The complexity of RNA biology features in the Federation of European Biochemical Societies congress in Turin.
go.nature.com/kvlglw

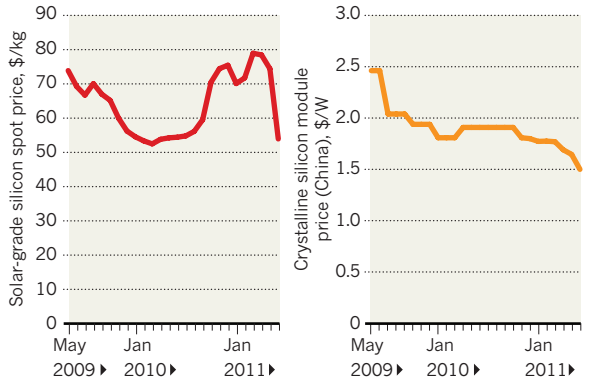
25 JUNE–2 JULY
An annual United Nations Food and Agriculture Organization meeting in Rome debates rising global food prices.
go.nature.com/l3naqa

TREND WATCH

The solar-energy industry's drive to cut costs got a sharp boost in June, after the spot price of solar-grade silicon fell by 28% from May. Raw materials for photovoltaics had been overpriced owing to high demand and national subsidies, says Jenny Chase, a solar-energy analyst at consultants Bloomberg New Energy Finance in London. But as subsidies are reduced and demand drops, "we're starting to see the bare bones of what it costs to make crystalline silicon modules — and it's low, which is exciting", she says.

SOLAR GETTING CHEAPER

The price of solar-grade silicon dropped sharply in June, while module prices continued their gradual decline.



SOURCE: BNEF

NATURE.COM
For daily news updates see:
www.nature.com/news

NEWS IN FOCUS

PATH/AMYNAH JANMOHAMED

A schoolgirl receives human papillomavirus vaccine as part of a feasibility study in India.

DRUG DEVELOPMENT

Vaccine trial's ethics criticized

Collapsed trial fuels unfounded vaccine fears.

BY PRIYA SHETTY

A clinical trial that came under fire in India threatens to have a dual legacy: inflaming unfounded fears about a lifesaving vaccine and raising new questions about the management of medical research in the country. After four teenage girls taking part in a test of human papillomavirus (HPV) vaccines died last year, the Indian government faced accusations that its citizens were being used as guinea pigs to test dangerous vaccines. A scientific investigation has exonerated the vaccines but uncovered a more familiar problem in India: ethical irregularities.

The study, funded by the Bill & Melinda Gates Foundation and run by the international health charity PATH and the Indian Council of Medical Research (ICMR), vaccinated more than 23,000 girls aged 10–14 against HPV, which can cause cervical cancer. The vaccines — Merck's Gardasil and GlaxoSmithKline's Cervarix — are already in widespread use in the developed world, and the study was designed to assess the feasibility of launching an HPV-immunization programme in the Indian health system. The researchers hoped to gauge public acceptance of the vaccines and assess the costs of administering it in different parts of the country.

A committee of three scientists from the All India Institute of Medical Sciences (AIIMS) in New Delhi, commissioned by the government to look into the trial, confirmed that the deaths were not linked to the vaccines — two of the girls died of poisoning, one of drowning and the fourth of a fever. But its report, leaked to India's media last month, said that the study involved several serious ethical violations. According to media reports, participants were recruited from vulnerable tribal populations, consent was improperly obtained — headmasters of the girls' schools signed the forms — and adverse events were poorly recorded.

The scientists also criticized Indian regulators for classifying the HPV study as an observational rather than a clinical trial, which meant that it was subject to different regulations, including looser reporting of side effects. The expert committee deemed it to be a clinical trial because it was a "study of a pharmaceutical product carried out on human participants" and "4 of 5 primary outcome measures proposed related to evaluation of the safety of the vaccine".

Vivien Tsu, director of PATH's HPV vaccines project, says that the procedures criticized in the report had all been approved by state ethics boards in India and an independent review board in the United States. "The problems the report raises, over the poor reporting of adverse events, for instance, were the sorts of issues that the study was intended to tease out," she says. Vishwa Katoch, director-general of the ICMR, says that his organization "had advised on ethical issues when the study was being planned. All necessary ethical approvals were there; the problem was how different individuals or teams implemented it."

Still, the verdict could pose a setback to the country's ambitions to become a hub for international clinical trials, luring drug developers with its large patient population and low costs. Rani Kumar, dean of the AIIMS, who assisted the investigating committee, declined to speak to *Nature*. But India's weak ethical infrastructure has been heavily criticized in the past for having few well-trained ethicists, and poorly run ethics boards. A clinical-trials registry was introduced in 2007 in a bid to better regulate clinical research, but "India still needs clear national guidelines on the ethical ▶

NATURE.COM
For more on vaccines and related issues, visit:
go.nature.com/a3nzqx

conduct of clinical trials", says Ramanan Laxminarayan, vice-president of policy and research at the Public Health Foundation of India in New Delhi. Shortly after the HPV report hit the headlines, the Drugs Controller General of India produced, for the first time, draft guidelines on the reporting of adverse events in clinical trials.

Heidi Larson, an anthropologist at the London School of Hygiene & Tropical Medicine who studies the social acceptance of vaccines, says that the collapse of the HPV trial highlights one of the key problems in research collaborations between developed and developing countries. Ultimately, she says, international researchers are obliged to work within the ethics capacity of the host country, regardless of whether or not it is robust. Trying to align different expectations over research ethics can be especially tricky, Larson adds. "How do you negotiate local versus national versus international tensions?"

Meanwhile, India's vaccination plans could suffer collateral damage from the controversy. India decided in 2008 to roll out a new pentavalent vaccine against diphtheria, pertussis, tetanus, *Haemophilus influenzae* type b (Hib) and hepatitis B. Public objections over fears of dangerous side effects reported in Sri Lanka and the vaccine's high cost delayed the programme. Vaccination is finally due to start next month, but only in Tamil Nadu and Kerala, southern states chosen because their routine immunization coverage is already high, says Ajay Khera, deputy commissioner of India's Universal Immunisation Programme.

Vaccination fears that have made headlines in the West are now taking hold in countries such as India, says Larson (H. J. Larson *et al. Lancet* doi:10.1016/S0140-6736(11)60678-8; 2011). Because most of the vaccines now in development are aimed at diseases common in the developing world, such as malaria, tuberculosis, leishmaniasis and helminth infections, anti-vaccination movements in such countries could have a major impact on public health, adds Richard Moxon, a paediatrician at the John Radcliffe Hospital in Oxford, UK.

Jacob Puliyel, head of paediatrics at St Stephen's Hospital in Delhi, has been a vocal opponent of both the HPV vaccines and the pentavalent vaccine. He does not endorse the vaccine fears that gripped the public after the four girls' deaths, but he told *Nature* that too little is known about the prevalence of Hib and HPV-related cervical cancer in India to justify the new vaccines. At a time when India is already struggling to achieve universal coverage with existing vaccines — coverage for basic childhood immunizations is just 63%, according to Khera — the country simply cannot afford them, Puliyel says. ■

JOURNALS

Open access comes of age

Publishing model enters phase of slower but steady growth.

BY JOHN WHITFIELD

A study of open-access publishing — published last week in the open-access journal *PLoS ONE* — has found that the number of papers in freely accessible journals is growing at a steady 20% per year (M. Laakso *et al. PLoS ONE* **6,** e20961; 2011). To many, the growth confirms the health of the free-access, author-pays model. But to a few it is a discouraging sign that open access is not about to take over the world of scholarly publishing.

The analysis, by information scientist Mikael Laakso of the Hanken School of Economics in Helsinki and his colleagues, also found that the number of fully open-access journals is growing at around 15% every year as new journals are founded and subscription journals switch to the open-access model (see 'Opening up'). By contrast, subscription journals are growing at about 3.5%. "Most indicators suggest growth is not slowing," says Laakso. "The open-access publishing model has proven itself to work."

Laakso divides the history of open access into three phases. First came the pioneering years of 1993–99, during which most open-access journals were, he says, "home-brew" efforts, set up by individuals and hosted on university servers. Next were the innovation years, which saw the birth of publishers such as the Public Library of Science and of software infrastructure that makes it much easier to launch a digital journal.

Since 2005, Laakso says, innovation has slowed but growth continues — the consolidation phase. Following this trend, last week Nature Publishing Group (NPG) launched *Scientific Reports*, an author-pays, open-access, online-only journal, which reviews papers on technical soundness rather than impact.

NPG's acknowledged inspiration is *PLoS ONE*, which in 2010 published 6,749 papers, making it the world's largest journal. It has been a "phenomenal success", says Jason Wilde, business development director at NPG. "It shows that authors and readers like the model of a broad-based journal with light peer review." *Scientific Reports* will provide *PLoS ONE* with a rival and help drive up standards, says Wilde. "In any market there should be competition."

The trends "show the success of open access", says Peter Suber, director of the Open Access project at the non-profit lobby group Public Knowledge in Washington DC. So far, he adds, the open-access movement has not imperilled commercial publishers. "The predictions of harm are being proven to be false."

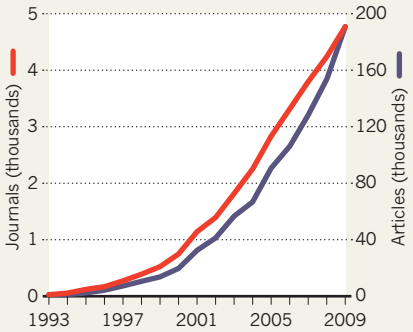
But not all advocates of open access are satisfied with its progress. "The growth rate is portrayed as dramatic, but it's not dramatic at all if the goal is 100% open access," says Stevan Harnad, a cognition researcher at the University of Southampton, UK. Other ways to make papers freely accessible, such as self-archiving and hybrid journals, which allow authors to choose whether to pay for open access, are also growing only linearly, he says. "The rate is much too low for the needs of research."

What is not known is whether open-access journals are competing with subscription journals, or whether they have opened up a publishing niche. Harnad believes that most open-access journals are new ventures. Because nearly all the must-have journals still charge subscription fees, the rise of the author-pays model actually imposes an extra expense on research funders, he says.

To escape this catch-22, says Harnad, institutions and funders — who have led the demand for open access — must mandate grantees to deposit papers published in subscription journals in open repositories. This would free up resources to support author-pays open access. "Publishers won't convert until the money is available to pay them," he says. ■

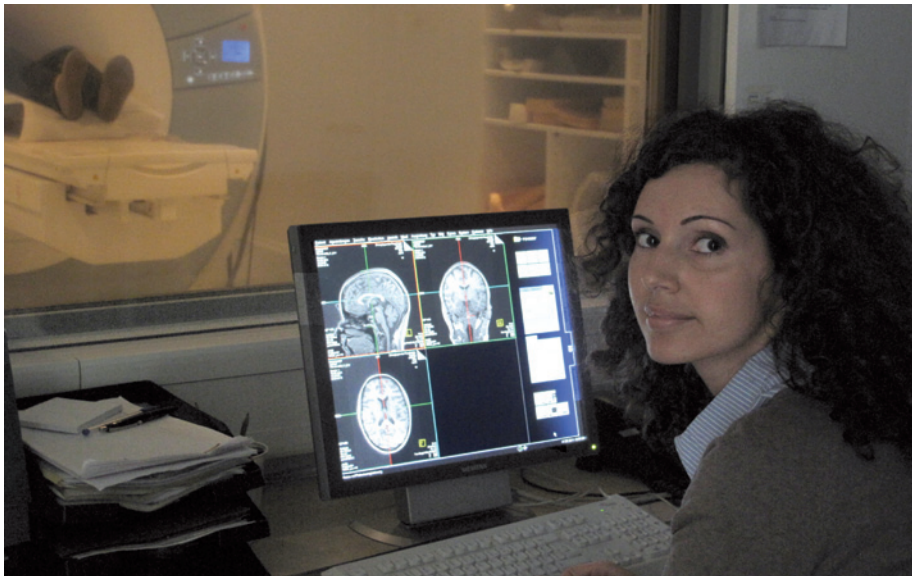
OPENING UP

The number of open-access articles and journals has been steadily increasing since 2000.



SOURCE: M. LAAKSO *ET AL. PLOS ONE* **6,** E20961 (2011)

S. WOLFF



At a research institute in Mannheim, Leila Haddad scolds subjects to stress them while imaging their brains.

NEUROSCIENCE

City living marks the brain

Neuroscientists study social risk factor for mental illness.

BY ALISON ABBOTT

Epidemiologists showed decades ago that people raised in cities are more prone to mental disorders than those raised in the countryside. But neuroscientists have avoided studying the connection, preferring to leave the disorderly realm of the social environment to social scientists. A paper in this issue of *Nature* represents a pioneering foray across that divide.

Using functional brain imaging, a group led by Andreas Meyer-Lindenberg of the University of Heidelberg's Central Institute of Mental Health in Mannheim, Germany, showed that specific brain structures in people from the city and the countryside respond differently to social stress (see pages 452 and 498). Stress is a major factor in precipitating psychotic disorders such as schizophrenia.

The work is a first step towards defining how urban life can affect brain biology in a way that has a potentially major impact on society — schizophrenia affects one in 100 people. It may also open the way for greater cooperation between neuroscientists and social scientists. "There has been a long history of mutual antipathy, particularly in psychiatry," says sociologist Craig Morgan at the Institute of Psychiatry in London. "But this is the sort of study that can prove to both sides that they can gain from each others' insights."

Meyer-Lindenberg works on risk mechanisms in schizophrenia, and previously focused on the role of genes. But although a dozen or so genes have been linked to the disorder, "even the most powerful of these genes conveys only a 20% increased risk", he says. Yet schizophrenia is twice as common in those who are city-born and raised as in those from the countryside, and the bigger the city, the higher the risk (see 'Dose response?').

So Meyer-Lindenberg set out to study how city life might increase the risk of mental illness. The team scanned the brains of 32 student volunteers while they performed arithmetic tests. At the same time, the students received negative feedback through headphones. "We'd tell individuals they were performing below average, and suggest impatiently they hurry up a bit, so they'd feel they were failing," he explains.

This 'social stress' activated many brain areas, two of them specifically correlated with the volunteers' history of urban living. The amygdala, which processes emotion, was activated only in people currently living in a city. And the cingulate cortex, which helps to regulate the amygdala and processes negative emotions, responded more strongly in those brought up in cities than in those who grew up in towns or rural areas.

The initial experiment showed such clear associations that Meyer-Lindenberg didn't think anyone would believe them. So he did a similar experiment on another 23 subjects, this time adding visual feedback that allowed participants to see the investigators' frowns. He found the same sturdy associations.

He now plans to repeat the work in the general population, where urban–rural differences are likely to be even stronger than in students. He also plans to study how other risk factors identified by social scientists — such as being an immigrant — affect stress processing. "We will use tools from social scientists to help us quantify things like perceived discrimination, social support networks, or stigma," he says.

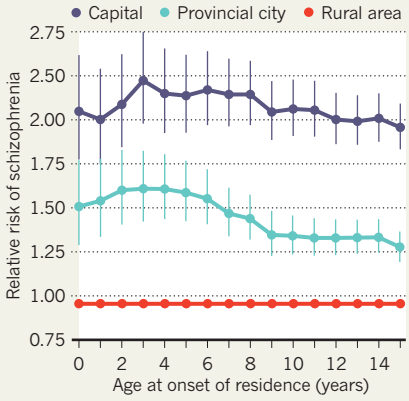
Yet he has had trouble interesting his social-science colleagues in setting up joint projects. Such lack of sympathy across the cultural divide is common, says Ernst Fehr, an economist at the University of Zurich, Switzerland, and a pioneer in the field of neuroeconomics, which studies the neurological basis of economic decisions. "But social problems have neurobiological effects, which, in turn, may exacerbate the social problems," he says.

The social sciences have as much to gain from crossing disciplinary boundaries as the biological sciences, says Morgan. "Sociologists and epidemiologists establish associations that are plausible — like immigrants may suffer more mental illness because of social isolation — but they are validated when neuroscientists demonstrate a robust biological mechanism."

For his future investigations, Meyer-Lindenberg is seeking urban planners who can help him to tease out how variables such as green space and population density contribute to the neurobiological impact of city living. Hans Wirz of the urban planning office in Basel, Switzerland, says that it took decades to integrate knowledge about the biomedical effects of the cityscape into his profession. "But when it comes to mental health we haven't a clue." ■

DOSE RESPONSE?

A 2001 study in Denmark showed that people who grew up in larger cities had a higher risk of schizophrenia.



SOURCE: C. B. PEDERSEN & P. B. MORTENSEN *ARCH. GEN. PSYCHIATRY* **58,** 1039–1046 (2001)

➲ **NATURE.COM**
For more on the environment and the brain, see:
go.nature.com/i6d1ch

E. CASTRO-MENDIVIL/REUTERS

Indigenous communities turned out in Lima to protest the proposed Inambari Dam in October 2010.

ENERGY POLICY

Worth a dam?

Voluntary agreement enables rating of hydroelectric impacts.

BY JEFF TOLLEFSON

Hydropower is booming in the developing world, but one megaproject faltered last week. On 13 June, after years of community protests, Peru announced that it was revoking an agreement with a Brazilian consortium to build the 2,000-megawatt Inambari Dam, which would have flooded 400 square kilometres of Amazonian forest.

Now, to foster a less confrontational way of advancing projects, the hydropower industry, environmental and human-rights organizations, and representatives from banks and governments have negotiated a mechanism for evaluating, and perhaps mitigating, the impact of dams before they are built.

Released on 16 June in Iguaçu Falls, Brazil, the Hydropower Sustainability Assessment Protocol provides a method for assessing dams in all phases, from development to operation. Projects would be ranked on a scale of one to five according to their likely effects on biodiversity, ecology, hydrology and erosion as well as on broader issues regarding regional planning, cultural heritage and effect on local inhabitants.

The protocol is voluntary, and a poor rating may not prevent a project from going ahead. Yet quantifying anticipated effects could generate pressure for managers to rethink plans to improve the outcome. "If we have some good results in a few test cases around the world, I think it will take off," says Pedro Bara, who works for WWF, one of the environmental groups that helped to develop the protocol. "It's very useful to have an international standard, especially for countries that don't have much experience in hydropower development."

The document has its roots in the World Commission on Dams, which produced comprehensive international guidelines in 2000. The International Hydropower Association (IHA), a trade group based in London, followed up with its own sustainability protocol in 2006, but continued criticism that the protocol was weak led the IHA to establish a formal dialogue with environmentalists and human-rights groups in 2008. Since then, the process has brought groups such as the WWF, Oxfam and Transparency International together with industry officials to hammer out a compromise.

Advocacy groups, governments and companies can use the protocol informally, but a governance council representing the stakeholder groups will oversee a more formal assessment carried out by trained auditors. Companies would pay for this assessment and would be required to publicly release the results.

How the protocol will be applied remains to be seen. Companies that sign it are not required to use it, or to alter their projects if the assessment identifies problems. This has split environmental groups, many of which called the protocol a dangerous public-relations tool that will allow companies to 'greenwash' their projects and will weaken existing standards.

The protocol is designed to be applied one dam at a time, missing cumulative impacts of development as well as opportunities to identify the best sites and coordinate energy production across an entire river system, such as the Amazon. "Where you have a cascade of dams, you really need a broader assessment," says Mathias Kondolf, a fluvial geomorphologist at the University of California, Berkeley.

Nonetheless, advocates hope that governments will require formal evaluations under the protocol and apply minimum standards to the projects that they support. "The aspirational end point would be some kind of independent certification system," says David Harrison, a senior adviser at the Nature Conservancy based in Salem, Oregon, who worked on the protocol.

Within industry, many see the protocol as a way to head off the kind of public opposition that can stall or quash a project after years of investment, says Cameron Ironside, a programme director at the IHA. "Everybody would be very happy if industry were able to apply the protocol early on," he says, "because that would solve a lot of the issues that people are running into downstream."

At least 140 major companies have signed on, including representatives of utility companies such as the China Three Gorges Dam Corporation in Yichang, Paris-based EDF and Eletrobras in Rio de Janeiro, Brazil. WWF officials have noted that many Chinese companies that are building dams around the world are not yet parties to the protocol. Yet companies on the sidelines may ultimately have to act in accordance with the protocol, says Bara, if it becomes standard practice. "If the banks get involved, it will be difficult not to jump in." ■

SEE EDITORIAL P.420

MORE ONLINE

Q&A

Tachi Yamada moves on from the Bill & Melinda Gates Foundation go.nature.com/mmyxoi

BILL & MELINDA GATES FOUNDATION

NEWS

- Emperor Hadrian's country estate is aligned to meet the solstices go.nature.com/jxbxo3
- Chemical in urine alerts prey species to a nearby predator go.nature.com/4udkoh
- Genome study solves twins' mystery condition go.nature.com/mfetk3

COMMUNICATION

Researchers tweet technical talk

In the latest trend in scientific discourse, journal clubs and data disclosures move to Twitter.

BY EUGENIE SAMUEL REICH

Robert Boyle would not have approved. The famously verbose seventeenth-century natural philosopher and pioneer of the scientific method argued that technical communication demands detail, not brevity. But there is no room for Boyle-like excess within the confines of the social-networking site Twitter, where users converse in 140-character posts, or 'tweets'. Researchers are now using the site's abbreviated messages to discuss papers in journal clubs and to share data in real time.

Scientists were among the early adopters when Twitter was launched five years ago, and many now use the site to tweet updates from conferences, promote links to papers, network and gossip. But using Twitter to engage in technical discussions with remote colleagues "is a novel thing", says Cornelius Puschmann, a linguist and information scientist at the University of Düsseldorf in Germany, who has tracked scientists' use of social media. Proponents say that Twitter enables fast-moving conversations, and allows users to post links pointing to more detail where required.

A medical-journal club started by Fiona Douglas, a medical student at the University of Cambridge, UK, and Natalie Silvey, a doctor at University Hospitals Coventry and Warwickshire in Coventry, UK, has held three weekly meetings since it began this month. A group for astronomers met for the first time last week.

In the inaugral session of the medical club on 5 June, around 60 participants discussed an influential paper on the treatment of acute sepsis and septic shock (E. Rivers *et al. N. Engl. J. Med.* **345,** 1368–1377; 2001). Silvey says that the participants encompassed a wider range of expertise and career stages than would be seen at a conventional journal club, and that the technical quality of the discussion was high. Some participants used data in the paper to calculate the 'number needed to treat' (NNT), a measure of the effectiveness of an intervention based on the average number of patients who have to be treated for one to benefit (see 'Medical chatter').

The format confounded some potential participants. Ves Dimov, an allergy doctor at the University of Chicago in Illinois, found the conversation hard to follow. "Only people who can wrap their minds around this somewhat chaotic environment will find Twitter journal clubs useful," he says.

On 11 June, Mike Brown, an astronomer at the California Institute of Technology in Pasadena, pioneered another use of Twitter when he 'live-tweeted' plots of observations he was making of the transit of the dwarf planet Haumea by its moon, Namaka, from the William Herschel Telescope on La Palma in the Canary Islands, Spain (see go.nature.com/me9trl). "There is an interest in getting discoveries out there quickly," he says.

Simon Schaffer, a historian of science at the University of Cambridge, says that Twitter is not as different from traditional scientific communication as it may seem. Journals began as letters between scientists, and were only later collated and published. "The idea of the person-to-person message is the essence of communication in the sciences," says Schaffer. And although Boyle's published reports were long-winded, he and other early scientists also kept notebooks in which they succinctly listed facts and observations that they had heard from others in coffee shops and elsewhere. "In a sense, Boyle was already tweeting," says Schaffer. ■

DISCUSSION DATA

Medical chatter

On 5 June, a medical-journal club had its first meeting on Twitter. The chat was brisk and technical, as this extract shows. NNT measures a treatment's effectiveness; #TwitJC is the hashtag that identifies posts as part of the club. The full transcript is at go.nature.com/iuidmy.

tobyhillman RT @TWSY: NNT to save one life? #twitjc -9:04 PM

drgrumble @amcunningham Not if we are still arguing about whether or not the result is meaningful. #TwitJC -9:04 PM

simonpeteryoung @drgrumble @twitjournalclub ah but then the non-sig results are left vulnerable #TwitJC -9:04 PM

aj0610 @anaestheticdoc @TWSY About 6 #twitjc -9:04 PM

AP/J. WILCOX, BOSTON HERALD

Massachusetts governor Deval Patrick (centre) listens with Boston mayor Thomas Menino as Pfizer's Sally Susman announces an expansion of the firm's research operations in Boston.

PHARMACEUTICALS

Drug buddies

The pharmaceutical industry is seeking stronger ties with academia in a bid to speed up drug development.

BY HEIDI LEDFORD

When pharmaceutical company Pfizer announced on 8 June that it is teaming up with eight research institutions in the Boston area to hunt for candidate drugs, the news was cheered from all sides. The governor of Massachusetts, Deval Patrick, praised the US$100-million, five-year deal for the jobs it would bring to the region. Eric Buehrens, interim chief executive of Beth Israel Deaconess Medical Center, one of the academic partners, extolled the benefits to scientific research. Pfizer says that the partnerships will accelerate development of the next generation of drugs.

The agreement is the latest sign of a growing trend in the pharmaceutical industry, which is trying to cut costs and improve efficiency by outsourcing the earliest phases of drug discovery. "All the drug companies are looking for a new model," says Mark Pepys, a professor of medicine at University College London who is collaborating with London-based GlaxoSmithKline (GSK). For academics facing tight research budgets, the deals bring financial benefits — and potentially fraught relationships with the companies and their academic peers.

"My big fear is that we're going to create a polarization within academic centres," says Kenneth Kaitin, director of the Tufts Center for the Study of Drug Development in Boston. "There will be those that partner with industry and in some cases will be looked on more negatively by their academic peers, and those that would never partner with industry because they feel that betrays their academic purity."

Pfizer's partnership is the third agreement it has brokered with institutions since November under a new programme that includes deals with a group of research centres in New York, and an $85-million collaboration with the University of California, San Francisco. Other major pharmaceutical firms, including GSK and AstraZeneca, have also been busy snatching up academic partners (see Table). In 2006, AstraZeneca initiated 271 interactions, including collaborations and other agreements, with academia. By 2010, that number had more than doubled to 594.

NATURE.COM
Follow the changing landscape of drug discovery at:
go.nature.com/ekv6qw

Discoveries made in academic labs have long fuelled industrial drug development. Greg Wiederrecht, vice-president for external scientific affairs at Merck, readily ticks off the company's drugs that originated in this way. The development of Gardasil, a vaccine against human papillomavirus, began in a lab at the University of Queensland in Australia, for example. RotaTeq, a rotavirus vaccine, was developed with technology from the University of Pennsylvania and the Wistar Institute in Philadelphia. But these relationships are becoming more important as the industry closes its research labs in response to falling profits (see *Nature* **470,** 154; 2011).

Tightening federal budgets are also putting financial strains on academic labs, making industry collaborations more attractive. "Every academic centre is looking at all this money flying around and asking: 'how do we get a piece of this action?'" says Kaitin.

Yet industry's need for secrecy, and its tendency to change its research focus abruptly, can conflict with the more open and stable academic environment. Pepys experienced this first-hand in the 1990s when the Swiss drugmaker Roche abruptly terminated its collaboration with his team. He faced a long and costly battle to retrieve the intellectual-property rights to a compound developed during the project. A little later, Roche agreed to work on the compound with him again, only to prematurely end the collaboration a second time. Once more, Pepys had to fight to continue his work. "It was a very expensive and tedious process that has delayed the drug by about ten years," he says. "And the clock on the patent is ticking."

Nevertheless, Pepys notes that without Roche's help, he would not have been able to develop a compound that, he hopes, will soon be ready for clinical tests in people with Alzheimer's disease. "Nobody except big pharma can make a medicine effectively," he says.

Industry, too, has had its share of frustrations. Shiv Krishnan, a senior director at French drugmaker Sanofi's branch in Bridgewater, New Jersey, notes that Hoechst, a German life-sciences company that Sanofi acquired, paid $70 million in the 1980s to fund research at Massachusetts General Hospital. In the end, however, the firm had little to show for it, Krishnan says. "And why?" he asks. "Because it wasn't collaborative. It was: I'll send you the cheque and you let me know when you have something."

In recent years, industry has taken a more focused and collaborative role in academic research. In its latest agreement, Pfizer says it will set up a lab in Boston that will house about 50 researchers — half of them Pfizer employees, the rest Pfizer-funded postdocs from participating academic labs. The team will work on projects selected by an oversight board comprised of academics and Pfizer executives. The company hopes to develop up to 30 projects from the three agreements, says Anthony Coyle, who is directing the programmes for Pfizer.

Pfizer's programmes are unique in the size of their financial commitments, but not in ▶

HIGHLIGHTS FROM INDUSTRY–ACADEMIA COLLABORATIONS ANNOUNCED THIS YEAR

Company	Academic institutions	Therapeutic area
Pfizer	Seven New York medical institutes	Biological drugs
Sanofi	University of California, San Francisco	Ageing, diabetes and inflammation
Gilead	Yale University, Connecticut	Cancer
GlaxoSmithKline and AstraZeneca	University of Manchester, UK	Inflammation
AstraZeneca	National Institute of Health and Medical Research, France	Cancer, inflammation and respiratory and autoimmune diseases
Takeda	Kyoto University, Japan	Obesity and schizophrenia

their pursuit of active collaboration: earlier this year, Sanofi announced similar agreements. Pepys, meanwhile, is working directly with GSK scientists to develop a drug against amyloidosis, a disease caused by a build-up of amyloid protein.

The various deals also aim to smooth over tensions between industry and academia. Duncan Holmes, who heads GSK's Discovery Partnerships with Academia initiative, says that the company will give research partners a year's notice if it chooses to end a collaboration and that, if it that happens, academics would be free to continue with the project. To ease worries about publication restrictions, many agreements stipulate the terms for publication ahead of time. Yet some academics wonder whether the trend towards industry collaboration will harm academic credibility. Some hospitals and universities, including Harvard University, have cracked down on industry relationships after it emerged over the past few years that researchers had received consulting and speaking fees from companies with a vested interest in their research.

Research agreements do not generally raise the same conflict-of-interest alarms as speaking fees, for example, which can be seen as marketing a product for a company, says Eric Campbell, a sociologist at Harvard Medical School in Boston. Also, many institutions vet the language in the contracts, he notes, and industry money is deposited into institutional accounts rather than given directly to investigators.

But Campbell also notes that industry collaborations can restrict or delay publication and lead to a publication bias in favour of a company's product. "You should not in any way accept the notion that these giant institutional agreements are without tremendous danger," he says.

Furthermore, academia's growing appetite for industry funds could tip the balance of power at the negotiating table. "The pressures on a university president are intense," says Howard Brody, a bioethicist at the University of Texas Medical Branch in Galveston. He advocates the creation of an external organization to oversee large-scale collaborations with the drug industry. "We have to remember that institutions have conflicts of interest, just like individuals do," he cautions.

Kaitin, however, says that industry is just as desperate to collaborate, if not more so. "Earnings at these companies are falling through the floor and investors are losing confidence," he says. ■

CORRECTION

The News story 'Egypt invests in science' (*Nature* **474,** 266; 2011) wrongly identified Susan Hockfield as a member of the board of trustees appointed to run a proposed non-profit science city near Cairo. She is not a member of this board.

WORD

PLAY

By mining a database of the world's books, Erez Lieberman Aiden is attempting to automate much of humanities research. But is the field ready to be digitized?

BY ERIC HAND

Erez Lieberman Aiden is standing on the sun deck of his town house, rocking back and forth on the balls of his bare feet as he belts out a blessing. The Hebrew words echo across the quiet courtyards of Harvard University in Cambridge, Massachusetts. The sky has turned indigo as the light and warmth leak away from this day in late April. *Shalom aleichem*, he sings. Peace be upon you.

Lieberman Aiden — molecular biologist, applied mathematician and, at 31 years old, the precocious doyen of the emerging field known as the digital humanities — could do with a little peace. The cries of his 10-month-old son have abated — for the moment — and he has had just enough time to throw on a pair of frayed black trousers and a shiny synthetic pullover before his guests arrive. A five o'clock shadow darkens the terrain between his thick goatee and unkempt hair. The night before, he caught a late train back from Princeton University in New Jersey, where he, the geeky scientist, had the delicate task of informing a room of erudite historians that his efforts at mining a database of 5 million books, about 4% of all those ever published, had made much of what they do trivially easy. The scrupulous tracking of ideas across history, for instance — work that has consumed entire careers — can be done in seconds with tools that Lieberman Aiden and his colleagues have invented.

SAM OGDEN

Yet his role as evangelist for change in the humanities — or doomsday prophet, depending on your point of view — is just one of the many parts played by Lieberman Aiden. He is also: the inventor of a groundbreaking protocol that reveals how DNA can be tightly wound and yet untangled enough to orchestrate life; the chief executive of iShoe, a company that is testing sensor-stuffed shoe inserts to help the elderly with their balance; and the co-founder, with his wife, of Bears Without Borders, which sends thousands of stuffed animals to children in the developing world. (Barely concealed in the couple's basement are mounds of donated animals awaiting delivery.) In pouring his energies into all the projects that excite him, Lieberman Aiden doesn't transcend disciplinary boundaries so much as ignore them. And although he is still technically a postdoctoral researcher at Harvard, Lieberman Aiden seems to publish the results of those projects almost exclusively on the covers of *Science* and *Nature*; hung in the stairwell below the sun deck, he has framed blow-ups of the magazine covers to prove it.

But that is work, and this is Shabbat dinner, the start of the Jewish Sabbath: a time for rest. The light switches will remain untouched, leaving the house illuminated through the night; the hot plate in the kitchen, on which the meal is being warmed, is on a timer. Three candles have been lit, one for each member of the household. Lieberman Aiden sings unabashedly in a hearty baritone that is not at all like his reedy, excitable speaking voice. He gazes at his wife, Aviva Presser Aiden, who grins back at him, holding her sweater tight to herself in the chilly night air. She too has reason to rest contentedly. The week before, she learned that she had won a US$100,000 grant from the Bill & Melinda Gates Foundation in Seattle, Washington, to build a microbial fuel cell that could charge mobile phones in Africa. The project means a year-long break from her studies at Harvard Medical School in Boston, where she is adding an MD to her PhD in genetics.

It is only by comparison with this academic power-couple that the other dinner guests — two young, self-assured Harvard physics graduates — look a bit lost, but

that probably has more to do with their unfamiliarity with the Shabbat rituals. They flip through the Hebrew prayer books and try to follow along. But Lieberman Aiden, who in his 20s toyed with becoming a rabbi, has no need for the book. These are the texts that he has studied for years. These are the words he knows best.

READING VERY NOT-CAREFULLY

As a reader with a finite amount of time, Lieberman Aiden likes to say, you pretty much have two choices. You can read a small number of books very carefully. Or you can read lots of books "very, very not-carefully". Most humanities scholars abide by the former approach. In a process known as close-reading, they seek out original sources in archives, where they underline, annotate and cross-reference the text in efforts to identify and interpret authors' intentions, historical trends and linguistic evolution. It's the approach Lieberman Aiden followed for a 2007 paper in *Nature*[1]. Sifting through old grammar books, he and his colleagues identified 177 verbs that were irregular in the era of Old English (around AD 800) and studied their conjugation in Middle English (around AD 1200), then in the English used today. They found that less-commonly used verbs regularized much more quickly than commonly used ones: 'wrought' became 'worked', but 'went' has not become 'goed'. The study gave Lieberman Aiden a first-hand lesson in how painstaking a traditional humanities approach could be.

But what if, Lieberman Aiden wondered, you could read every book ever written 'not-carefully'? You could then show how verbs are conjugated not just at isolated moments in history, but continuously through time, as the culture evolves. Studies could take in more data, faster. As he began thinking about this question, Lieberman Aiden realized that 'reading' books in this way was precisely the ambition of the Google Books project, a digitization of some 18 million books, most of them published since 1800. In 2007, he 'cold e-mailed' members of the Google Books team, and was surprised to get a face-to-face meeting with Peter Norvig, Google's director of research, just over a week later. "It went well," Lieberman Aiden says, in an understatement.

Working with Google and his chief collaborator, 29-year-old Harvard psychology postdoc Jean-Baptiste Michel, Lieberman Aiden built a software tool called the *n*-grams viewer to chart the frequency of phrases across a corpus of 500 billion words. A 'one-gram' plots the frequency of a single word such as 'feminism' over time; a 'two-gram' shows the frequency of a contiguous phrase, such as 'touch base' (see 'Think outside the box').

Google unveiled the tool on 16 December 2010, the same day that Lieberman Aiden and his colleagues published a paper in *Science*[2] describing how the tool could be used, for example, to identify the verb that has regularized the fastest: 'chid' and 'chode' to 'chided' in some 200 years (see 'The fastest verb on the planet'). "We found 'found' 200,000 times more often than we finded 'finded'," they wrote, with characteristic playfulness. "In contrast, 'dwelt' dwelt in our data only 60 times as often as 'dwelled' dwelled." Interspersed between the jokes were real discoveries — many of which had nothing to do with verbs. By comparing German and English texts from the first half of the twentieth century, the team showed that the Nazi regime suppressed mention of the Jewish artist Marc Chagall, and that the *n*-grams tool could be used to identify artists, writers or activists whose suppression had hitherto been unknown. Lieberman Aiden and Michel called their approach culturomics, a reference to the genomics-like scale of the book database, and a nod to the future, when they hope that more of the media that underpin culture — newspapers, blogs, art, music — will be folded in.

In the first 24 hours after its launch, the *n*-grams viewer (ngrams.googlelabs.com) received more than one million hits. Dan Cohen, director of the Roy Rosenzweig Center for History and New Media at George Mason University in Fairfax, Virginia, calls the tool a "gateway drug" for the digital humanities, a field that has been gaining pace and funding in the past few years (see 'A discipline goes digital'). The name is an umbrella term for approaches that include not just the assembly of large-scale databases of media and other cultural data, but also the willingness of humanities scholars to develop the algorithms to engage with them. "These tools are revolutionizing the way we work and the kinds of arguments we can make," says Dan Edelstein, a historian at Stanford University in California, who has used mapping software to show unexpected patterns in the way that Voltaire's letters spread through Europe during the Enlightenment.

"THESE TOOLS ARE REVOLUTIONIZING THE WAY WE WORK."

Yet some humanities researchers in the traditional camp complain that their field can never be encapsulated by the frequency charts of words and phrases produced by the *n*-grams tool. "I think saying all books equal the DNA of human experience — I think that's a very dangerous parallel," says Cohen. How do you factor in the cultural contributions of furniture, or dance, or ticket stubs at a movie hall, he asks. What about all the books that were never published? Or the culture as experienced by the world's vast illiterate populations?

Other scholars have deep reservations about the digital humanities movement as a whole — especially if it will come at the expense of traditional approaches. "You can't help but worry that this is going to sweep the deck of all money for humanities everywhere else," says Anthony Grafton, a historian at Princeton and president of the American Historical Association, who uses a giant, geared wooden reading wheel to help him manage his oversized, Renaissance texts. He wants researchers to hold onto the power that comes with intimately knowing their primary sources, right down to the scribbled notes in the margin that would elude the book scanners. "You don't want to give up what is your own core activity," he says.

FOLLOWING TRADITION

Back at the Aiden house, the Shabbat dinner guests have all laved their hands with a glass of water and returned to the sun deck for matzo-ball soup. Lieberman Aiden explains some of the trepidation he felt when he and Michel talked to the historians at Princeton about their work. "I was a little bit nervous going in," he says. "I really thought that we were going to get denounced at one point."

Although Lieberman Aiden and Michel are sensitive to the feelings of traditional humanities scholars, they are also too young, restless and deeply ambitious to slow their own pursuits. Lieberman Aiden says that the influence of technology on the humanities is already past a tipping point. The tools and methods that it provides, he says, will be impossible for researchers to ignore. And yet he doesn't think that the old approaches will ever disappear. "I think you should use the best methods available — and all of them," he says. "And I think that includes carefully reading texts and trying to get behind what authors think."

Daniel Koll, one of the dinner guests, shyly interrupts. "Erez? Do you think you're maybe partly influenced in that kind of thinking by your religious upbringing? From my limited outsider's perspective, Judaism has a very strong interpretive component. There is no single authority on a text, and so on." He wonders whether Lieberman Aiden, like any good humanities scholar, enjoys wrestling with the ambiguities of religious texts as much as he enjoys cool, hard data.

Clearly, the answer is yes — why else would his host have spent a year of his life at Yeshiva University in New York, studying the Talmud and Jewish case law? But Lieberman Aiden, who prefers to talk about other people and their

A DISCIPLINE GOES DIGITAL

The humanities mine cultural databases

The digital humanities — the use of algorithms to search for meaning in databases of text and other media — have been around for decades. Some trace the field's origins to Roberto Busa, an Italian priest who, in the late 1940s, teamed up with IBM to produce a searchable index of the works of thirteenth-century theologian Thomas Aquinas.

But the field has taken on new life in recent years. Journals have sprouted up and professional societies are blooming. Some universities are now requiring graduate students in the humanities to take statistics and computer-science courses. Funding — far harder to come by in the humanities than in the sciences — flows slightly more generously to those willing to adopt the new methods. This year, the US National Endowment for the Humanities, in collaboration with the National Science Foundation and research institutions in Canada and Britain, plans to hand out 20 grants in the digital humanities, worth a total of US$6 million.

Many researchers in the digital humanities use textual databases composed primarily of books — as Erez Lieberman Aiden does in his 'culturomics' project (see 'Heavy-duty data'). Franco Moretti, a literary scholar at Stanford University in California, has shown that genres of fiction — Gothic novels, for example, or romance — have a textual 'fingerprint' that is apparent even in simple frequency counts of nouns, verbs and prepositions. "These genres are different at every scale," he says, "not only in the huge scene of being held captive by a Count."

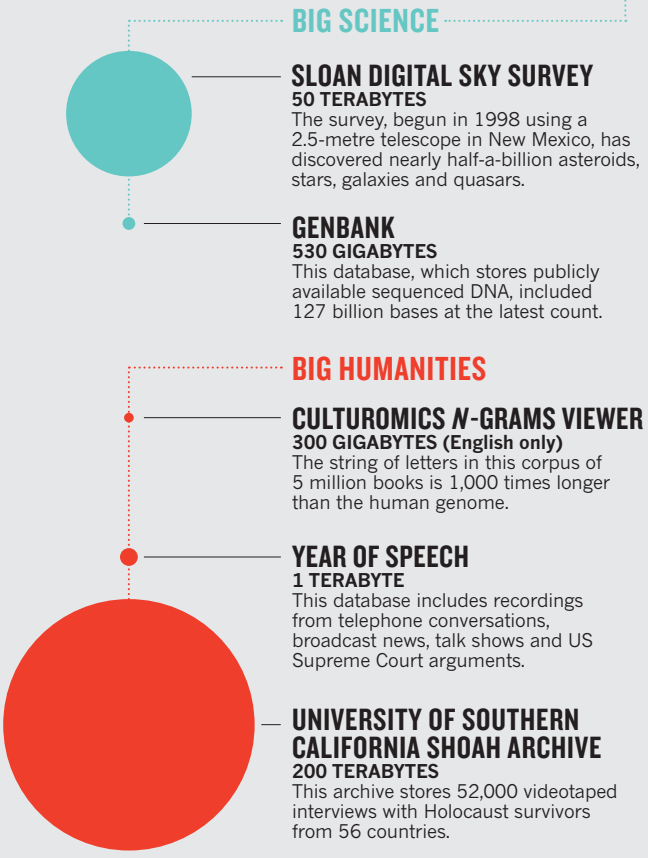
Some researchers are busy digitizing other forms of cultural data. John Coleman, a phonetician at the University of Oxford, UK, is putting 5 million spoken words — about 3 months of speech, end to end — into a database, down to the level of the individual phonemes. Collected largely as recordings made with Sony Walkmen in the 1990s, it contains all sorts of things typically ignored by linguists: neologisms, slurring and sub-verbal honks and snorts. Coleman is already learning how conversation partners take pacing cues from each other, and how pitch of voice reflects attitude. And, he says, he can prove that women and men talk at the same speed. The linguistics textbooks, he says, "are going to have to be rewritten".

Ichiro Fujinaga, a music technologist at McGill University in Montreal, Canada, is trying to do something similar for music. In a project known as SALAMI (Structural Analysis of Large Amounts of Music Information), Fujinaga is finding the common structural patterns (such as verse–chorus) in 350,000 pieces of music from all over the world. With more than 7,000 hours of Grateful Dead recordings in the database, he says, his team will be able to answer the all-important question: "Did the guitar solos get extended over the years or did they get shorter?" **E.H.**

LARGE HADRON COLLIDER
13 PETABYTES (2010)
The proton collider, near Geneva, Switzerland, generates about 15 petabytes of data per year — even after rejecting 99.9995% of collisions.

HEAVY-DUTY DATA

The computer-storage space required to support projects in the digital humanities is now starting to rival that of big-science projects.

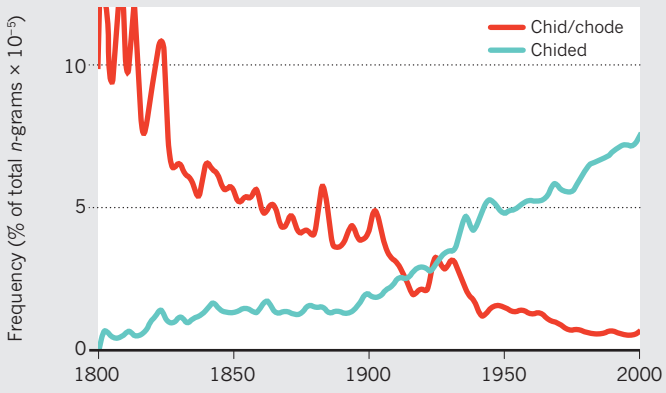


1 petabyte = 1,024 terabytes = 1,048,576 gigabytes

SOURCE: REF. 2

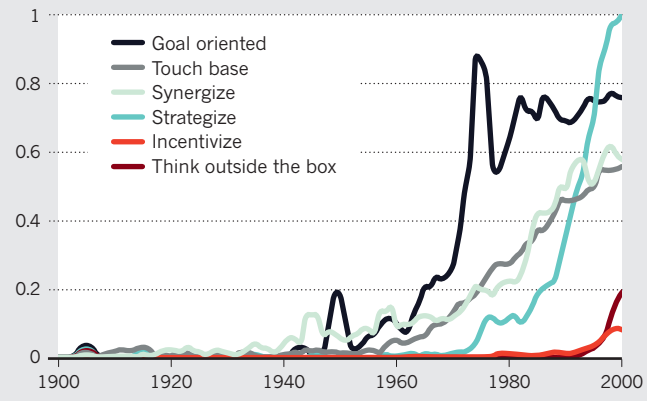
THE FASTEST VERB ON THE PLANET

Rarely used verbs regularize quickly; the *n*-grams viewer reveals that 'chide' has changed fastest of all.



THINK OUTSIDE THE BOX

Text analysis using the *n*-grams viewer shows the infiltration of corporate speak into the English language.



SOURCE: GOOGLE N-GRAMS VIEWER

ideas rather than himself, provides an indirect answer by way of history. He tells the story of Isaac Casaubon, a sixteenth-century Protestant scholar, who undermined the presumed Egyptian provenance of a set of religious texts by identifying a reference to a Greek play on words — something that could only have been written hundreds of years later. "That point is as objective an interpretive remark as any remark a scientist might make," says Lieberman Aiden. "So the methods of humanists are very, very formidable. And I think the degree of insecurity they have over whether these methods are here to stay is not really befitting."

TWO CULTURES

From the day he was born in a New York City hospital, Lieberman Aiden was steeped in the cultures of both language and technology. The son of a Hungarian mother and Romanian father, both émigrés by way of Israel, Lieberman Aiden grew up in a community of Satmar Jews, a branch of Hasidic Orthodox Judaism. English was his third language, after Hungarian and Hebrew, and by the age of nine, he was helping his father, a self-taught inventor, with the English contracts for the family saw-manufacturing business. Lieberman Aiden studied at a religious high school in Brooklyn, but soon found that video games held more allure. In his second year there, he found himself flunking classes, and his addiction to *X-COM: UFO Defense* was consuming so much time that he eventually had to quit, cold turkey. "It was an amazing game, actually," he says, ruefully.

Lieberman Aiden soon found more edifying outlets for his energies: he was allowed to skip school one day a week to study in a molecular-biology lab at Brooklyn College, and he began his own computer-repair business. The family was quite secular by Hasidic standards, going to synagogue only for the High Holidays of Rosh Hashana and Yom Kippur. One day in high school, he went to Burger King for his usual bacon cheeseburger, and decided to respect kosher rules by forgoing the bacon — not realizing that mixing dairy and beef in the cheeseburger itself was not kosher at all.

Lieberman Aiden went to Princeton as an undergraduate, where he wasn't content to study just maths and physics, but also fulfilled all the requirements for a philosophy degree. And even as he took five, six, seven classes a term, he managed to squeeze in creative-writing courses, specializing in haiku poetry. While at Yeshiva University after Princeton, he taught maths on the side to pay for a master's degree in history, and completed the first year of rabbinical studies. "He's a non-conformist by design and he revels in it," says his Talmudic study partner, Avi Bossewitch. And yet, he says, "he's the least arrogant person I've ever met".

The allure of science eventually proved too strong. Lieberman Aiden left Yeshiva to begin a PhD at the Broad Institute of the Massachusetts Institute of Technology (MIT) and Harvard in Cambridge, under the supervision of famed geneticist Eric Lander. But even while mastering molecular biology, he put his maths skills to use. He realized that a knot-free shape described in a 120-year-old maths paper — a fractal globule — could describe the way that the 2-metre-long human genome folds up into the cell nucleus, a space one million times smaller. He then developed a protocol to prove that this was true. The results, published in the first of his papers to make it onto the cover of *Science*[3], showed that the fractal globule allowed widely separated sections of DNA to unfold and interact. "There are no bounds to what he can be interested in," says Lander, who, like others, suspects that culturomics might eventually be merely a sideline for Lieberman Aiden in his ascent in mathematical biology.

> "HE'S A NON-CONFORMIST BY DESIGN AND HE REVELS IN IT."

It was during his time in the Lander lab that he met Aviva Presser, a shy young woman from Los Angeles who was also working towards her PhD. They married in 2005, with Lander giving one of the blessings. Rather than one of them taking the other's name, they decided to share a common new name: Aiden, which means Eden in Hebrew and, in Gaelic, little fire.

Those fires stoked, they were faced with another naming issue last June: what to call their son. The working title, *in utero*, was Snedley Balagan (Balagan means 'fiasco' in Hebrew). But they soon settled on Gabriel Galileo Aiden. Presser Aiden says that no one believes them when they say that they had no idea that their son's initials formed the DNA code for a sweet amino acid.

WORK AND PLAY

Towards the end of the Shabbat dinner, Gabriel decides to wake up, adding to Presser Aiden's obvious exhaustion. But her husband doesn't want to miss out on the highlight of the night, an Aiden-family staple: Dessert Face-off, in which each guest competitively designs a slice of brownie within a theme. Given that guests Koll and his girlfriend, Larissa Zhou, are interested in molecular gastronomy, Lieberman Aiden decides on the theme of food science. A box of edible Betty Crocker decorations is laid out.

Koll turns his brownie into the cross-section of a wok. Zhou transforms hers into a pig — definitely not kosher. But Lieberman Aiden's construct is perplexing: he has created a starry night-time skyscape, with multicoloured sprinkles as stars and the Milky Way. What does this have to do with food science? "Well, you know," he says, a little proud of himself, "gastronomy is just one letter away from astronomy."

Gabriel has returned to bed, and Presser Aiden looks ready to follow, but her husband isn't quite finished yet. As midnight approaches, Lieberman Aiden is holding forth on the mathematical beauty of ramen noodles — as depicted in the 2008 film *Kung Fu Panda* — and remarking on a piece of maths software, KnotPlot, that could help his wife to braid some really radical challah bread. The guests finally do Presser Aiden a favour and excuse themselves to go home.

But Lieberman Aiden just might keep revving and riffing through the night: he can work for 70–80 hours straight, fuelled by Diet Coke and junk food. He has big plans for culturomics, as he and Michel add more languages, books and other media to the *n*-grams database. He also has new projects to think over, such as one currently under way with Ed Boyden, a young, acclaimed neurobiologist at MIT, in which the pair are developing a way to detect the genes expressed in thousands of individual cells at a time. But tonight and tomorrow he'll keep his computer off, although it is not religious conviction that makes him abide by the Sabbath rules. Doing so forces him to detach, clear his mind and go for walks in the park with his wife and son.

And yet the boundary between work and play — just like that between the sciences and the humanities — is not one that Lieberman Aiden respects. That might just be what makes him successful, says Lander. For centuries, the best science has come from the most playful scientists, he says. Think of Watson and Crick shirking the lab in favour of tennis; think of Einstein and his wild-haired bike rides.

"What do children do?" says Lander. "They learn, they're curious, they're stimulated. The problem is, at some point, many people get in a rut. They're not really interested in learning more. They're not able to be fascinated and delighted by everything around them. Erez — he hasn't lost the playfulness." ■ SEE EDITORIAL P.420

Eric Hand *is a reporter for* Nature *in Washington DC.*

1. Lieberman, E., Michel, J.-B., Jackson, J., Tang, T. & Nowak, M. A. *Nature* **449,** 713–716 (2007).
2. Michel, J.-B. *et al. Science* **331,** 176–182 (2010).
3. Lieberman-Aiden, E. *et al. Science* **326,** 289–293 (2009).

COMMENT

CORBIS

H. NEW/REUTERS

The wave that hit Miyako City on Japan's east coast during the 11 March tsunami caught researchers by surprise.

Hidden depths

A staggering lack of undersea data hampers our understanding of earthquakes and tsunamis. Geophysicists must put more instruments offshore, says **Andrew V. Newman**.

The magnitude-9.0 Tohoku-Oki earthquake and tsunami that struck Japan on 11 March were devastating not just to the Japanese people, but also to the scientific community. Seismologists had underestimated the earthquake size by a factor of 4 or more, and the tsunami hazard by so much that the protective walls around the Fukushima nuclear power plant were overtopped.

The past 20 years have seen great strides in understanding earthquake faults and volcanoes, largely thanks to technological advances allowing a huge increase in ground-deformation measurements. Precise measurements of ground movement can provide nearly direct information about the strain energy accumulating along a fault, which can be released by a giant quake. The number of researchers interested in this field has skyrocketed.

But underwater monitoring lags behind. The tools for monitoring subtle ground movements rely largely on the Global Positioning System (GPS), and so only work when they have a clear view of the satellites they rely on. Their radio waves cannot penetrate water. This makes underwater monitoring expensive and time-consuming, so few opt for it. Instead, most of the community has been drawn to rely on cheaper, land-based tools for understanding earthquake behaviour, including that around underwater faults — where the real action happens offshore. Seismometers, which detect shaking during earthquakes, are routinely deployed on land and at sea. But ground-deformation monitors have been restricted almost entirely to land. Geophysicists are relying on the data they have rather than the data they need.

This needs to change. We must improve undersea monitoring and make it cheaper, increasing measurements of the sea floor 100-fold. By so doing, we will vastly improve both our understanding of plate-boundary dynamics and volcanic processes underwater, and our assessment of earthquake and tsunami hazards.

LOCKED AND LOADED

Earthquakes are the result of runaway release of stored strain energy from 'locked patches', where the tectonic plates get ▶



D. PARKER/SPL; EARTHSCOPE; J. WARK, VISUALS UNLIMITED/SPL

Land-based deformation monitors (left and middle) are crucial, but only help to understand above-ground stretches of faults like California's San Andreas.

stuck along faults. As the plates creep past each other, the ground deforms over these locked patches in a way that is easily modelled. So, by monitoring and quantifying regions undergoing deformation along tectonic-plate boundaries, researchers can determine the locations and sizes of locked patches. That in turn helps to determine the earthquake potential of a given fault.

Along the section of the San Andreas Fault in Parkfield, California, for example, scientists have taken measurements of ground deformation at a dozen GPS stations spaced 5–10 kilometres apart, some of which have been in place since the early 1990s. These stations mapped out the locked patch that failed in the 2004 magnitude-6.0 earthquake at Parkfield[1].

More than 90% of known tectonic-plate boundaries are underwater. Only rarely do large, coastal faults pop up above the water to places where we can more easily monitor them, as in the San Andreas Fault, the North Anatolian Fault in Turkey or numerous faults in central China. These locales are important because they yield detailed imaging opportunities. But they provide only glimpses of these faults' processes, and offer little solace to those studying the environments that create the world's largest earthquakes: subduction zones, where an oceanic plate slides beneath another plate, underwater and out of sight.

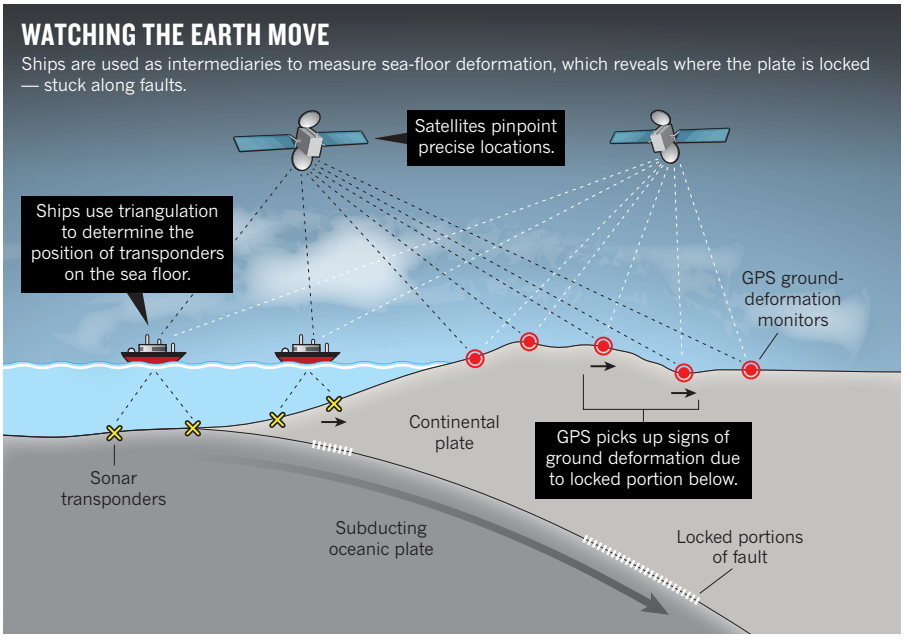
As with tectonic boundaries, many major tectonic plates are largely underwater, which makes tracking their motions difficult. The Cocos plate, for example — which collides into, and is forced under, the North American and Caribbean plates below southern Mexico and much of Central America — holds only a single island. This leads to significant uncertainty in our estimates of convergent plate velocities and direction — parameters important for assessing the long-term earthquake potential of the region.

Subduction-zone faults extend downwards from a trench on the sea floor, often hundreds of kilometres offshore (200 kilometres in the case of the Japan Trench). From here, the fault slices downwards at an angle towards and underneath land (see 'Watching the Earth move'). With land-based measurements alone, scientists can do a fair job of imaging the locking in the deeper portions, which typically lie 30–50 kilometres below the surface. This can be helpful for assessing the strain energy likely to be released by strong shaking on land. But we cannot assess the quake's tsunami potential from land measurements.

A growing body of research is now showing that large tsunamis are most easily generated by earthquake movement near the trench. In Sumatra last October, for example, a magnitude-7.8 earthquake very near the sea-floor trench created waves approximately 10 metres high, comparable to those from the quake in Chile in February that year, which was magnitude-8.8 but 20 kilometres deeper[2,3].

To get offshore ground-deformation data, scientists mostly use sonar transponders on the sea floor. Using GPS, a ship pinpoints its own location, and uses triangulation to determine the transponders' positions on the sea floor. Repeat measurements are made, years apart, to determine any changes in the transponders' locations. This works, but it is expensive. The development and deployment of a transponder may cost several hundreds of thousands of US dollars, and the use of a large ship for several days to make measurements would cost another hundred thousand. Depending on the ships used and the cost of fuel, a single determination of sea-floor deformation may cost half a million dollars.

This has kept the number of measurements to just a handful, globally. Studies off the coast of Japan before and after the Tohoku-Oki



N. JONES/A. V. NEWMAN

earthquake[4], and earlier off Peru and western North America[5,6], have shown deformation in a few spots along a trench using a combination of GPS and acoustic tools. In Japan, three points were observed to be moving landwards in the 10 years prior to the earthquake: a signal indicator of strain build-up. The region with the most such build-up suddenly jumped at least 24 metres seawards as the Tohoku-Oki quake ruptured along the fault[7] (see 'Lopsided measures'). With numerous similar measurements, particularly closer to the trench and before the event, it would have been possible to map out the shallow locked area, and better estimate both the earthquake and tsunami potential.

SCALING UP

Geophysicists should be working with government and intergovernmental agencies to develop and test cheaper technologies for tracking continuous, long-term sea-floor strain accumulation. Autonomous systems that can run without human intervention will be much cheaper in the long run. Real-time transmission of data would be possible by hooking in to existing systems of underwater cables, such as those used by NEPTUNE Canada, the scientific observatory offshore from Vancouver Island. Or it could be achieved through satellite communication with buoys on the ocean's surface, as is currently being developed at Scripps Institution of Oceanography in San Diego, California. It is not yet clear exactly what the cost savings of these methods might be.

There will likely be trade-offs between tools capable of making many accurate measurements in a matter of weeks, versus those more suited to making one measurement a year over many years. Some instruments need not based on GPS: devices can measure changes in tilt precisely, for example, rather than changes in absolute position.

It is vital that all subduction boundaries — particularly those with the capacity for causing massive damage to nearby cities — be densely measured routinely enough to capture the extent of locking. How many sensors are needed? For an underwater volcanic system, up to a dozen instruments might be sufficient to capture magmatic unrest. To identify the offshore locking that causes big earthquakes and tsunamis in subduction zone environments, stations should ideally be installed 20–40 kilometres apart, in a grid, from the shore to over the trench. For the Japan Trench, this would mean 100–400 new sensors on top of the handful there today. For larger-scale studies of global plate dynamics, stations distributed every 5,000–15,000 kilometres would be sufficient.

It should be possible to bring instrument manufacturing and deployment costs to under US$50,000 per station within a decade, if research and development is stepped up on available and emerging technologies. It would then cost $5 million to $20 million to equip an environment like the Japan Trench. That's far less than the current spend on national, land-based scientific-grade GPS infrastructure in countries such as the United States, Japan, and New Zealand. (The western US Plate Boundary Observatory, for example, cost $100 million to install, and needs about $10 million per year in upkeep.) Put another way, the figure is less than 0.01% of the projected costs of the Japanese tsunami, which may exceed $300 billion.

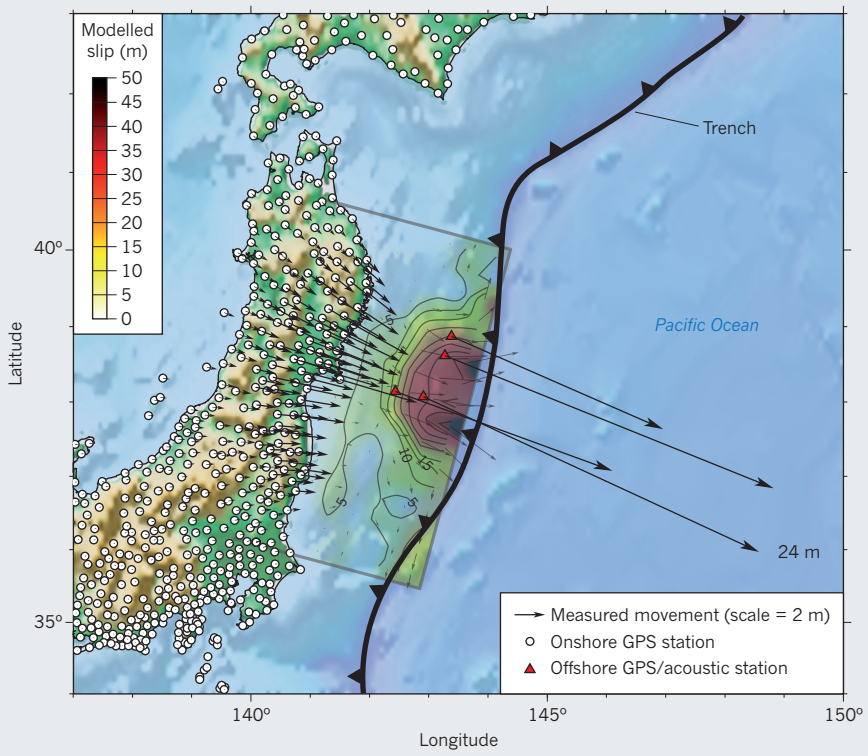
Geoscientists need to overcome the 'cognitive inertia' that is holding back progress in this field. Although the community is starting to recognize the importance of offshore measures, ideas and common practices are slow to change. Scientists and engineers should prod funding agencies for significantly more support towards developing new and cheaper tools.

Because earthquakes and tsunamis do not stop at political borders, it is essential that agencies such as the United Nations Educational, Scientific and Cultural Organization and the World Bank play a part. Intergovernmental agencies should catalyse the international research community, first through funding workshops and exploratory research, and then through the deployment and upkeep of extensive sea-floor measurement systems. Wealthier countries could offer support through programmes such as the United States Agency for International Development.

We live in something of a golden age for earthquake geophysicists: there has been a burst of large quakes to study with advanced research tools in the past 20 years. Data sharing is also getting better. Scientists worldwide had access to local seismic and geodetic data within a few days of the recent Japanese quake. Globally, the seismology community can examine real-time data from sensors almost anywhere on land only a few seconds after it is recorded. It is time for underwater monitoring to catch up. ■

LOPSIDED MEASURES

Most of the action during the 11 March 2011 tsunami-forming earthquake that hit Japan was offshore, but the vast majority of ground-deformation sensors are on land.



NEWMAN, A. V. *ET AL. SEISMO. RES. LETT.* **82** (2011).

Andrew V. Newman *is at the School of Earth and Atmospheric Sciences, Georgia Institute of Technology, 311 Ferst Drive, Atlanta 30332, USA.*
anewman@gatech.edu

1. Murray, J. & Langbein, J. *Bull. Seismo. Soc. Am.* **96,** S283–S303 (2006).
2. Newman, A. V., Hayes, G., Wei, Y. & Convers, J. A. *Geophys. Res. Lett.* **38,** L05302 (2011).
3. Lorito, S., Romano, F., Atzori, S. & Tong, X. *Nature Geosci.* **4,** 173–177 (2011).
4. Sato, M. *et al. Geophys. Res. Lett.* **38,** L01312 (2011).
5. Spiess, F. N. *et al. Physics Earth Planet. Inter.* **108,** 101–112 (1998).
6. Gagnon, K., Chadwell, C. D. & Norabuena, E. *Nature* **434,** 205–208 (2005).
7. Sato, M. T. *et al. Science* 10.1126/science.1207401 (2011).

NG HAN GUAN/AP

Presidents Barack Obama and Hu Jintao meeting in 2009. Space security was on the agenda.

US and China need contact, not cold war

Attempts to isolate the Chinese space community undermine US interests, says **Gregory Kulacki**.

This April, a US congressman used budget negotiations to ram through a potentially unconstitutional assault on the president's ability to conduct scientific diplomacy. A bill was passed stipulating that, until September 2011 at least, no appropriated funds may be used by NASA or the White House Office of Science and Technology Policy (OSTP) "to develop, design, plan, promulgate, implement, or execute a bilateral policy, program, order, or contract of any kind to participate, collaborate, or coordinate bilaterally in any way with China".

The move, instigated by Representative Frank Wolf (Republican, Virginia), chair of the commerce, justice and science subcommittee of the House appropriations committee, which funds NASA and the OSTP, is part of a decades-old congressional tradition of concerns about China's space programme. Nevertheless, this latest shot has wide-ranging implications. It has already led to the suspension of a geodynamics research project between the Chinese Academy of Sciences and NASA, for example. And it will impede ongoing bilateral negotiations on climate change and nuclear security that are part of the US–China Strategic and Economic Dialogue, a high-level forum established by presidents Barack Obama and Hu Jintao in 2009. My experience working in China for the Union of Concerned Scientists suggests that the ban will have a chilling effect on both government-funded scientific cooperation and on non-governmental activities.

The ban should be lifted. The progress of Chinese space activity during the previous US administration suggests that the prohibitions that have stifled Sino–American scientific cooperation for decades have not achieved their aims, and have arguably been counterproductive. China has shown that it has the talent and resources to go it alone. The sanctions have only severed links between the countries and made a new generation of Chinese intellectuals resentful and suspicious of the United States. And they stand in contrast to the tradition of scientists strengthening diplomatic relations.

There are signs that things could change. In 2009, Hu and Obama signed a joint statement to "enhance security in outer space" by "expanding discussions on space science cooperation". NASA administrator Charles Bolden visited Chinese launch facilities in October 2010, and presidential science adviser and OSTP director John Holdren has travelled to China three times in the past two years to discuss US–Chinese scientific and technical cooperation in many areas, including space science and exploration. Although no concrete programmes emerged from these travels, a second joint statement was signed in January 2011, committing the two countries to "deepen dialogue and exchanges in the field of space" and to "continue discussions on opportunities for practical future cooperation in the space arena, based on principles of transparency, reciprocity, and mutual benefit".

Wolf's bill, however, aims to prevent the Obama administration from implementing this statement. Holdren questioned the constitutionality of the ban when he testified before Wolf's subcommittee. He said that the Department of Justice had advised that the language in the resolution "should not be read as prohibiting interactions that are part of the president's constitutional authority to conduct negotiations". Representative John Culberson (Republican, Texas) accused Holdren of planning to violate the law and threatened legal and financial reprisals. He questioned Holdren about Chinese spies at the OSTP and NASA, and claimed that Holdren was "blindly ignoring the threat posed by China".

Why do members of Congress fear that advocates of scientific cooperation such as Holdren are naively helping China to win a new cold war?

SHUTTLE DIPLOMACY

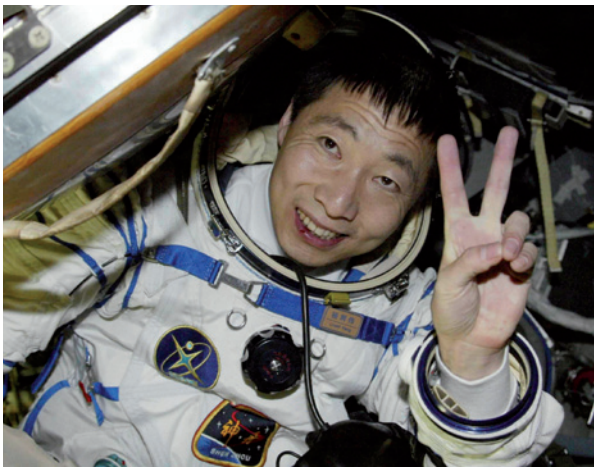
The US–Chinese relationship in space began with cold-war intrigue. In 1950, the US government detained the Chinese-born rocket scientist Qian Xuesen on suspicion of communist sympathies. Qian was a US citizen with top-security clearance who served as director of the Jet Propulsion Laboratory at the California Institute of Technology in Pasadena. The charges were never substantiated. Under virtual house arrest, his career crumbling, Qian returned to China in 1955, where he spent the rest of his career leading China's efforts to develop ballistic missiles and space-launch vehicles. To this day, Chinese American scientists working for the US government do so under suspicions created by the legacy of this affair.

In 1979, diplomatic relations between the United States and China were normalized. Scientific and technological cooperation then became a tool of US military policy, as the United States sought to play China off against the Soviet Union. In 1984, President Reagan travelled to what he later described as "so-called communist China"

to consolidate an extensive programme of scientific exchanges, including technical support for the modernization of China's armed forces.

Relations warmed further in 1986, when the US space-shuttle fleet was grounded by the *Challenger* disaster. The Reagan administration agreed to let US satellite manufacturers contract with China to arrange much-needed space-launch services. But after Chinese troops opened fire on unarmed protesters in the Tiananmen Square massacre of 1989, Congress imposed sanctions prohibiting those launches and exports of space and nuclear technology. These sanctions are still in effect, giving weight to arguments by some that the US–China Joint Statements of 2009 and 2011 go against congressional wishes and challenge congressional authority.

From the late 1980s to the early 1990s, President George H. W. Bush granted waivers of these sanctions, allowing the United States to use Chinese launches. During his campaign for presidency, Bill Clinton accused Bush of "coddling dictators". But after his election, he too issued waivers and even simplified the waiver-approval process. In 1998, *The New York Times* columnist William Safire accused President Clinton of selling space technology to the Chinese military in exchange for campaign contributions, and Congress launched an investigation. Administration officials admitted that the waivers were political — they were an incentive for China to honour its agreement not to sell missile technology to Iran and Syria. No charges were filed, but the issue of cooperation with China in space became highly politicized.



AFP/GETTY IMAGES

Yang Liwei piloting the first manned Chinese space mission in 2003.

Then, in February 1996, a US communications satellite was destroyed during a failed Chinese launch. It was reported that US workers had given Chinese officials investigating the accident sensitive technical information with potential military uses. This too led to an investigation, which figured prominently in the final report of a select House committee on commercial and scientific relations between the United States and China. The 1999 Cox Report, named after chairman Christopher Cox (Republican, California), ended the satellite launches and led to tighter restrictions on all scientific and technical cooperation with China. George W. Bush did not challenge these restrictions.

China took them in its stride, increasing domestic spending on space and completing its first human space flight in October 2003. When the United States pressured Europe to limit Chinese participation in the Galileo project, the European rival to the Global Positioning System, China began building its own system. US refusals to allow China to join the International Space Station prompted Chinese leaders to green-light a national space-station plan. In January 2007, Chinese tests of anti-satellite weaponry included the dramatic destruction of a defunct Chinese weather satellite used for target practice, scattering debris that threatens to strike other satellites and the International Space Station. In sum, the sanctions meant to inhibit Chinese progress in space have done no such thing.

COMMON GROUND

Some see a moral argument behind the bans on US–China collaboration, arguing that China's human-rights record necessitates a certain distance. Wolf is one such person. He challenged Holdren to justify cooperation with a regime still guilty of the same human-rights abuses that precipitated the Tiananmen sanctions. Holdren explained that the administration's pursuit of scientific and technological cooperation was not an endorsement of the Chinese government or a reward for good behaviour, but was necessary to address complicated threats to US interests, including climate change, nuclear terrorism and space debris. These are problems that require international solutions, and if China fails to address them it will harm the US public.

Scientific diplomacy has long brought together individuals interested in solving problems, who share a common language and methodology that helps them to overlook national and cultural differences. Of course, scientists are not immune to economic interests, bureaucratic entanglements and political passions, but my personal and professional experience suggests that they are able to set them aside more easily than politicians, diplomats or soldiers.

The Union of Concerned Scientists has more than two decades of experience working with scientists from across the world on public-policy issues that require international scientific and technical cooperation, notably: the pursuit of nuclear disarmament; nuclear-power safety and security; and security in outer space. More than 100 Chinese scientists and engineers have participated in our annual Summer Symposiums on Science and World Affairs, and many have gone on to hold important positions in Chinese institutes. The China Project that I manage for the union has used these relationships to expand dialogue and pursue joint research on the contentious security issues that have long divided Chinese and US scientists.

In the process, we have established relationships with China's defence-science community that our counterparts in government cannot. We have used these relationships to ease Chinese anxieties about US missile defences, US discussions of developing new nuclear weapons, and the Pentagon's overblown rhetoric about 'space control'. We are starting discussions aimed at helping Chinese scientists to make decisions that reduce the risks of nuclear power. Scientists with access to good support networks and information can make a positive difference to government policies, or at least try to. The Union of Concerned Scientists has learned that the January 2007 anti-satellite test was approved in part thanks to distorted representations of data on space debris provided by the Chinese military. We watched with admiration as our Chinese research colleagues risked their careers to provide their government with more reliable information.

US scientific and technological cooperation is unlikely to reform China's oppressive political system. But experience shows that it can contribute to mitigating some of the dangers to US citizens. It can also foster communication and understanding, setting the stage for cooperation on issues that require international solutions.

Wolf is to be commended for his support for human rights, and congressional concerns about US–China rivalry are understandable. But President Obama's decision to use scientific cooperation as an instrument of diplomacy is not only his constitutional prerogative, it is the best way to advance US interests. The two joint statements and the efforts of NASA and the OSTP are hopeful signs that the Obama administration is pursuing a healthy approach to collaboration. Gaining congressional support for such efforts requires a broader awareness of the constructive role that scientists can have in international relations. ■

Gregory Kulacki *is a senior analyst and the China Project manager for the Union of Concerned Scientists.*
e-mail: gkulacki@ucsusa.org

From UFOs to conspiracy theories, we form beliefs and then look for reasons to support them.

J. BERGMARK/GLOWIMAGES.COM

PSYCHOLOGY

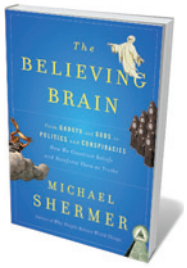
How we form beliefs

Religions and superstitions may stem from the brain's ability to spot patterns and intent, finds **A. C. Grayling**.

Two long-standing observations about human cognitive behaviour provide Michael Shermer with the fundamentals of his account of how people form beliefs. One is the brain's readiness to perceive patterns even in random phenomena. The other is its readiness to nominate agency — intentional action — as the cause of natural events.

Both explain belief-formation in general, not just religious or supernaturalistic belief. Shermer, however, has a particular interest in the latter, and much of his absorbing and comprehensive book addresses the widespread human inclination to believe in gods, ghosts, aliens, conspiracies and the importance of coincidences.

Shermer is well equipped for this task. He is a psychology professor, the founder of *Skeptic* magazine and resident sceptical columnist for *Scientific American*. Once an evangelical Christian, he lost his faith largely as a result of his college studies of psychology and cognitive neuroscience.



The Believing Brain: From Ghosts and Gods to Politics and Conspiracies — How We Construct Beliefs and Reinforce Them as Truths
MICHAEL SHERMER
Times Books: 2011. 400 pp. $28, £19.99

The important point, Shermer says, is that we form our beliefs first and then look for evidence in support of them afterwards. He gives the names 'patternicity' and 'agenticity' to the brain's pattern-seeking and agency-attributing propensities, respectively. These underlie the diverse reasons why we form particular beliefs from subjective, personal and emotional promptings, in social and historical environments that influence their content.

As a 'belief engine', the brain is always seeking to find meaning in the information that pours into it. Once it has constructed a belief, it rationalizes it with explanations, almost always after the event. The brain thus becomes invested in the beliefs, and reinforces them by looking for supporting evidence while blinding itself to anything contrary. Shermer describes this process as "belief-dependent realism" — what we believe determines our reality, not the other way around.

He offers an evolution-based analysis of why people are prone to forming supernatural beliefs based on patternicity and agenticity. Our ancestors did well to wonder whether rustling in the grass indicated a predator, even if it was just the breeze. Spotting a significant pattern in the data may have meant an intentional agent was about to pounce.

Problems arise when thinking like this is unconstrained, he says. Passionate investment in beliefs can lead to intolerance and conflict, as history tragically attests. Shermer gives chilling examples of how dangerous belief can be when it is maintained against all evidence; this is especially true in pseudoscience, exemplified by the death of a ten-year-old girl who suffocated during the cruel 'attachment therapy' once briefly popular in the United States in the late 1990s.

Shermer's account implies that we are far from being rational and deliberative thinkers, as the Enlightenment painted us. Patternicity leads us to see significance in mere 'noise' as well as in meaningful data;

NATURE.COM
Pascal Boyer on religion and neuroscience:
go.nature.com/gpiibe

agenticity makes us ascribe purpose to the source of those meanings. How did we ever arrive at more objective and organized knowledge of the world? How do we tell the difference between noise and data?

His answer is science. "Despite the subjectivity of our psychologies, relatively objective knowledge is available," Shermer writes. This is right, although common sense and experience surely did much to make our ancestors conform to the objective facts long before experimental science came into being; they would not have survived otherwise.

Powerful support for Shermer's analysis emerges from accounts he gives of highly respected scientists who hold religious beliefs, such as US geneticist Francis Collins. Although religious scientists are few, they are an interesting phenomenon, exhibiting the impermeability of the internal barrier that allows simultaneous commitments to science and faith. This remark will be regarded as outrageous by believing scientists, who think that they are as rational in their temples as in their laboratories, but scarcely any of them would accept the challenge to mount a controlled experiment to test the major claims of their faith, such as asking the deity to regrow a severed limb for an accident victim.

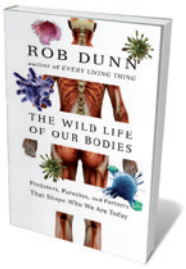
Shermer deals with the idea that theistic belief is an evolved, hard-wired phenomenon, an idea that is fashionable at present. The existence of atheists is partial evidence against it. More so is that the god-believing religions are very young in historical terms; they seem to have developed after and perhaps because of agriculture and associated settled urban life, and are therefore less than 10,000 years old.

The animism that preceded these religions, and which survives today in some traditional societies such as those of New Guinea and the Kalahari Desert, is fully explained by Shermer's agenticity concept. It is not religion but proto-science — an attempt to explain natural phenomena by analogy with the one causative power our ancestors knew well: their own agency. Instead of developing into science, this doubtless degenerated into superstition in the hands of emerging priestly castes or for other reasons, but it does not suggest a 'god gene' of the kind supposed for history's young religions with their monarchical deities.

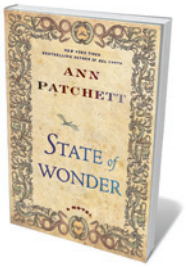
This stimulating book summarizes what is likely to prove the right view of how our brains secrete religious and superstitious belief. Knowledge is power: the corrective of the scientific method, one hopes, can rescue us from ourselves in this respect. ■

A. C. Grayling *is professor of philosophy at Birkbeck College, University of London. His latest publication is* The Good Book.
e-mail: a.grayling@philosophy.bbk.ac.uk

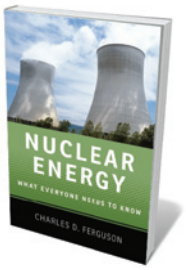
Books in brief



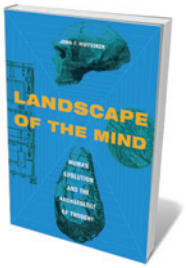
The Wild Life of Our Bodies: Predators, Parasites, and Partners That Shape Who We Are Today
Rob Dunn HARPER *304 pp. $26.99 (2011)*
Ridding our bodies of microbes is one of the goals of modern medical science and sanitation. But antibiotics and cleanliness might be causing us more harm than good, according to biologist Rob Dunn. In his provocative book, he sings the praises of intestinal worms for keeping our guts healthy, and urges us to 'rewild' our internal ecosystems. By distancing ourselves from other wildlife, he argues, humans are harming themselves both mentally and physically. The cure is to get our bodies back in touch with other species.



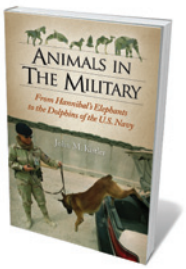
State of Wonder: A Novel
Ann Patchett HARPER *368 pp. $26.99 (2011)*
A research scientist is the heroine of best-selling author Ann Patchett's latest novel. Marina Singh is called away from her lab bench at a large pharmaceutical company in Minnesota to search in the Brazilian rainforest for her old mentor, who has disappeared while looking for medicinal plants. The last person sent out to Brazil on the same mission was Singh's research partner and friend, who never returned. Singh tracks down her mentor, but what she finds in the jungle raises questions about her friend's fate, her company's future and her own past.



Nuclear Energy: What Everyone Needs to Know
Charles D. Ferguson OXFORD UNIVERSITY PRESS *240 pp. $16.95 (2011)*
With concerns about nuclear energy on the rise after Japan's Fukushima disaster, security scholar Charles Ferguson's accessible overview is timely. He sets out how nuclear power is generated, which countries use it and how much electricity they produce. He discusses safety issues, from defending plants against military attacks to technical failures. And he notes the roller-coaster of public attitudes to nuclear power, which have oscillated in the past decade from growing acceptance of its utility in reducing greenhouse-gas emissions to recent proposals to abandon the industry.



Landscape of the Mind: Human Evolution and the Archaeology of Thought
John F. Hoffecker COLUMBIA UNIVERSITY PRESS *288 pp. $50 (2011)*
The origins of the human mind may lie in feedbacks between our brains, hands and tools, suggests palaeoanthropologist John Hoffecker. Social, physical and mental experiences combined long ago to form a 'superbrain' network, he suggests, giving rise to consciousness and creativity. Technologies were integral to expanding our horizons and to abstract thinking. Archaeological evidence links the rise of symbolism with tool development, following people's adaptation to different climates as they spread worldwide.



Animals in the Military: From Hannibal's Elephants to the Dolphins of the U.S. Navy
John M. Kistler ABC CLIO *343 pp. £44.95 (2011)*
Animals have a long history in the military, as writer John Kistler recounts. Horses have lugged men and equipment across the battlefields of history and are still in use today in combating Taliban fighters in remote regions of Afghanistan. Elephants, dogs and stinging insects have been used to attack enemies. Kistler relates the heroics and horrors of animals in warfare, from the dolphins of the US Navy to message-delivering pigeons and mine-sniffing rats.

DEMOGRAPHICS

The growth of nations

Michael Sargent enjoys a social history of how height and lifespan increased during the Industrial Revolution.

The idea that a wild mammal might reveal an unsuspected genetic potential for greater body size and for doubling its life expectancy in only 12–15 generations would astonish most naturalists. Yet this is exactly what happened to humans of the developed world after 1700.

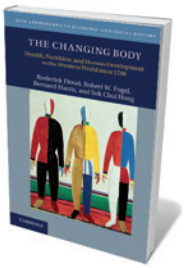
In a meticulous reconstruction of the recent evolution of human stature and longevity, *The Changing Body* convincingly links crucial episodes in social history with biological consequences using data extracted from obscure public records, such as military recruitment logs. The book's authors — a mix of economists and social historians — ascribe these trends in the development of the human body to a synergy between technological advance and physiological adaptation, which led to improved health, agricultural yields and industrial productivity.

This virtuous circle of progress produced extraordinary improvements in lifespan. The increase in longevity — one of the most important outcomes of the Industrial Revolution — originated largely in the greater robustness and capacity of vital organ systems, changes that are endorsed by recent insight into human development.

> ***"Since 1950, the height of US-born males has declined, whereas Dutch males have emerged as the world's tallest."***

The book explains how advances in health and longevity were aided by a mysterious increase in fertility that began in the seventeenth century. This, coupled with the improved food production and distribution of the agricultural revolution, powered a population explosion without provoking a subsistence crisis. Improved sanitation and public-health management in the late nineteenth century helped to prevent the spread of infectious disease, further boosting productivity by massively reducing the time workers lost to sickness.

The authors surmise that most eighteenth-century Europeans barely consumed enough calories to meet the minimal maintenance requirements of the human body, based on estimates of food available and contemporary descriptions of diet. One-fifth of the population subsisted on even less: beggars and paupers, too weak to work effectively. From 1800, the authors argue, a higher calorie intake and reduced risk from infectious disease led to increased industrial efficiency. Slave masters in the old American plantations understood this issue well, providing substantial rations of cornmeal, salted herrings and pork to get more work out of their labour force. By contrast, many Europeans were stunted, except for the better-fed social elite. For example, between 1750 and 1860, officer cadets at Sandhurst, the British army's training college, were 14 centimetres taller on average than recruits to the British merchant navy, who came from the poorest strata of society.



The Changing Body: Health, Nutrition, and Human Development in the Western World since 1700
RODERICK FLOUD, ROBERT W. FOGEL, BERNARD HARRIS & SOK CHUL HONG
Cambridge University Press: 2011. 456 pp. £19.99, $32.99 (pbk)

The authors make use of charts showing the relationship between height, weight and risk of death in modern Norwegians and in veterans of the American Civil War, which they extrapolate to Europeans from 1700 onwards. Improved nutrition, they say, was the biggest factor in reducing death risk until the 1870s, and declining exposure to infection became more significant thereafter.

Average stature, as a quantitative measure of human welfare, gives insight into circumstances affecting an entire population. For example, following US independence in 1766, early male settlers became the world's tallest men, with longer lifespans than their European peers (57 years compared with 40 years in Europe) thanks to a greater abundance of food. But by the mid-nineteenth century, US food producers were struggling to feed the rapidly rising population, and the average height of US males fell by 4.4 cm.

Childhood diarrhoeal infections also caused stunting of this generation: epidemics were rife in the densely populated cities of the eastern seaboard, which teemed with immigrants. Forty years later, the drop in calorie intake that affected many during the Great Depression of the 1930s had little effect on stature because by then, the country had in place a fully developed system for controlling the spread of infectious disease.

NATURE.COM
For more on how nutrition affects fetal development:
go.nature.com/kwxtkb

Since 1950, the height of US-born males has declined slightly, whereas Dutch males have emerged as the world's tallest, by 4.5 cm on average. This could indicate greater equality of nutritional opportunity in childhood in the Netherlands compared with the United States. However, the emergence of a genetic difference previously obscured by undernutrition is also possible — a hypothesis that could be tested on people of Dutch descent living in the United States. Another possibility not considered by the authors is that shorter migrant families could be contributing to the statistics.

The complex and original narrative in *The Changing Body* is the culmination of decades of detective work, embracing the diverse histories of many nations. The statistical analysis involved is subtle, sophisticated and must be taken on trust by most of us. The conclusions, however, are convincing and reveal aspects of our biological history that could not have emerged in any other way. They are also consistent with current understanding of the determinants of stature and longevity; prenatal malnutrition undermines full development of the vital organs of the adult body, making premature age-related diseases more likely.

Using data showing that US Civil War veterans succumbed to age-related diseases well in advance of comparable groups in modern society, the authors provide striking support for the idea that the longevity of this age group within a population is affected by similar experiences in early life. Developmental biologists are accumulating evidence showing that the function of crucial genes is modified *in utero* by an epigenetic process — that does not alter the DNA sequence — which reduces organ development and has consequences for robustness in later life. These changes can be transmitted to the next generation, explaining the need for several generations to elapse after a nutritional improvement before a small-stature trait is lost.

The anthropometric approach to history described in this book is impressive and skilful, and is likely to become routine in the study of future influences on human development. ■

Michael Sargent *is at the National Institute for Medical Research, Mill Hill, London and author of* Biomedicine and the Human Condition.
e-mail: msargen@nimr.mrc.ac.uk

Mark Dion's cabinet of curiosities at the Oceanographic Museum of Monaco, illuminated by chandeliers modelled on zoologist Ernst Haeckel's jellyfish sketches.

NOUVEAU MUSÉE NATIONAL DE MONACO/OCEANOGRAPHIC MUS. MONACO/D. MILLE

ART

Treasures fit for a prince

A Monaco exhibition showcases the marvels of the city-state's oceanic museum, discovers **Alison Abbott**.

Like many modern royals keen to demonstrate their relevance to society, Prince Albert II of Monaco has adopted a good cause — the protection of the marine environment. For his marriage on 2–3 July, he even tried to side-step protocol to hold his wedding dinner, not in his glamorous sixteenth-century palace, but inside the Oceanographic Museum of Monaco.

Clinging to a rocky cliff above the Mediterranean, the 1910 museum is a jewel in an anachronistic city-state scarred by shabby high-rises. Had its inadequate kitchens not scotched the prince's plans at the last minute, wedding guests would next weekend be dancing beneath antique chandeliers modelled on drawings of jellyfish and radiolaria by zoologist, artist and passionate Darwinist, Ernst Haeckel.

They could also have mingled with imposing modern artworks — including one of the world's largest oceanographic art installations, by US artist Mark Dion; and French–Chinese artist Huang Yong Ping's 25-metre octopus–cuttlefish hybrid, whose tentacles curl around the main hall's marble columns. Dion's 'cabinet of curiosities' is part of a major exhibition, *Oceanomania,* of his and other artists' work, running at the museum until the end of September, and extending to the New National Museum of Monaco at the Villa Paloma.

Noted for its historic subterranean aquarium and its world-class research on coral, the Oceanographic Museum was founded by the prince's great-great grandfather, Albert I. It was run by the French Academy of Sciences until 2006, when its management returned to Monaco under Albert II, who wishes to unite within it scientific and artistic approaches to examining the ocean.

Dion is famous for his interest in collections of all kinds and was naturally drawn to the museum's splendid scientific collection. Albert I led 28 expeditions in the North Atlantic and Mediterranean between 1885 and 1915, and brought back all manner of exotic flora and fauna. These formed the museum's initial collection together with his oceanographic instruments. Extended by large acquisitions of objects and artworks, the holdings now exceed 400,000 items.

Oceanomania
Oceanographic Museum and the New National Museum (Villa Paloma), Monaco. Until 30 September.

Dion scoured the collection for items to display in a giant cabinet of curiosity, or *Wunderkammer*, modelled on those of the sixteenth and seventeenth centuries. His installation has an overall beauty, but the viewer is also drawn to the individual, wonder-inspiring items within it, such as an 1863 underwater landscape painted by the Belgian artist Eugène de Ransonnet from within a diving bell, or a magnificent stuffed polar bear.

Close up, you see that the objects are categorized according to different — often non-scientific — criteria on different shelves. One grouping contains just teeth, another is of diverse objects assembled by one collector, another contains objects of similar colour. In this way, Dion aims to challenge the conventional mindset of the museum establishment.

But there is a sense of repetition here. Dion has created other such cabinets — his largest and most magnificent was for an exhibition celebrating 300 years of science in Berlin (*Nature* **467,** 660; 2010). Paradoxically, he has become a darling of the very establishment he seeks to challenge.

Dion's overarching theme for *Oceanomania* is the contrast between our wonder at ocean biodiversity and fear that we are destroying it. Other Dion works on display include a model of a sturgeon — fished to near extinction for caviar — beached on a mound of discarded coins, and a stuffed flamingo coated in tar.

Given its location, *Oceanomania* is likely to attract thousands of visitors, which I hope will trigger sympathy for neglected historical collections elsewhere. The endorsement of a royal wedding will do it no harm. The bride, incidentally, also fits the aquatic theme: she is a former Olympic swimmer. ■

Alison Abbott *is* Nature*'s Senior European Correspondent.*

CORRESPONDENCE

Don't base Spanish funding on citations

The Spanish government announced in March that Spain's scientific research quality has overtaken that of Australia and Switzerland (see go.nature.com/si76qc). It drew this conclusion from a 2011 report by Britain's Royal Society (see go.nature.com/q8hmxw). I question the government's interpretation, given that the Swiss win at least one Nobel prize in science every decade and Spain has not won one for 100 years.

The Royal Society report uses total citation counts as a measure of quality and impact. But citation figures can be misleading, depending on how they are derived. For example, Switzerland and Australia emerge well ahead of Spain when adjustment is made for high-profile, but low-frequency, breakthroughs among the routine results that dominate citation counting (A. Rodríguez-Navarro *PLoS ONE* **6,** e20510; 2011).

The Spanish government's apparent misinterpretation has led to a reduction in tenure for young scientists, among other funding cuts justified on the grounds that Spain has too many researchers. The government should look instead to the excessive number of university professors, who are not always picked for their research record.

Spain's scientific societies should oppose this misguided research policy if the country's science is to improve.

Alonso Rodríguez-Navarro
Polytechnic University of Madrid, Spain. alonso.rodriguez@upm.es

Is China producing too many PhDs?

China produced an estimated 117,000 PhD graduates in 2010 — more than any other nation, and a tenfold increase on 1999 figures (see go.nature.com/xbdqm9; in Chinese). It looks set to fulfil the government's goal of becoming a leader in higher education — except that the quantity of PhDs awarded has diluted their quality.

This boom has overburdened the country's education infrastructure. Standards are compromised by a lack of proper quality control and of elimination mechanisms for trainee PhDs (*Nature* **472,** 276–279; 2011).

Employment opportunities for home-grown PhDs are dwindling. Universities and research institutes want to recruit Chinese people with foreign PhDs, and corporations prefer graduates with a master's degree because they cost less.

Given the global excess of PhDs (*Nature* **472,** 280–282; 2011), China must remodel its education system and introduce rigorous elimination mechanisms in PhD training.

Zhu Liu, Yong Geng *Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, China. liuzhu@iae.ac.cn*

Reward research that informs policy

We agree that science policy can and should encourage research programmes that deliver public values — the core aspirations that motivate and justify public spending on science (*Nature* **473,** 123–124; 2011).

The public expects climate research to generate useful, reliable knowledge. This is more likely when programmes engage with the policy and public context in which they are embedded. Such engagement is not easy, or popular. It demands creativity, resources and collaboration with diverse communities.

Scientists working on important societal problems are usually subject to the same metrics of success as theoretical physicists and mathematicians — peer-reviewed articles. Nevertheless, some scientists and organizations generate useful information for policy-makers and for public benefit. More must be done to recognize and reward their contribution, and encourage others to follow their example.

Thus, it is not a question of whether the United States has enough institutional capacity, as you ask, but how it can develop appropriate kinds of institutions. Pouring money into the same institutional framework is unlikely to bring climate research into line with society's expectations.

Ryan Meyer, Peat Leith
University of Tasmania, Australia. ryan@ryanmeyer.org

Promoting science careers in Brazil

The prestigious State University of Campinas, which accounts for almost 15% of Brazil's output of scientific research papers, has been running a successful programme since 2003 that has similar goals to *Scientific American*'s '1,000 Scientists in 1,000 Days' initiative (*Nature* **473,** 123; 2011).

'Vocations in Science and the Arts' helps public high schools in the city. Students spend a month in one of the university's research laboratories, learning how to conduct a simple research project and how to interpret their findings. They then present these at a university 'congress'. They feel like scientists from start to finish, and get a sense of the highs and lows of research work.

Our molecular-genetics lab has so far helped 18 students under this programme, of whom at least one-third are intending to pursue a career in biology.

Lucas Leite Cunha, Laura Sterian Ward, *State University of Campinas, Brazil. lucasleitecunha@gmail.com*

Community is key to REDD success

As workers with indigenous communities in the Solomon Islands, we believe a fundamental issue must be addressed if REDD (the United Nations' Collaborative Programme on reducing emissions from deforestation and forest degradation) is to respect the rights of indigenous peoples meaningfully (*Nature* **472,** 390; 2011).

Indigenous landowners have set up the Tetepare Descendants' Association (TDA) to protect the large, uninhabited island of Tetepare, an important carbon sink. The association has successfully countered threats posed by timber companies contracting with individuals across the Solomons for timber rights that belong to clans or tribes.

Such community organization is essential for engagement with REDD. But, ironically, it has become harder for the TDA to certify its carbon under the Voluntary Carbon Standard (VCS), a carbon-credit verification programme. This is partly because the TDA's short-term success has made it more complicated to prove that a VCS project would result in additional carbon sequestration.

The existence of a community-based organization such as the TDA should not hinder participation in REDD. Community commitment to preventing deforestation is essential if UN programmes are to succeed.

Unless REDD provides the incentive for legitimate community-governance structures, indigenous rights will not be respected.

Brian Weeks *New York, USA. bcweeks@gmail.com*
Christopher Filardi *Missoula, Montana, USA.*

SOCIAL NEUROSCIENCE

Stress and the city

Many of us were raised or currently live in an urban environment. A neuroimaging study now reveals how this affects brain function when an individual is faced with a stressful situation. SEE LETTER P.498

DANIEL P. KENNEDY & RALPH ADOLPHS

The landscape of human society is changing drastically. In 1950, only 30% of the world's population lived in urban regions; today, more than 50% of us do so; by 2050, this figure is expected[1] to have climbed to almost 70% (Fig. 1). And, just as social isolation is well known to have harmful effects[2], so does the opposite extreme: overcrowding can induce stress and illness in species ranging from insects to rodents[3] to primates, including humans[4]. In particular, mental illness in humans has been linked to the urban environment: living in a city increases the risk of depression and anxiety, and the rate of schizophrenia is markedly higher in people born and brought up in cities[5]. Writing on page 498 of this issue, Lederbogen *et al.*[6] use functional magnetic resonance imaging to investigate for the first time the specific human brain structures that are affected by urban living.

The study's participants lived or had lived in locations ranging from rural areas to large cities (Fig. 2). The authors measured regional brain activation while participants performed a social-stress test — solving hard arithmetic problems under time pressure and with negative feedback from the experimenter. This task not only increased the participants' heart rate, blood pressure and salivary levels of the stress-associated hormone cortisol, but also resulted in significant activity in brain structures known to be involved in emotion and stress.

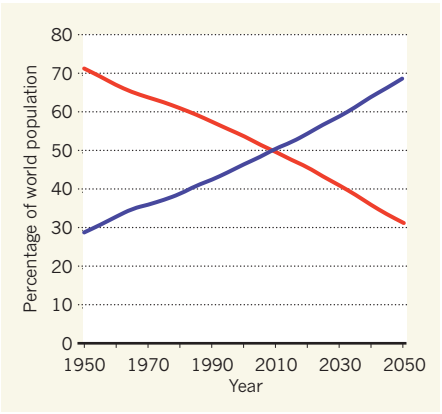


Figure 1 | The city allure. The percentage of the global population living in urban (blue) and rural (red) regions since 1950, with projected[1] figures up to 2050.

Of the activated brain regions, two were of particular interest: activation in the amygdala correlated with the size of the city in which an individual currently resided, and activation of the perigenual anterior cingulate cortex (pACC) correlated with how long a participant had lived in a large city during their childhood. Urban upbringing also affected the strength of the functional coupling between the amygdala and the pACC: those who had spent more time growing up in large cities had reduced functional connectivity between these two regions.

Intriguingly, a similar pattern of reduced amygdala–pACC coupling has previously been associated[7] with genetic risk for psychiatric disorders, and the amygdala has recently been linked both to social-network size[8] and to the sense of personal-space violation[9]. Taken together, the findings suggest that the cingulate–amygdala circuit is one on which genetic and environmental risks for mental illness may converge.

The sheer number and complexity of the factors involved in studies of real-world society raise concerns about both the reliability of effects and the possibility of confounding explanations. To address the first of these concerns, Lederbogen *et al.*[6] replicated their findings in several separate samples, used two different stress-inducing tasks, and demonstrated that there were no effects of urbanicity on brain activation when participants performed a non-stressful cognitive task.

The second concern — that urbanicity may be confounded by other variables associated with but causally separable from it — is difficult to address, given the enormous

PHOTOLIBRARY.COM; a, R. TOMLINSON; b, S. GRANDADAM; c, EYE UBIQUITOUS

Figure 2 | Regional categories. In their study[6], Lederbogen *et al.* categorized living conditions as those asssociated with rural areas (**a**), towns with more than 10,000 inhabitants (**b**) and cities with more than 100,000 inhabitants (**c**). Their data suggest that city living affects the brain's response to stress.

number of such variables. To probe some of the possibilities, the authors examined participants' age, education, income, marital and family status, as well as aspects of their health, mood, personality and the amount of social support they had. None of these factors significantly influenced the effects of urbanicity, suggesting that living in a city environment changes brain response during a social stressor by a distinct, although mysterious, mechanism.

Given that Lederbogen and colleagues' study is purely correlational, an obvious next step will be to conduct larger-scale, longitudinal studies that measure more variables and that attempt to trace a causal factor linking brain activity to more fine-grained and individualized aspects of city living. Such studies could, for example, measure individuals' perceived position in a social hierarchy and frequency of encounter with strangers, as well as population density, amount of space and type of housing.

There are wide variations in individuals' preferences for, and ability to cope with, city life: some thrive in New York City; others would happily swap it for a desert island. Psychologists have found[10] that a substantial factor accounting for this variability is the perceived degree of control that people have over their daily lives. Social threat, lack of control and subordination are all likely candidates for mediating the stressful effects of city life, and probably account for much of the individual differences seen.

But although the negative aspects of city living have been highlighted extensively, city life is by no means always bad. In many countries, for example, studies on the complex relationship between urbanicity and suicide show higher rates of suicide in rural areas than in cities[11]. Although there are a number of possible explanations for this observation, it could relate to cities' provision of a richer, more stimulating and more interactive social environment, a larger social-support network and easier access to medical care.

Future work complementary to Lederbogen and co-workers' study[6] could investigate the positive effects of city living in more detail and begin to make recommendations for urban planning and architecture. Given the world's increasing population (estimated to hit 7 billion this autumn), the fact that we will be living mostly in cities seems inescapable. This highlights the importance of understanding the effects that such living conditions will have on human mental health. ■

Daniel P. Kennedy *and* **Ralph Adolphs**
are at the California Institute of Technology, Pasadena, California 91125, USA.
e-mail: radolphs@hss.caltech.edu

1. http://esa.un.org/unpd/wup
2. Harlow, H. F., Dodsworth, R. O. & Harlow, M. K. *Proc. Natl Acad. Sci. USA* **54,** 90–97 (1965).
3. Calhoun, J. B. *Ann. NY Acad. Sci.* **51,** 1113–1122 (1950).
4. Hall, E. T. *The Hidden Dimension* (Doubleday, 1966).
5. Krabbendam, L. & van Os, J. *Schizophr. Bull.* **31,** 795–799 (2005).
6. Lederbogen, F. *et al. Nature* **474,** 498–501 (2011).
7. Pezawas, L. *et al. Nature Neurosci.* **8,** 828–834 (2005).
8. Bickart, K. C., Wright, C. I., Dautoff, R. J., Dickerson, B. C. & Barrett, L. F. *Nature Neurosci.* **14,** 163–164 (2011).
9. Kennedy, D. P., Gläscher, J., Tyszka, J. M. & Adolphs, R. *Nature Neurosci.* **12,** 1226–1227 (2009).
10. Fleming, I., Baum, A. & Weiss, L. *J. Pers. Soc. Psychol.* **52,** 899–906 (1987).
11. Hirsch, J. K. *Crisis* **27,** 189–199 (2006).

MATERIALS SCIENCE

Graphene moiré mystery solved?

In systems consisting of just a few layers of graphene, the relative orientation of adjacent layers depends on the material's preparation method. Light has now been shed on the relationship between stacking arrangement and electronic properties.

ALLAN H. MACDONALD & RAFI BISTRITZER

Moiré patterns appear whenever two regular templates are overlaid at an angle — commonly as artful designs in textiles, and unfortunately also as annoying distractions in digital images. More than 20 years ago, periodic modulations (superlattices) were discovered[1] in scanning tunnelling microscopy (STM) studies of graphite surfaces and correctly interpreted as moiré patterns caused by misorientations between subsurface graphene layers. (Graphene is a two-dimensional, pure carbon material that forms three-dimensional graphite when stacked in layers.)

The graphitic moiré pattern phenomenon has recently reappeared in studies of materials consisting of only a few layers of graphene, especially in systems prepared by the technique of epitaxial growth on a silicon carbide (SiC) substrate[2] or by the process of chemical-vapour-decomposition growth on metallic substrates. But as in the graphite case[3], the relationship between stacking rotations and electronic properties has been difficult to sort out. Writing in *Physical Review Letters*, Lucian *et al.*[4] provide insight into this relationship by combining advanced microscopy and spectroscopy techniques.

The problem of identifying the relationship between stacking rotations and electronic properties is particularly intriguing in the few-layer case, both because it should be tractable and because it is known[5] that stacking arrangements can radically alter a material's electronic properties. For few-layer graphene systems, the main sticking point has been that studies[6–8] of epitaxial graphene layers grown on SiC have given the impression that adjacent layers that are misoriented by only a few degrees could, in stark contradiction to theory[9], be very weakly coupled electronically.

By combining high-magnetic-field STM and Landau-level spectroscopy on samples consisting of a few layers of graphene grown by chemical-vapour deposition, Lucian and colleagues[4] now convincingly establish that electronic decoupling occurs between layers misoriented by more than about 20° and that strong coupling occurs between layers rotated by less than about 2°. Moreover, the authors point to hints in previous epitaxial graphene data[8] which suggest that the layer pairs identified as weakly coupled may not have been adjacent. If confirmed by additional STM studies, this conjecture would allow for a consistent interpretation of all current data.

The authors' findings[4] provide an attractive jumping-off point for systematic studies of interlayer coupling in misoriented few-layer graphene systems. Many important and subtle mysteries remain, both at large and small rotation angles. For example, what does effective decoupling at rotation angles larger than about 20° mean quantitatively? Complete decoupling at the typical separation between adjacent graphene sheets of only approximately 0.3 nanometres would be truly unexpected.

When two graphene sheets, each with a honeycomb lattice structure, are overlaid at an angle (Fig. 1), the result is almost always not a crystal. A crystal forms only at a discrete set of commensurate rotation angles[9–12]. In graphene, most observable properties depend on electrons close to the Dirac point, the point at which the gap between the material's conduction and valence energy bands vanishes. These electrons have an unusual ultrarelativistic, massless behaviour, which is characterized by a velocity that does not vanish with momentum. And in the presence of a magnetic field they show a characteristic pattern of discrete energy states, known as Landau levels, that is disturbed when layers couple.

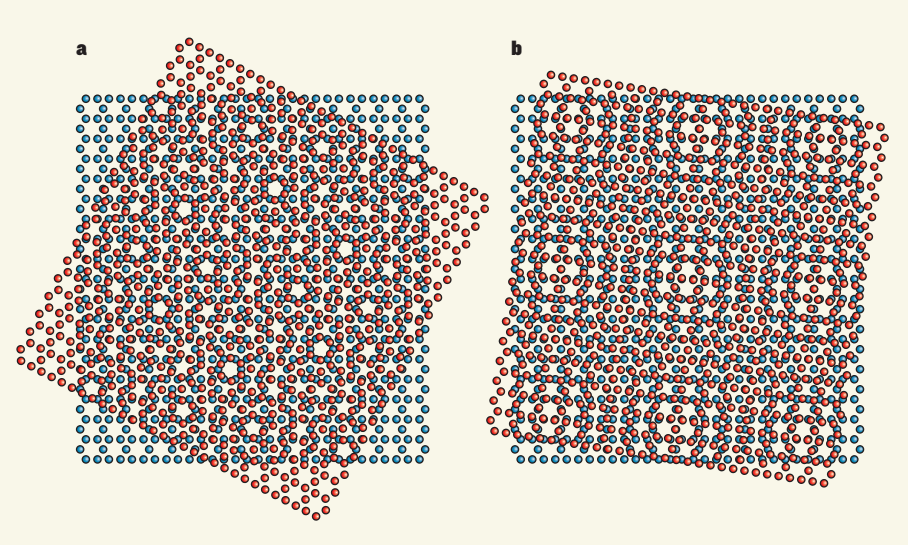


Figure 1 | Two examples of misoriented honeycomb lattices. **a**, The two lattices are overlaid at an angle of about 27.8°. The resulting atomic arrangement is precisely periodic, but has more atoms per period than has perfectly oriented (no rotation) Bernal stacking. **b**, Lattices rotated by 9°. Although the atomic arrangement never precisely repeats, there is a periodic pattern of points in space at which atoms from the two layers are nearly on top of each other, making the structure appear more open in this top view. When the local stacking arrangement varies slowly in space, electronic properties are insensitive to the atomic details that distinguish commensurate and nearby incommensurate rotation angles.

Because Dirac-point electronic states at crystalline misorientations are equal-weight sums of components in the two layers[9–12], it is not immediately obvious in what sense the layers can be considered to be decoupled. Theory answers this question in the first place in terms of energy gaps induced by energy-level repulsion between Dirac-point states in different layers. Numerical *ab initio* calculations find, however, that these gaps rapidly decrease with increases in the unit-cell area of commensurately rotated bilayers. In fact, for all but the commensurate structures with the smallest crystalline unit cells[9–12] (which occur at rotation angles near 21.8°, 27.8°, 32.2° and 38.2°), the gaps are smaller than the energy uncertainty implied by Heisenberg's uncertainty principle and the typical time between electron–sample-impurity collisions (the quantum lifetime), making them unobservable. A sample's quantum lifetime can be extracted[13] from STM Landau-level spectroscopy measurements. So far, gaps due to coupling between rotated layers have never been observed, suggesting either that they are smaller than generally expected[14] or that the rotation angles found experimentally do not correspond to short-period commensurate structures.

When the energy uncertainty is larger than the Dirac-point gap, the strength of interlayer coupling is more appropriately characterized by the interlayer conductance. This conductance is proportional to the area of the sample, increases without limit with quantum lifetime at commensurate rotation angles, and vanishes with quantum lifetime for incommensurate rotation angles. When the conductance is small, it can be characterized by the RC time, the time it takes for a voltage difference between the layers to relax to zero. The RC time is expected[14] to be very sensitive to quantum lifetimes in the individual graphene sheets, and long enough to be experimentally accessible — as long as about 10^{-8} seconds — when sample disorder is weak.

In their study, Lucian *et al.*[4] find that, at small rotation angles, the local density of electronic states develops a dependence on position within the moiré-pattern unit cell and no longer exhibits the Dirac-like, decoupled-layer, Landau-level pattern. Layer coupling becomes strong in this sense for rotation angles less than about 2°, corresponding to moiré-pattern periods longer than about 10 nanometres. Here it is tempting to conjecture — from the spatial dependence of the density of electronic states — that bilayer wavefunctions have become localized, so that an STM measurement at one position reflects the stacking arrangement only at that position.

But continuum models, which ignore local atomic arrangements and are expected to be more accurate at small rotation angles, suggest otherwise. We have separately used[15] continuum models to argue that Dirac-point states in the small-rotation-angle limit are characterized by narrow moiré energy bands whose width oscillates as a function of the reciprocal of the rotation angle. Greater understanding of these small-angle states is an interesting theoretical challenge, which may be informed by future experiments using samples that have weaker disorder than has so far been available, or by improved resolution of STM spectroscopy. Continuum models capture the moiré-pattern periodicity but ignore some details of the atomic arrangement. It is interesting that Bloch's theorem, which allows the electronic structure to be calculated relatively simply for periodic structures, can also be applied to these systems despite the fact that, crystallographically, they are quasi-periodic rather than periodic[15].

It has long been recognized[16] that commensurability between a unit-cell area in a periodic system and the area that encloses one quantum of magnetic flux should theoretically lead to complex, fractal electronic structures. This physics has, however, been difficult to address in the real world. On the one hand, typical unit-cell areas for crystalline periodicity are about 0.1 nm^2, dramatically smaller than the magnetic-flux quantum areas achievable experimentally (about 3,000 nm^2 divided by the magnetic field measured in tesla). On the other hand, any artificial periodicity fabricated lithographically tends to be too large. The moiré-pattern periods seem to fall nicely in the middle ground. This could allow[17] the physics of periodic electronic systems in external magnetic fields to be studied more fully than has previously been possible.

The extraordinary sensitivity of the electronic properties of few-layer graphene systems to the relative orientations of their layers could prove useful in various applications, for example in ultra-sensitive strain gauges, pressure sensors or ultra-thin capacitors. Further progress requires an improved understanding of both large and small rotation-angle limits, and also improved experimental control of rotation angles. ■

Allan H. MacDonald *and* **Rafi Bistritzer** *are in the Physics Department, University of Texas at Austin, Austin, Texas 78712, USA.*
e-mail: macdpc@physics.utexas.edu

1. Kuwabara, M., Clarke, D. R. & Smith, D. A. *Appl. Phys. Lett.* **56,** 2396–2398 (1990).
2. First, P. N. *et al. MRS Bull.* **35,** 296–305 (2010).
3. Pong, W.-T. & Durkan, C. J. *Phys. D* **38,** R329–R355 (2005).
4. Luican, A. *et al. Phys. Rev. Lett.* **106,** 126802 (2011).
5. McCann, E. & Falko, V. I. *Phys. Rev. Lett.* **96,** 086805 (2006).
6. Sadowski, M. L. *et al. Phys. Rev. Lett.* **97,** 266405 (2006).
7. Sprinkle, M. *et al. Phys. Rev. Lett.* **103,** 226803 (2009).
8. Miller, D. L. *et al. Science* **324,** 924–927 (2009).
9. Lopes dos Santos, J. M. B., Peres, N. M. R. & Castro Neto, A. H. *Phys. Rev. Lett.* **99,** 256802 (2007).
10. Shallcross, S. *et al. Phys. Rev. B* **81,** 165105 (2010).
11. Trambly de Laissardière, G., Mayou, D. & Magaud, L. *Nano Lett.* **10,** 804–808 (2010).
12. Mele, E. J. *Phys. Rev. B* **81,** 161405 (2010).
13. Song, Y. J. *et al. Nature* **467,** 185–189 (2010).
14. Bistritzer, R. & MacDonald, A. H. *Phys. Rev. B* **81,** 245412 (2010).
15. Bistritzer, R. & MacDonald, A. H. Preprint at arXiv:1009.4203 (2010).
16. Hofstadter, D. R. *Phys. Rev. B* **14,** 2239–2249 (1976).
17. Bistritzer, R. & MacDonald, A. H. Preprint at arXiv:1101.2606 (2011).

METABOLISM

A lipid for fat disorders

A high-fat diet often leads to metabolic disorders such as diabetes, fatty liver disease and obesity. One lipid, however, might mitigate these effects through an unexpected signalling role in the nucleus. SEE LETTER P.506

HOLLY A. INGRAHAM

Lipids are best known for their integral role in biological membranes and as signalling molecules in the cytoplasm. Rarely is their importance in nuclear processes appreciated, even though it has been known for decades that lipids and the machinery that modifies them can be found in the nucleus[1]. On page 506 of this issue, Lee and colleagues[2] explore the nuclear activities of lipids, showing that a particular species of the phospholipid phosphatidylcholine (which is a component of the food supplement lecithin) controls transcriptional programs. Their data also suggest that targeting lipid signalling in the nucleus might be of value for treating human metabolic diseases.

Structural studies[3,4] have shown that phospholipids bind to the NR5A subclass of nuclear receptor. Building on these observations, Lee *et al.* set out to find a phospholipid ligand that activates the nuclear receptor LRH-1 (also known as NR5A2), which is important for bile-acid homeostasis[5]. They reasoned that such an agonist would increase bile-acid levels and reverse conditions associated with fatty liver disease. Non-alcoholic fatty liver disease can lead to hepatic steatosis, which is often associated with other metabolic disorders, including obesity, insulin resistance and type 2 diabetes.

After screening several different phospholipids in cell-based assays, the authors identified two short-chain phosphatidylcholine species that, at high concentrations (100 μM), strongly activate LRH-1. Of these, they focused on dilauroyl phosphatidylcholine (DLPC) — a species with two saturated 12-carbon fatty acyl chains (Fig. 1).

A previous structural study[6] showed that bound phosphatidylcholine nestles tightly into LRH-1's closest homologue, SF-1 (also known as NR5A1). It would therefore be expected that both DLPC and a longer-chain phosphatidylcholine called DPPC would also bind to LRH-1. Intriguingly, however, Lee *et al.*[2] report that DLPC, and not DPPC, increased LRH-1 activity. Consistent with this selective activation, they also found that a 1:1 molar ratio of DLPC, but not DPPC, could displace phospholipids bound to LRH-1. Moreover, when the authors administered these phosphatidylcholines to mice orally, DLPC elevated the levels of some, but not all, LRH-1 targets in the liver and increased the serum levels of bile acid.

Emboldened by these positive results, Lee and co-workers put DLPC to a final test to investigate whether this phospholipid might reverse diet-induced insulin resistance and fatty liver disease in mice. They fed the mice a high-fat diet (the equivalent of a continuous rich, greasy human diet) for nearly four months, to make them fat and diabetic. The authors then gave these metabolically challenged animals an oral dose of DLPC (100 milligrams per kilogram of body weight per day) for another three weeks.

The physiological effects of DLPC were truly impressive, reversing the metabolic problems commonly observed with a high-fat diet and obesity. Compared with control animals, glucose homeostasis and insulin signalling were substantially improved, and the histological hallmarks of fatty liver significantly diminished.

This paper[2] firmly establishes the beneficial actions of DLPC. But how does this trace component of lecithin perform its metabolic magic? The authors posit that DLPC functions as a natural ligand for LRH-1. In favour of this hypothesis, they discovered that almost all of DLPC's therapeutic benefits disappear when LRH-1 is genetically ablated in the liver. Coupled with their cellular data, this observation makes for a compelling case that DLPC is a ligand for LRH-1. But there is ample room for sceptics and alternative models, especially given the extremely high levels of DLPC that Lee and colleagues used in their cellular studies. For example, DLPC might target LRH-1 indirectly, either by triggering a signalling cascade or by serving as a metabolic precursor for the 'real' lipid ligand of LRH-1.

The work also raises other questions, beginning with the basic one of how DLPC enters the cell. Could a dedicated flippase enzyme or a transporter facilitate its entry? And once inside the cell, how does it reach the nucleus? Is it shuttled there by specialized phospholipid-transfer proteins? If DLPC does make its way into the nucleus to bind LRH-1, could existing synthetic ligands[7] mimic its cellular and physiological effects?

Equally perplexing are the selective effects of DLPC compared with DPPC, which share identical head groups and differ only in their acyl-chain length (12 and 16 carbons, respectively; Fig. 1). How does this difference affect these phospholipids' binding affinities and change receptor activation? If we assume, on the basis of structural studies, that the head groups are similarly positioned at the 'mouth' of the ligand-binding pocket, what path do the buried acyl chains take to account for the vastly different activities observed for DLPC and DPPC? Such questions merely underscore both the challenge and the mystery of lipid signalling in the nucleus.

What emerges from Lee and colleagues' work[2] is that, surprisingly, phosphatidylcholines reverse some of the consequences of a high-fat diet in rodents. Consistent with this is the earlier finding[8] that 1,2-dilinoleoyl-*sn*-glycero-3-phosphocholine — the major component in soya bean lecithin — was partially effective in treating fatty liver disease (note that this phosphatidylcholine species is also abbreviated to DLPC but differs from the DLPC used in the current study[2]). And, more recently, it has also been reported[9] that injection of another phosphatidylcholine called POPC into the portal vein decreases hepatic steatosis in mice, possibly by binding and activating another nuclear receptor, PPARα. On the basis of this evidence[2,8,9], dietary phosphatidylcholines could certainly offer a new option for treating human

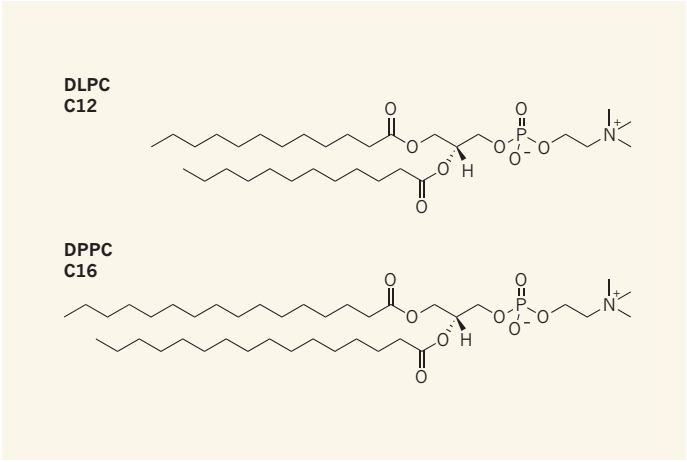


Figure 1 | Length matters. Lee *et al.*[2] investigated the effects of two phosphatidylcholine species, DLPC and DPPC, on activation of the nuclear receptor LRH-1. They found that only DLPC could activate LRH-1 and lead to improved glucose homeostasis in fat, diabetic mice. The two phospholipids differ only in the length of their fatty acyl chains (12 compared with 16 carbons), which could account for their different binding affinities to LRH-1.

50 Years Ago

The Borneo earless monitor lizard (which forms, with two American lizards, the family Helodermatidae) is known from less than ten specimens … A live specimen measuring 13 in. (about average size to date) was obtained only a mile from our own archaeological base camp … In most of its behaviour it resembled a nocturnal snake. Though taken from a hole in the ground, the front legs are so weak that it is difficult to conceive of its burrowing with these. The strong snout and head were used to enlarge any ground weakness, however … It showed no inclination to bite either the handler or anything else (including food). It seems unlikely, therefore, that it is poisonous as has often been suggested.
From *Nature* 24 June 1961

100 Years Ago

Britain's Birds and their Nests — Another gorgeous volume on Britain's birds and their nests! … Happy the publishers, and authors we presume, supported by a public with so insatiable an appetite for British ornithology … We must, however, confess to considerable disappointment in the volume before us. The text is excellent. Indeed, the various biographies are pleasantly written … But it is with the plates that fault is chiefly to be found. They are all "very pretty," but we have more of art than of nature in them. Without exception the species … depicted are the most "proper" series of British birds we have ever made the acquaintance of. They never foul the ground, when 'tis their nature to; they never disturb a blade of grass or a single petal of the beautiful flowers that emborder their nests in nearly every case. They are indeed the most aesthetic company we have yet met with, in the choice of nesting sites.
From *Nature* 22 June 1911

metabolic disorders. But regardless of whether phospholipids can mitigate decades of bad eating habits, this study illustrates a potentially powerful role for phospholipid signalling in the nucleus. ■

Holly A. Ingraham *is in the Department of Cellular and Molecular Pharmacology, University of California, San Francisco, San Francisco, California 94143-2611, USA.*
e-mail: holly.ingraham@ucsf.edu

1. Irvine, R. F. *Nature Rev. Mol. Cell Biol.* **4,** 349–361 (2003).
2. Lee, J. M. *et al. Nature* **474,** 506–510 (2011).
3. Krylova, I. N. *et al. Cell* **120,** 343–355 (2005).
4. Ortlund, E. A. *et al. Nature Struct. Mol. Biol.* **12,** 357–363 (2005).
5. Lee, Y.-K. & Moore, D. D. *Front. Biosci.* **13,** 5950–5958 (2008).
6. Sablin, E. P. *et al. Mol. Endocrinol.* **23,** 25–34 (2009).
7. Whitby, R. J. *et al. J. Med. Chem.* **54,** 2266–2281 (2011).
8. Lieber, C. S. *et al. Nutr. Res.* **27,** 565–573 (2007).
9. Chakravarthy, M. V. *et al. Cell* **138,** 476–488 (2009).

QUANTUM PHYSICS

Correlations without parts

Quantum correlations between the parts of composite systems have long fascinated physicists. There is now compelling evidence that such correlations can also occur in systems in which no parts can be identified. SEE LETTER P.490

ADÁN CABELLO

Quantum mechanics is arguably the most accurate and successful theory in the history of science. But unlike the case for special relativity, for which two physical principles suffice to derive the whole theory, physicists are still seeking the entire set of underlying principles for quantum mechanics. Recently[1,2], they have been trying to understand one of the most intriguing predictions of quantum mechanics: that quantum correlations violate mathematical relationships known as Bell inequalities, which are valid for any local realistic (classical) theory, but that they do so only up to a certain value, whereas more general theories allow violations up to greater values. On page 490 of this issue, Lapkiewicz *et al.*[3] describe an experiment suggesting that a wider perspective, beyond Bell inequalities, is needed to understand why quantum correlations can attain only certain values.

In Bell-inequality experiments (Fig. 1a), tests are performed on two widely separated parts of a composite system. The experimenters then extract the correlations between the outcomes of each of several pairs of tests. In any theory in which the outcomes of these tests are pre-established, the sum of these correlations cannot take a value beyond a certain upper limit. However, quantum mechanics predicts greater values.

In Bell-inequality experiments, the physical separation between the tests has a crucial role: if it is large enough, then the decision of what test is performed in one location cannot influence the outcome of the test performed in the other location, unless there is an instantaneous influence of the two tests on each other. If the outcomes were pre-established, then instantaneous influences would be required to explain quantum correlations. But this is too high a price to pay, because it is impossible to fit instantaneous influences into any theory in which such influences travel at a finite speed.

Quantum correlations have been experimentally observed in tests that are separated widely enough to prevent any influence that travels at the speed of light[4] (Fig. 1a). However, they have been found to have the same values whether the distance between the two experiments is one metre[5] or a few micrometres[6]. What's more, quantum correlations display the same values when two compatible tests are performed on a single system[7] (Fig. 1b, c). Therefore, although distance makes quantum correlations more fascinating, it apparently plays no part in the values that quantum correlations can attain.

Why should one care about quantum correlations between compatible sequential tests on the same physical system instead of about Bell experiments? There are two reasons. The first is that, to violate a Bell inequality, a particular type of quantum state is needed; these are called entangled states and cannot be prepared by local operations and classical communication. This might suggest that composite systems and entangled states are essential for quantum correlations. However, before Bell inequalities were introduced, Kochen and Specker[8] noticed that quantum mechanics is in conflict with classical physics even for non-composite systems. This conflict can be converted into experimentally testable violations of classical correlation inequalities[9] and

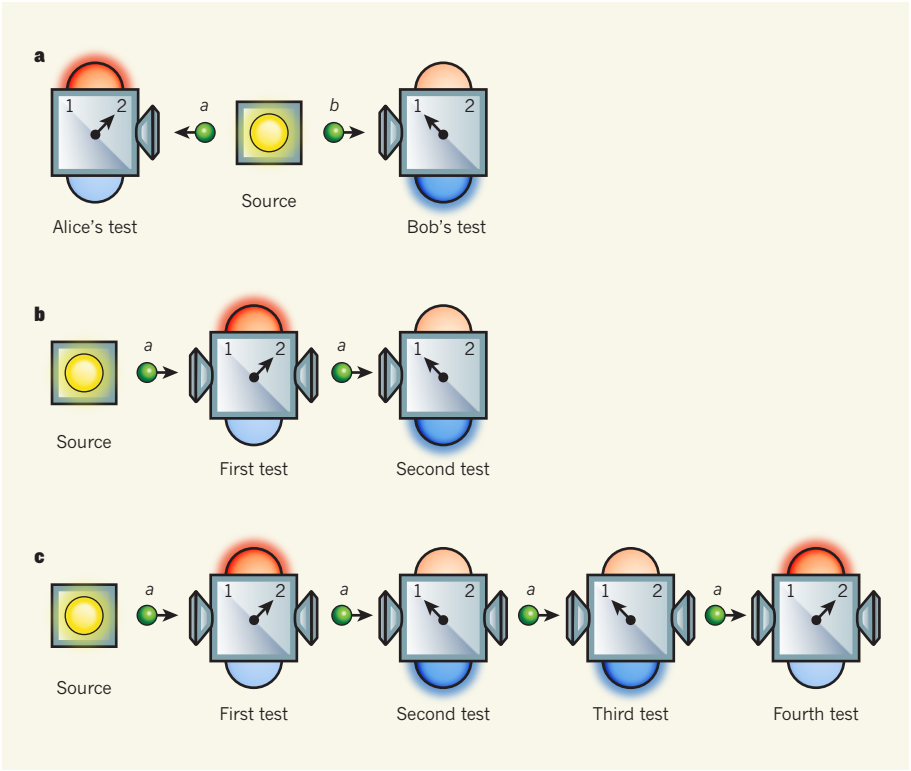


Figure 1 | Comparing separate and sequential tests. **a**, In a Bell experiment, a source emits a pair of particles (*a* and *b*), and an observer (Alice) performs one of two possible measurements (1 or 2) on particle *a*. The measurement has two potential outcomes (either the red light or the blue light flashes). Similarly, a second observer (Bob) performs one of the two measurements on particle *b*. In this example, the red light flashes as a result of measurement 2 on particle *a* and the blue light flashes as a result of measurement 1 on particle *b*. **b**, In an experiment involving sequential compatible measurements, such as that performed by Lapkiewicz and colleagues[3], a source emits particle *a* on which compatible measurements 2 and 1 are performed sequentially. **c**, Measurements 1 and 2 are compatible when, for each particle prepared by any source, each measurement always gives the same outcome, no matter how many times the measurements are performed or in which order.

into experiments[10,11] showing that quantum correlations occur for any quantum state — not necessarily just for entangled ones.

The second reason is the one that makes Lapkiewicz and colleagues' experiment[3] special. Whereas all previous experiments were performed on systems in which two parts can be defined, the work of Kochen and Specker suggests[8] that quantum correlations should occur even in simpler systems, in which no parts can be defined. They identified[8] a physical system in which three states can be distinguished (a 'qutrit') as the simplest one in which the predictions of quantum mechanics clash with those of theories in which unperformed experiments have pre-established outcomes. The authors' experiment[3] provides compelling evidence for quantum correlations in just such a system.

The experiment[3] is conceptually simple: a photon that can travel along three different paths is subjected to several pairs of compatible measurements (such as in Fig. 1b). If the results of these measurements were pre-established and were independent of the compatible measurements, then the correlations would not exceed a certain number. However, the experiment shows a clear violation of this limit, in agreement with the predictions of quantum mechanics.

Lapkiewicz and colleagues' results can still be explained using 'contextual' models, in which the outcome of one measurement depends on the previous (compatible) measurement. But there is no difficulty in converting quantum correlations produced in sequential compatible tests on single systems into correlations between separated systems in which contextual models become 'non-local'[12].

The authors' findings are therefore of fundamental importance, because they confirm that quantum correlations also occur in systems in which entanglement, which is supposed to be the most emblematic feature of quantum mechanics, cannot be defined. It seems that Bell experiments, composite systems and entangled states are not enough to provide a complete understanding of the physical principles behind quantum mechanics: quantum correlations exist without them. ■

Adán Cabello *is at the University of Seville, E-41012 Seville, Spain. He is also at*

Stockholm University, Sweden.
e-mail: adan@us.es

1. Pawłowski, M. *et al. Nature* **461,** 1101–1104 (2009).
2. Oppenheim, J. & Wehner, S. *Science* **330,** 1072–1074 (2010).
3. Lapkiewicz, R. *et al. Nature* **474,** 490–493 (2011).
4. Weihs, G. *et al. Phys. Rev. Lett.* **81,** 5039–5043 (1998).
5. Matsukevich, D. N. *et al. Phys. Rev. Lett.* **100,** 150404 (2008).
6. Rowe, M. A. *et al. Nature* **409,** 791–794 (2001).
7. Hasegawa, Y. *et al. Nature* **425,** 45–48 (2003).
8. Kochen, S. & Specker, E. P. *J. Math. Mech.* **17,** 59–87 (1967).
9. Cabello, A. *Phys. Rev. Lett.* **101,** 210401 (2008).
10. Kirchmair, G. *et al. Nature* **460,** 494–497 (2009).
11. Amselem, E. *et al. Phys. Rev. Lett.* **103,** 160405 (2009).
12. Cabello, A. *Phys. Rev. Lett.* **104,** 220401 (2010).

AERODYNAMICS

The cost of flight in flocks

There are well-known aerodynamic and energetic benefits to flying in an orderly formation. By contrast, it seems that the flocking flight seen in pigeons is metabolically expensive. So why do they do it? SEE LETTER P.494

GEOFFREY SPEDDING

Formation flight has long been known to confer aerodynamic advantages on appropriately spaced fixed-wing aircraft. Flying with a wing positioned in an updraft is a little like finding a free source of lift, which, in turn, reduces drag. Drag is directly related to fuel consumption, so formation flight in birds is seen as a way for these creatures to increase their migratory range or cut the costs of general commuting. All a bird must do to reap the rewards of formation flight is stay in formation. The potential benefits of the V-formation[1] or of certain more complex clusters[2] have been noted in idealized mathematical models. However, many bird flocks apparently lack the order and precision required to make such energy savings, and it is far from obvious how to formulate a tractable theoretical model for such complex patterns.

On page 494 of this issue, Usherwood *et al.*[3] describe how they made the first measurements of body accelerations in individual birds involved in voluntary, loosely formed flocking flights. The reasonable inference from the assembled data is that such flights do not save energy, but rather come at a cost. Energy saving is not of overriding importance in such flight excursions, and the flocks must form for other reasons.

Forty years ago, Lissaman and Shollenberger[1] pointed out that the aerodynamic advantages of formation flight could be especially accessible to birds: local wing twist and wing flexibility allow these animals to configure their aerodynamic profile according to the local air-flow field. The positioning accuracy required seemed reasonable, and the stable and preferred shape of V-formations was explained as the best configuration for evening out the drag distribution in a flock. Planar V-shaped formations, as observed in migrating geese for example (Fig. 1), could increase migratory range by as much as 70%; similar energetic advantages have been proposed for fish schooling[4]. And the potential cost savings in full-scale aircraft[5], and in fleets or swarms of unmanned autonomous vehicles in the air or underwater, are topics of renewed interest.

Noting that bird flocks are not always in neat, linear arrays, Higdon and Corrsin[2] analysed a more general cluster formation. In contrast to Lissaman and Shollenberger[1], they ignored details of the air-flow distribution on the wing, and replaced each bird with a mathematically convenient function, with almost identical far-field properties. They showed that, in three-dimensional flocks, drag savings could be either positive or negative, depending on the spanwise or vertical positions of the flock members. Their tentative conclusion was that "improved flight efficiency is not an important reason for migration in large, three-dimensional flocks".

There are many possible reasons for flying in a flock, which may include mutual observation, collective guidance and navigation, enhanced security as a result of greater numbers of individuals or of eyes, fitness display, and assessment of group numbers. Energy saving may be of paramount, or little, importance. Even if energy saving is not an explicit goal, then avoiding excessive energy

N. TAYLOR WILDLIFE/ALAMY

S.-E. ARNDT/PHOTOLIBRARY.COM

Figure 1 | Flight formations and clusters. Canada geese migrate in a characteristic V-formation (left); such an orderly, planar arrangement can reduce drag, resulting in energy savings. Complex swirls and flocks of organisms, such as those of pigeons (right), have less apparent order, and in their research with pigeons Usherwood *et al.*[3] find that flocking flight patterns are energetically costly. Group travel in flocks (birds), schools (fish) and herds (large vertebrates) is common, but there are probably several, often overlapping, reasons for such behaviours.

cost may at least be a consideration.

Usherwood *et al.*[3] measured the wing-beat frequency and body accelerations of 18 trained racing pigeons when they left their home loft in voluntary excursions, which involved quite irregular clusters with varying densities and flight paths. Quite often the cluster would circulate in a tight circle or spiral. Backpacks containing Global Positioning System equipment relayed data back at rates sufficient to correlate wing-beat and body accelerations with flock position and density.

Several interesting observations followed. First, sharp turning manoeuvres, with centrifugal accelerations comparable to gravitational acceleration, are themselves costly. Second, pigeons flap their wings faster when in a cluster than when flying alone. Third, the flapping frequency correlates strongly with the proximity of neighbouring birds. The authors[3] argue that the average aerodynamic downdrafts are probably comparatively small, and propose that the high-frequency flapping is more likely to be an adaptation to increased demands on flight control and collision avoidance. Regardless of the cause, because the flapping frequency can be very roughly used as a surrogate for power consumption, the implication is that flying in such a flock is more costly than flying alone.

The dynamics (social and physical) of flying in flocks is not easy to simplify. Even in seemingly orderly flocks of pelicans or geese, the measured precision in wingtip–wingtip spacing is often quite far from the mathematical ideal[6,7]. This study[3], like most others, ignores the effect of the complex wake disturbances that are undoubtedly generated by each pair of flapping wings. Yet the empirical evidence suggests that, because energetic savings are negative, in this instance we may have to search elsewhere for the reasons for flying in flocks. Perhaps the episodic flights of racing pigeons allow the birds to test and exercise their locomotory and control machinery. As with many problems in biology, it is quite possible that more than one reason conspires to create any given bird flock. ■

Geoffrey Spedding *is in the Department of Aerospace and Mechanical Engineering, University of Southern California, Los Angeles, California 90089-1191, USA.*
e-mail: geoff@usc.edu

1. Lissaman, P. B. S. & Shollenberger, C. A. *Science* **168,** 1003–1005 (1970).
2. Higdon, J. J. L. & Corrsin, S. *Am. Nat.* **112,** 727–744 (1978).
3. Usherwood, J. R., Stavrou, M., Lowe, J. C., Roskilly, K. & Wilson, A. M. *Nature* **474,** 494–497 (2011).
4. Weihs, D. *Nature* **241,** 290–291 (1973).
5. Ray, R. J., Cobleigh, B. R., Vachon, M. J. & St John, C. AIAA Flight Mechanics Conf. Exhibit, Monterey 2002–4492 (2002).
6. Hainsworth, F. R. *J. Exp. Biol.* **135,** 431–444 (1988).
7. Cutts, C. J. & Speakman, J. R. *J. Exp. Biol.* **189,** 251–261 (1994).

ORGANIC CHEMISTRY

Triumph for unnatural synthesis

Nature crafts many molecules from common precursors, but this approach isn't always possible in chemical synthesis. A strategy for synthesizing a family of natural products succeeds by ignoring nature's blueprint. SEE ARTICLE P.461

STÉPHANE QUIDEAU

Polyphenols are a group of structurally diverse compounds found in fruits, vegetables and plant-based food products such as tea, wine and chocolate. Because many polyphenols are antioxidants, they have been acclaimed as natural health-protecting agents, although the benefits to humans have yet to be proven. Nevertheless, some polyphenols have biological activities that make them potentially useful leads in the search for drugs against illnesses such as heart disease, cancer and Alzheimer's disease. Pharmaceutical research has been thwarted, however, by the fact that the most complex polyphenols are available from their natural sources in only limited quantities.

An obvious solution would be to prepare large quantities of polyphenols using chemical synthesis. This might seem a trivial task, given that many polyphenols are oligomers derived from just one or two precursor molecules, of which resveratrol (Fig. 1) is perhaps the best-known example. In fact, the structural complexity of resveratrol oligomers makes their chemical synthesis a daunting challenge. But on page 461 of this issue, Snyder *et al.*[1] describe a major advance in polyphenol research: a daring but clever synthetic strategy that has enabled them to prepare a series of resveratrol trimers and tetramers — the highest-order resveratrol oligomers prepared to date.

Traditionally, strategies for total synthesis target one specific, naturally occurring compound, rather than a series of biosynthetically or structurally related compounds. Furthermore, the large number of steps involved in many classical total syntheses often makes the synthetic routes economically unviable for industrial-scale processes. Synthetic organic chemists are therefore now using their *savoir faire* to devise short, practical natural-product syntheses, with the additional challenge of finding routes that have minimal impact on the environment. One approach is to identify better ways of exploiting the inherent chemical reactivity of starting materials and/or synthetic intermediates[2]. This is the approach used by Snyder and colleagues[1], although the resveratrol oligomers that they have made are not traditional targets for total synthesis — indeed, only a few chemists have attempted to make complex plant polyphenols[3,4].

Although most polyphenols are biosynthetically derived from the metabolism of only one or two parent molecules, the structural diversity generated within each class of polyphenol is enormous[5]. A couple of hundred oligomeric constructs are known to be derived from resveratrol, for example[6]. The biosynthesis of these compounds presumably involves the initial dimerization of resveratrol (which can produce several structurally different dimers), structural rearrangement of the dimers and then further transformations to make higher resveratrol oligomers. But biomimetic strategies that involve making resveratrol oligomers by treating resveratrol with chemical or enzymatic oxidants have generally produced low yields of the desired products and/or led uncontrollably to complex mixtures of compounds.

Nevertheless, a few dimeric members of the resveratrol oligomer family have been made using molecular building blocks other than resveratrol. But none of those syntheses, elegant though they may be[7], provides a common route that could generate multiple, structurally very different members of the family. This is what Snyder *et al.*[1] have achieved.

Snyder's group previously reported[8,9] the synthesis of several different resveratrol dimers from a common building block that is distinct from, and much more chemically controllable than, resveratrol (Fig. 1). To reach the next level of complexity[1] (trimers and tetramers), the authors decided to try to attach bromine atoms to specific sites in some of the previously prepared dimeric compounds. Once installed, the bromine atoms could be used as 'handles' to introduce resveratrol-based groups known as *trans*-dihydrofurans, making trimers (by the addition of one *trans*-dihydrofuran) or tetramers (by adding two *trans*-dihydrofurans). This ambitious goal required the means not only to differentiate selected sites for bromination from all the other sites that shouldn't be brominated, but also to selectively brominate different positions in compounds at will, as required for the particular trimer or tetramer being targeted.

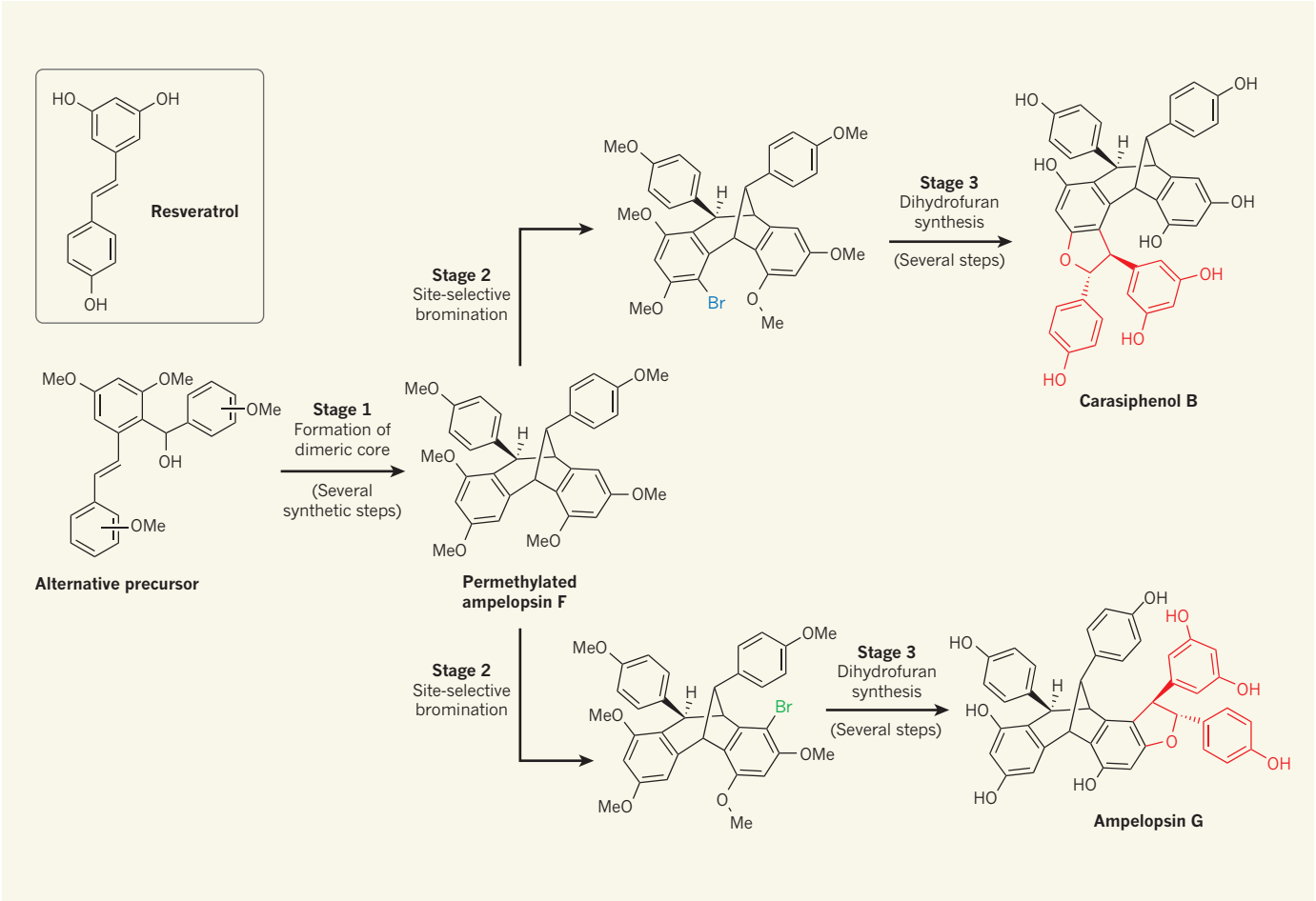


Figure 1 | A synthetic strategy for making resveratrol-derived polyphenols. Nature constructs a family of oligomeric polyphenol compounds from resveratrol, but this approach does not work well for the chemical synthesis of these compounds. Snyder *et al.*[1] report a three-stage strategy for making resveratrol oligomers that starts from an alternative precursor molecule. In the first stage, the precursor is converted into a resveratrol dimer, such as permethylated ampelopsin F. The second stage involves introducing a bromine atom (Br) at selected positions in the dimer; the two examples shown were achieved by using different bromination agents. The third stage is to introduce new resveratrol units in the form of *trans*-dihydrofurans (red), as shown here to make the resveratrol trimers carasiphenol B and ampelopsin G. The authors prepared other trimers and tetramers in a similar way using different resveratrol dimers.

By choosing the appropriate substrates and bromination agents (a process that required extensive experimental trials), Snyder *et al.* successfully synthesized three naturally occurring resveratrol trimers and two tetramers in good yields. For example, when the authors reacted a commonly used bromination agent, *N*-bromosuccinimide, with permethylated ampelopsin F (a resveratrol dimer), they obtained a product in which a single bromine atom had become attached to the dimer (Fig. 1) — a remarkable outcome, given that this was one of four possible products. They then converted this product into a trimeric compound, carasiphenol B. But when they reacted permethylated ampelopsin F with a different brominating agent[10], they obtained a product in which the bromine atom was attached to a different site in the dimer. The authors converted this into another resveratrol trimer, ampelopsin G.

Snyder and colleagues' syntheses yield products as racemic mixtures (one-to-one mixtures of mirror-image isomers), which would need to be separated into individual isomers for some applications. Even so, their work constitutes a spectacular, highly efficient and general solution to the chemical synthesis of many — if not all — resveratrol-derived polyphenols. This approach will also be useful for preparing analogues of the natural products, which could be necessary for finding compounds possessing optimal biological properties in drug development programmes. The structural diversity of resveratrol oligomers that can be prepared controllably from a simple and easily modulated starting material is truly remarkable, far surpassing what can be achieved using biosynthetic or biomimetic approaches starting from resveratrol itself. Moreover, the authors' strategy of performing divergent, site-selective reactions using lower-order oligomers might be applicable to the synthesis of other oligomeric families of natural products[11]. ■

Stéphane Quideau *is at the Institut des Sciences Moléculaires (CNRS-UMR 5255), Université de Bordeaux, Institut Européen de Chimie et Biologie, 33607 Pessac Cedex, France.*
e-mail: s.quideau@iecb.u-bordeaux.fr

1. Snyder, S. A., Gollner, A. & Chiriac, M. I. *Nature* **474,** 461–466 (2011).
2. Shenvi, R. A., O'Malley, D. P. & Baran, P. S. *Acc. Chem. Res.* **42,** 530–541 (2009).
3. Feldman, K. S. & Lawlor, M. D. *J. Am. Chem. Soc.* **122,** 7396–7397 (2000).
4. Ohmori, K., Shono, T., Hatakoshi, Y., Yano, T. & Suzuki, K. *Angew. Chem. Int. Edn* **50,** 4862–4867 (2011).
5. Haslam, E. *Practical Polyphenolics: From Structure to Molecular Recognition and Physiological Action* (Cambridge Univ. Press, 1998).
6. Lin, M. & Yao, C.-S. in *Studies in Natural Products Chemistry* Vol. 33 (ed. Atta-ur-Rahman) 601–644 (Elsevier, 2006).
7. Nicolaou, K. C., Wu, T. R., Kang, Q. & Chen, D. Y.-K. *Angew. Chem. Int. Edn* **48,** 3440–3443 (2009).
8. Snyder, S. A., Zografos, A. L. & Lin, Y. *Angew. Chem. Int. Edn* **46,** 8186–8191 (2007).
9. Snyder, S. A., Breazzano, S. P., Ross, A. G., Lin, Y. & Zografos, A. L. *J. Am. Chem. Soc.* **131,** 1753–1765 (2009).
10. Snyder, S. A. & Treitler, D. S. *Angew. Chem. Int. Edn* **48,** 7899–7903 (2009).
11. Snyder, S. A., ElSohly, A. M. & Kontes, F. *Nat. Prod. Rep.* **28,** 897–924 (2011).

ARTICLE

doi:10.1038/nature10197

Regioselective reactions for programmable resveratrol oligomer synthesis

Scott A. Snyder[1], Andreas Gollner[1] & Maria I. Chiriac[1]

Although much attention has been devoted to resveratrol, a unique polyphenol produced by plants and credited as potentially being responsible for the 'French paradox'—the observation that French people have a relatively low incidence of coronary heart disease, even though their diet is high in saturated fats—the oligomers of resveratrol have been largely ignored despite their high biological activity. Challenges in achieving their isolation in sufficient quantity from natural sources, coupled with an inability to prepare them easily synthetically, are seen as the main obstacles. Here we report a programmable, controlled and potentially scalable synthesis of the resveratrol family via a three-stage design. The synthetic approach requires strategy- and reagent-guided chemical functionalizations to differentiate two distinct cores possessing multiple sites with the same or similar reactivity, ultimately leading to five higher-order natural products. This work demonstrates that challenging, positionally selective functionalizations of complex materials are possible where biosynthetic studies have indicated otherwise, it provides materials and tools with which to unlock the full biochemical potential of this family of natural products, and it affords an intellectual framework within which other oligomeric families could potentially be accessed.

Resveratrol is to oligostilbenes what isoprene is to terpenes. Whether through highly evolved enzymatic pathways or diversification opportunities through uncontrolled reactions or both, these starting points lead to hundreds of unique frameworks (such as **1**–**3** and **5**–**7**, in Fig. 1) with wide-ranging biological properties relevant to the treatment of cancer, AIDS and bacterial infections, among other diseases[1–4]. For laboratory synthesis, although strategies, tactics and logic exist to access most terpenes controllably and thereby facilitate full biochemical and pharmaceutical explorations[3], the same cannot be said of resveratrol-derived oligostilbenes[5,6].

These unique materials are produced by plants throughout the world as phytoalexins, serving as a first line of chemical defence against stresses such as fungal infections. That role may derive from the fact that these molecules, including **8**–**20** (Fig. 2), can be formed readily from resveratrol (**4**) through its self-merger two or more times through numerous distinct C–C and C–O bond constructions. Indeed, to put that bond-forming diversity into perspective, most other similarly produced oligomer families have just one or two specific combination modes[6]. Although the biogenesis of this family is still unknown, a number of studies[4,7,8] point to a reasonable hypothesis based on two main phases of construction, stages that could be viewed as similar to the cyclase and oxidase paradigm of terpene biosynthesis[2,3].

In the first, radical and cation-based reaction cascades convert resveratrol (**4**) into multiple dimeric frameworks, including **8**–**15**. These cores may result either through direct dimerization of **4** or rearrangement reactions of a given dimer (especially ε-viniferin, **8**) into additional frameworks under appropriate conditions to create an initial base of architectural and stereochemical diversity[7,8]. Whether enzymes are involved in these processes is unknown, though the

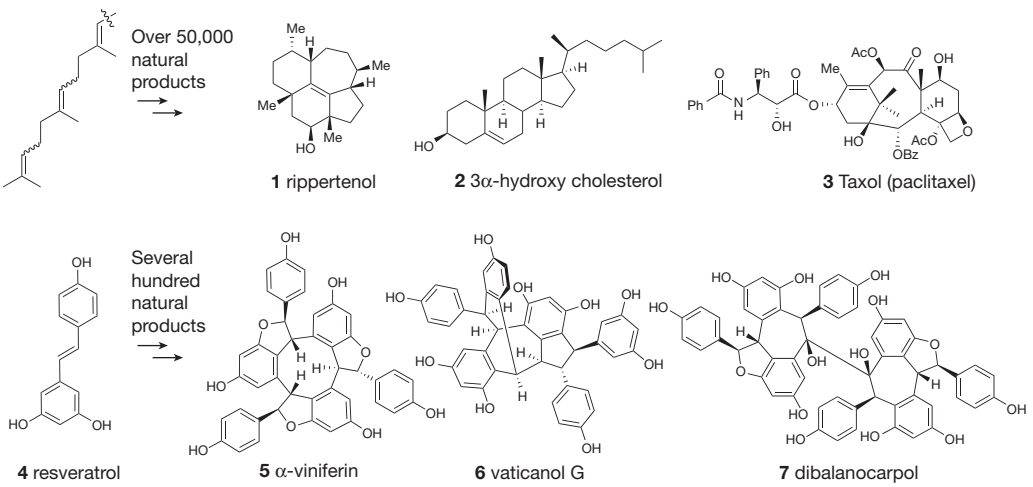


Figure 1 | The diversity of selected terpene and polyphenolic oligostilbene natural products: products of privileged starting materials. Me, methyl; Ph, phenyl; Ac, acetate; Bz, benzoate.

[1]Department of Chemistry, Columbia University, Havemeyer Hall, 3000 Broadway, New York, New York 10027, USA.

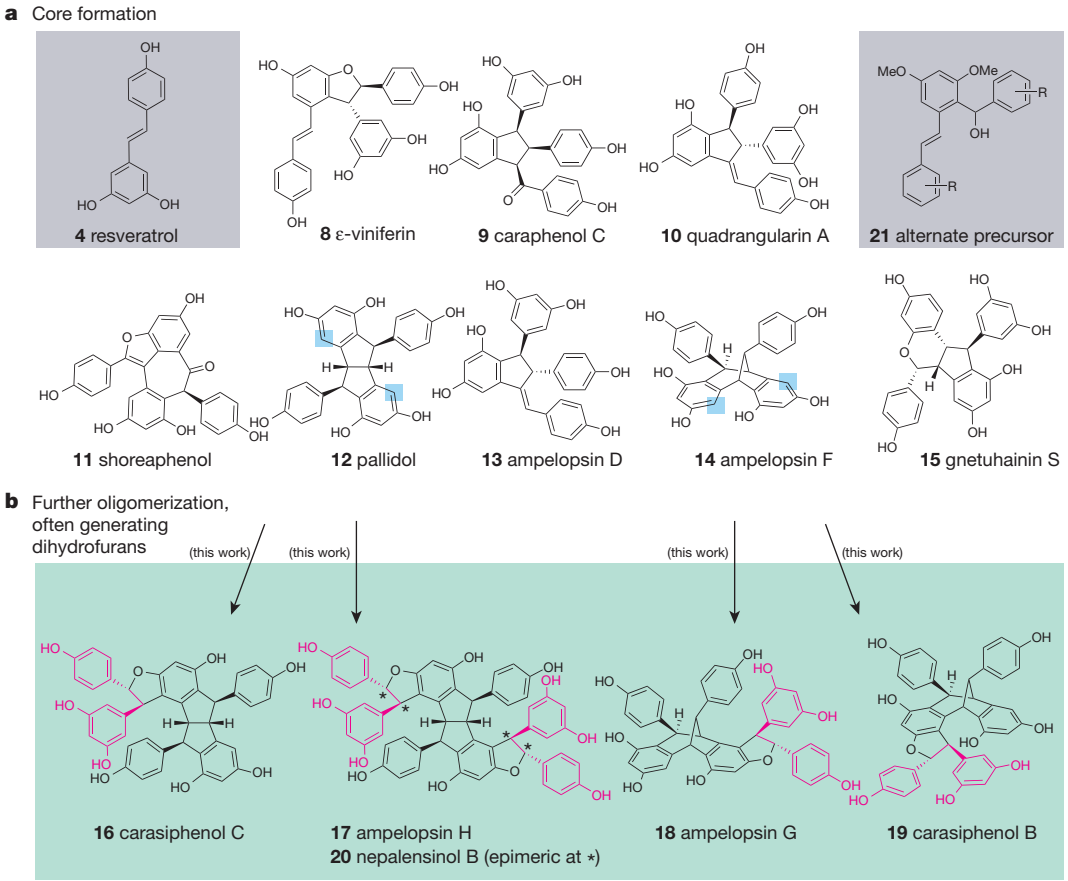


Figure 2 | Nature's putative biogenesis of the resveratrol family and our specific plan for achieving their controlled assembly. **a**, Core formation (stage 1 of nature's synthesis). **b**, Further oligomerization, often generating dihydrofurans (stage 2 of nature's synthesis, ring systems added shown in pink). In the laboratory, we anticipated that three distinct goals (steps) would be needed to forge these materials in a controlled fashion. The first was the use of a unique starting material (**21**) well removed from resveratrol itself to generate the dimeric cores as shown in **a**. We then required regioselective functionalizations that could differentiate the highlighted positions on a symmetric and non-symmetric core (goal 2, blue boxes) and a dihydrofuran formation protocol with divergent stereocontrol (goal 3). These final challenges are the subject of this work (green shading).

participation of some chiral entity at some stage seems reasonable given that many of these materials are isolated in an optically active form.

From these staging areas, further oligomerization can occur in a second synthesis stage, with one seemingly prevalent mode being the addition of new resveratrol units to generate dihydrofuran rings. As indicated by the representative structures shown for compounds **16**–**20** (refs 9–13) derived from the pallidol (**12**) and ampelopsin F (**14**) cores, these new ring systems are added with inconsistent regio- and stereospecificity. Although we do not know the degree to which a given plant species can dictate the synthetic distribution of these variants (and whether enzymes are involved), from a global perspective the possibility of concurrently garnering such diversity is probably advantageous because it affords a broader range of architectures and potential biochemical properties. Indeed, initial screens have shown that activity is correlated with both size and stereochemistry. For example, although resveratrol (**4**) has broad-spectrum activity in mouse models (and is implicated in the 'French paradox' in humans)[14,15], the addition of increasing numbers of resveratrol units produces compounds with greater specificity and potency[16–18]. Similarly, although ampelopsin H (**17**) and nepalensinol B (**20**) have the same cores, their stereochemical differences lead to different mechanisms of anti-cancer activity[19,20].

Given this general mode and purpose of synthesis, these materials are ideal compound collections for biochemical screening if they could be obtained in quantity. Unfortunately, achieving their isolation from natural sources even in minute amounts is challenging and labour-intensive, as is modulating their multiple reactive groups in a laboratory flask. Indeed, biosynthetic explorations based on exposing resveratrol (**4**) or varied protected forms to a single chemical or oxidase usually lead to complex product mixtures because the chemist cannot differentiate, at will, between the dozens of reaction pathways available, which is necessary to prepare a single structure[21–26]. In fact, of those materials that can be isolated and characterized from such studies, most (and sometimes all) are non-natural; those of natural origin make up only a small percentage of the total family. The same outcome occurs if higher-order materials are exposed to resveratrol (**4**) under similar conditions[27]. By contrast, wholly stepwise, non-biomimetic sequences empowered by retrosynthetic analysis can produce single dimeric members with control, but do not possess the strategic power to deliver a family-level solution for all frameworks[5,6,28–31]. Here we establish an approach with the potential to lead to the synthesis of the entire resveratrol collection controllably, one member at a time.

Programmable synthesis design

We anticipated that the clearly diverse, complicated, and potentially chaotic bond constructions of nature's resveratrol oligomer synthesis could be regulated in a laboratory flask if three separate goals proved achievable. The first goal—the rational, systematic, and controlled synthesis of the various dimeric cores—was accomplished by using **21** (Fig. 2) as an alternate starting point. Indeed, we have already converted this material into structures **10**, **12**, **13** and **14**, a reduced form of **8**, analogues of **9**, several natural product variants of **11**, and several additional unique dimeric natural products in both published[32,33] and

unpublished work; some of these syntheses have proceeded on the gram scale. The second and third goals are the subject of this study, tasks for which we initially targeted compounds **16**–**19** to address the challenges anticipated in adding resveratrol units controllably to any core. This process required (1) a means of differentiating the highlighted positions within symmetric and non-symmetric frameworks **12** and **14**, not only from all the other non-highlighted sites but also from each other[34,35], and (2) a robust sequence that could potentially advance a given positionally differentiated material into either *trans*-dihydrofuran diastereoisomer. Both the second and third goals are easily stated, but we expected them to be quite challenging, especially the second, because it requires differentiation that nature may not deploy (or can perform only with enzymes) and that biomimetic approaches have not achieved.

Differentiation of the pallidol core

We began our studies by attempting to functionalize selectively the symmetric, protected pallidol core (**22** in Fig. 3) using electrophilic aromatic substitution with bromine as our preferred tool. To prepare the dihydrofurans relevant to the resveratrol class, this task required selective functionalization of atom H_A over atom H_B, with H_A needing to be replaced selectively both once and twice. Although theoretical calculations—performed with the B3LYP functional in conjunction with the 6-31G(d) basis set—indicated that the atoms labelled H_A and H_B were the most likely to be replaced of all those on the framework, the electron density maps and electrostatic potentials were equivalent within error for H_A and H_B. From an inspection of various three-dimensional models, such as the one shown in the inset to Fig. 3, H_B appeared to be more accessible owing to the proximal nature of the hydrogens on the aliphatic core relative to H_A.

Experimentally, however, H_A proved to be the natural site of electrophilic aromatic substitution with a synthesis of dibromide **23** achieved in 99% yield using two equivalents of *N*-bromosuccinimide (NBS). Despite this selectivity, it proved difficult to obtain the corresponding monobromide even at a yield of 50%, because the resultant material (not shown) readily underwent additional halogenation (generating **23**) even when substoichiometric amounts of electrophilic bromine were used.

We thus needed an alternative strategy to desymmetrize compound **22**, given that the goal appeared to be outside the power of reagent control. A compound obtained previously (**25**, formed from **24** through three consecutive regiospecific brominations and a Friedel–Crafts cyclization)[32] solved the problem because it is non-symmetric by virtue of its aliphatic bromine atom. Consequently, through initial removal of its aryl bromides followed by a site-selective rebromination controlled by the aliphatic bromide, we formed compound **26** in

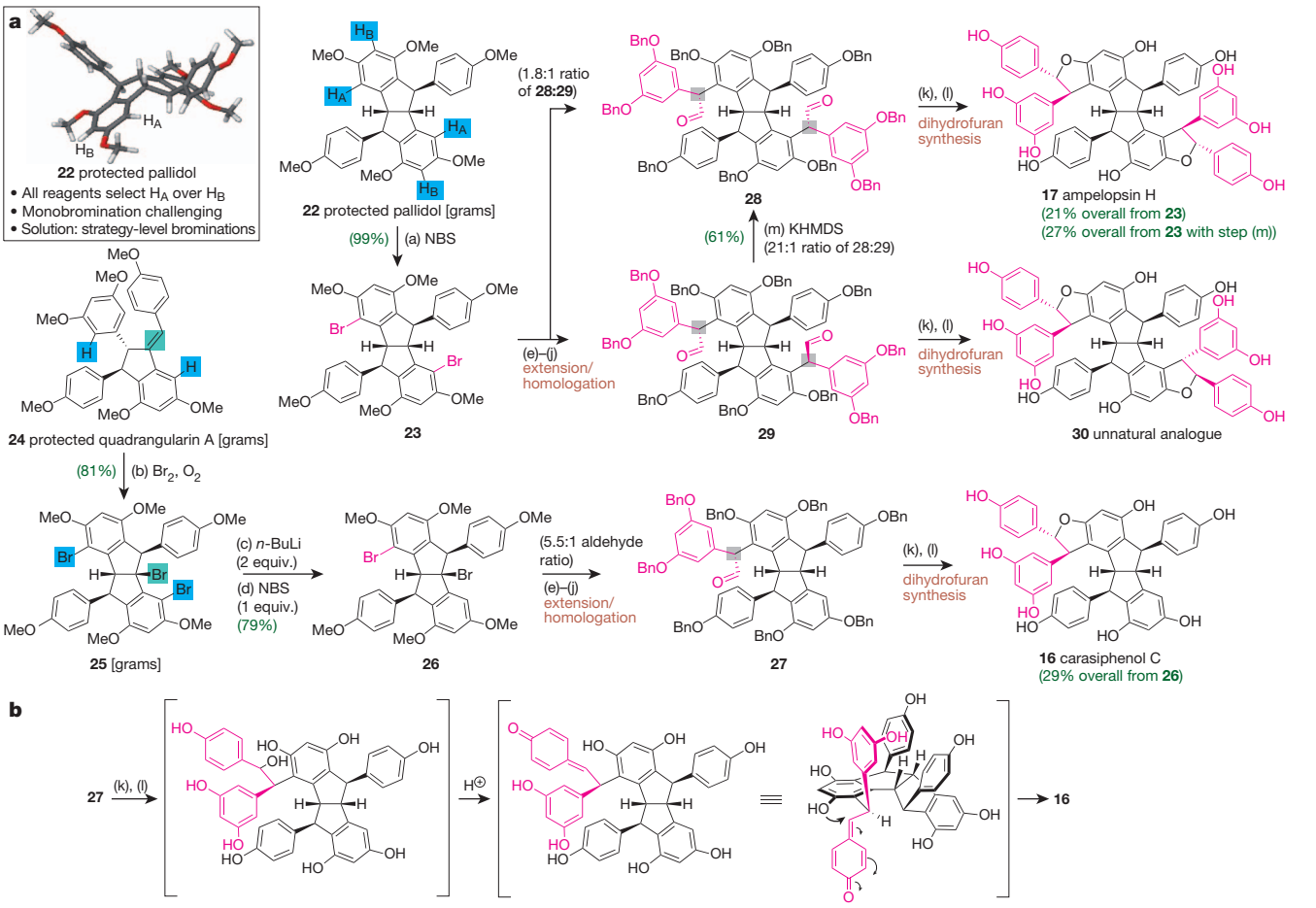


Figure 3 | Use of substrate-guided halogenations to synthesize two resveratrol trimers and tetramers (16 and 17) and an unnatural analogue (30) from protected pallidol (22) and 25. a, Overall synthetic sequence. **b**, Critical details of the final dihydrofuran formation cascade. The inset shows a three-dimensional model of the pallidol core (**22**). Blue shading indicates aromatic positions of **24** that are brominated, green shading indicates an alkene of **24** that is brominated, grey shading indicates new chiral centres generated, and pink shading indicates atoms added synthetically to the generic pallidol core. Reagents and conditions: (a) NBS, THF, −78 °C; (b) Br_2, CH_2Cl_2, −78 °C → 25 °C; (c) *n*-BuLi, THF, −78 °C; (d) NBS, THF, −78 → 25 °C; (e) *n*-BuLi, 3,5-dimethoxybenzaldehyde, THF, −78 °C → 25 °C; (f) Dess–Martin periodinane, $NaHCO_3$, CH_2Cl_2, 25 °C, (g) BBr_3, CH_2Cl_2, 70 °C; (h) BnBr, *n*-Bu_4NI, K_2CO_3, acetone, 70 °C; (i) *n*-BuLi, Me_3SI, THF, 0 °C; (j) ZnI_2, benzene, 25 °C; (k) 4-benzyloxyphenylmagnesium bromide, THF, 25 °C; (l) H_2, 30% Pd/C, EtOAc/MeOH (1:1), 25 °C, then Amberlite IR-12-OH, 25 °C; (m) KHMDS, −78 °C. NBS, *N*-bromosuccinimide; *n*-BuLi, *n*-butyllithium; Bn, benzyl; KHDMS, potassium hexamethyldisilazane. Yields are shown (green text).

79% yield as an effective surrogate of the originally targeted monobrominated congener of **22**.

With this critical stage of selective functionalizations complete, we turned next to dihydrofuran formation. Although many methods exist to forge such rings from aryl bromides[36], our efforts revealed that these approaches do not proceed on the highly electron-rich and sterically encumbered frameworks of the resveratrol class. Nevertheless, we were able to identify a reliable and novel solution to these challenges. These operations began (Fig. 3A) through the initial incorporation of additional 3,5-dimethoxyphenyl ring systems and a protecting group exchange to produce diaryl ketones (see Supplementary Information for structures); the exchange in phenol protecting groups was made at this stage to ensure that mild deprotection conditions could be deployed in the presence of sensitive and highly polar late-stage intermediates.

These steps were then followed by a Corey–Chaykovski epoxidation[37] and Lewis-acid-induced rearrangement (ZnI_2 or $BF_3{\bullet}OEt_2$)[38] to generate aldehydes **27**–**29** from **26** and **23**, respectively. Although these operations collectively seem conventional, they were designed to both respond to, and take advantage of, the uniqueness of the resveratrol system while also ensuring effective diastereocontrol. Indeed, because of the electron-wealth of the neighbouring aromatic rings, the ketone precursors to **27**–**29** could not be engaged by aryl-containing organometallic, Wittig or diazo-forming reagents; they could, however, react with a sulphonium-ylide, and once converted into aldehydes, the adjoining electron-wealth proved invaluable in that it probably prevented **27**–**29** from epimerizing readily, as often occurs with such materials. For example, in the case of **29** only exposure to strong base (potassium hexamethyldisilazane, KHMDS) in tetrahydrofuran (THF) altered that stereochemistry, with this moderately yielding reaction indicating the inherent reprotonation bias of this system.

Thus, with the ability to set these initial stereocentres, the stereodetermining portion of the sequence was complete, given that the remaining operations (Grignard addition and a one-pot deprotection/cyclization sequence through the specific intermediates shown in Fig. 3B for **16**)[39] produced the desired *trans*-disposed dihydrofuran rings of carasiphenol C (**16**), ampelopsin H (**17**) and the non-natural analogue **30**. Crude ^{1}H nuclear magnetic resonance (NMR) spectra revealed that these materials were generated with >10:1 preference over any other compounds, with the very minor, unassigned products potentially being *cis*-dihydrofurans on the basis of a comparison to published NMR coupling constant values. We attribute the overall high level of *trans*-selectivity to a tendency to minimize strain, as shown in the precyclization intermediate in Fig. 3B, although we cannot rule out the possibility of equilibration of any *cis*-disposed products into the probably thermodynamically stable *trans*-materials under the acidic conditions used[40].

Finally, we note that while we have not explicitly attempted a large-scale synthesis of any of these final natural products (most reactions were performed with fewer than 100 mg of material for the sake of convenience, though the dimeric cores are available in gram quantities), we did advance enough material to obtain a 55 mg isolate from the final deprotection/cyclization sequence containing carasiphenol C (**16**) with at least 90% purity (based on ^{1}H and ^{13}C NMR analysis). To put that amount of synthetic material into context, only 10 mg of **16** was obtained from 70 kg of dried plant material in the original isolation effort[9].

Differentiation of the ampelopsin F core

To assess whether these two final stages of synthesis had family-level applicability, we decided next to target natural products **18**, **19** and **38** (Fig. 4) starting from the non-symmetric ampelopsin F core (in form

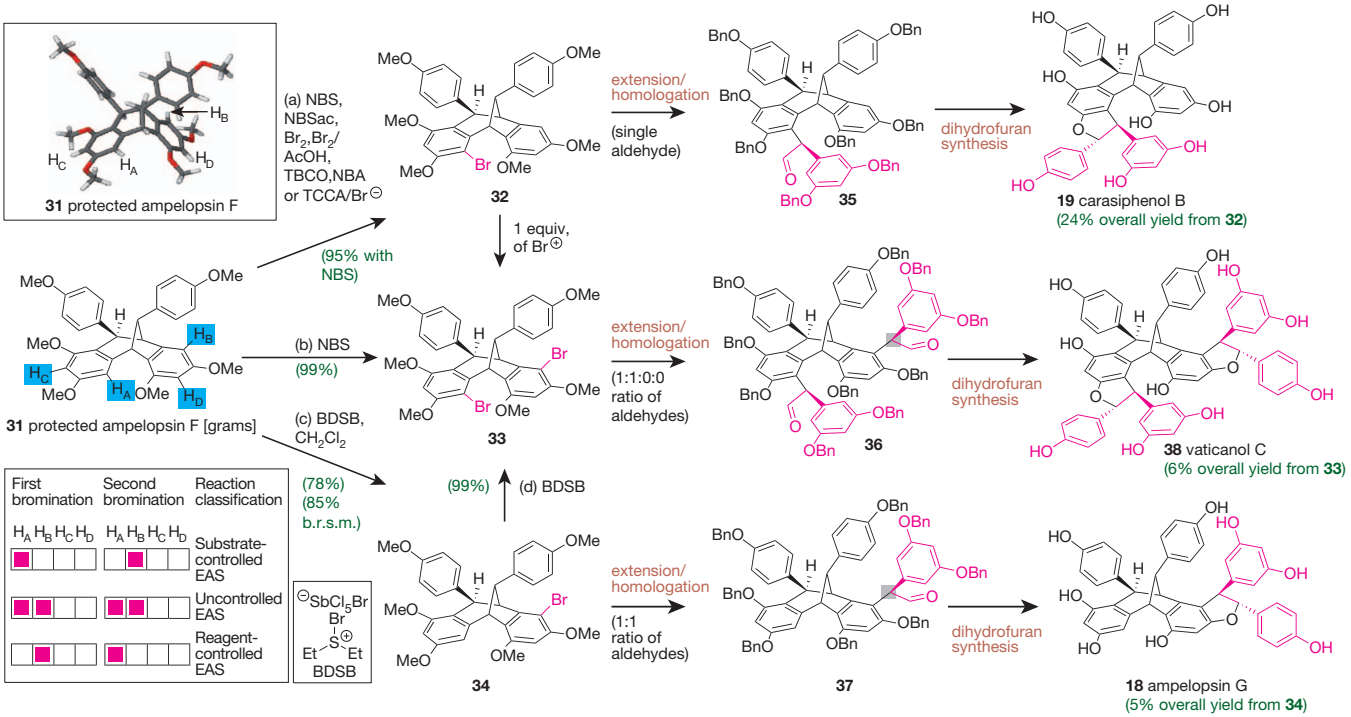


Figure 4 | Use of reagent-guided halogenations to synthesize three resveratrol trimers and tetramers (18, 19 and 38) from protected ampelopsin F (31). The upper inset shows a three-dimensional model of the ampelopsin F core (**31**). The lower inset shows the patterns of halogenation observed with various reagents and their implications for chemical reactivity. EAS, electrophilic aromatic substitution. Blue shading indicates aromatic positions of **31** that could be brominated, grey shading indicates new chiral centres generated, and pink shading indicates atoms added synthetically to the generic ampelopsin F core. Reagents and conditions: (a) NBS, CH_2Cl_2, −78 °C → 25 °C; (b) NBS, CH_2Cl_2, −78 °C → 25 °C; (c) BDSB, CH_2Cl_2, −78 °C; (d) BDSB, CH_2Cl_2, −78 °C. Yields are shown (green text); b.r.s.m., based on recovered starting material; AcOH, acetic acid; NBSac, *N*-bromosaccharin; TBCO, tetrabromocyclohexadienone; NBA, *N*-bromoacetamide; TCCA, trichloroisocyanuric acid; BDSB, structure and polarity as shown.

31). We expected that synthesis of these compounds would present an even greater challenge in positional selectivity because it necessitates the site-selective replacement of positions H_A and H_B (in either order) among four highly electron-rich positions. Physical models and theoretical calculations (using B3LYP/6-31G(d), as described above) revealed no obvious differences between these sites, indicating that all four hydrogens within **31** were equally likely to be replaced.

Experimentation soon revealed, however, that position H_A could readily be functionalized with a variety of bromine sources (including under acidic conditions), producing **32** at a usually high yield (95% with slightly less than 1 equivalent of NBS); use of two equivalents of these reagents smoothly provided dibromide **33** (in 99% yield using NBS). Given the consistency of this outcome with reagents of varying size and reactivity, this functionalization pattern (H_A then H_B) probably reflects substrate control. Thus, accessing **34** (H_B before H_A) required reagent control, a result we felt was potentially achievable given that the use of a collidine/Br_2 complex afforded **34** along with roughly equal amounts of **32** in an uncontrolled electrophilic aromatic substitution reaction.

After much experimentation, we found that a bromonium source we had recently developed[41]—$Et_2SBr{\bullet}SbCl_5Br$ (bromodiethylsulphide bromopentachloroantimonate, BDSB)—was uniquely able to afford **34** cleanly in CH_2Cl_2; the yield for this process was 78% (85% based on recovered starting material) when we performed the reaction on the >0.5 g scale. Only a small amount of **33** was formed from this process, with no **32** discernible by ^{1}H NMR analysis of the crude reaction mixture. To our knowledge, this reaction constitutes the most complex, positionally selective reagent-based bromination described thus far, complementing other catalyst-controlled processes[42]. We are currently investigating the basis for this selectivity, an important issue given that BDSB could potentially provide other uniquely functionalized aromatic building blocks.

Pleasingly, application of the same dihydrofuran formation sequence with these bromides led to diastereocontrolled syntheses of ampelopsin G (**18**), carasiphenol B (**19**), and vaticanol C (**38**); the last of these targets (vaticanol C) is an exciting lead for cancer treatment and prophylaxis[17]. We found that high control in aldehyde stereochemistry was always obtained at position H_A (as for **35**), but was consistently a 1:1 mixture at position H_B, as highlighted in **36** and **37**, suggesting different inherent biases of the system.

Strategically, however, it is important to note that we were also able to advance successfully the undesired aldehyde epimer from the ampelopsin G sequence into the diastereomeric dihydrofuran, synthesizing what is, at present, a non-natural analogue (see Supplementary Information). Thus, though we have not outlined a specific solution to forging these epimeric aldehydes controllably (given that natural products were the specific goal of this study), we believe the robustness of the terminating sequence provides a family-level solution for the clean preparation of any *trans*-dihydrofuran, dependent only on initial control of aldehyde stereochemistry. As such, the potential exists to access resveratrol-based diversity and biochemical reactivity in compounds not yet isolated from nature.

Conclusions and discussions

This work demonstrates, for the first time, that the controlled synthesis of higher-order resveratrol oligomers (and potentially the entire class) can be achieved. Salient features of the design are its reliance on a unique starting point coupled with challenging functionalizations on diverse and complex cores. Indeed, only through the introduction of a novel reagent (BDSB) was appropriate positional control achieved on one framework, while a strategic solution was required on the other framework. As such, these complex and positionally selective approaches indicate that the same and/or similar functionalities can be differentiated even on highly complex frameworks, countering the notion (based on previous biosynthetic explorations) that such processes could not be achieved. Moreover, the developed sequences provide opportunities to optimize biological activity and alter physicochemical properties by making analogues otherwise inaccessible from resveratrol alone.

To put the developed approach for the resveratrol class into strategic context, it reflects a hybrid of retrosynthetic analysis[43] and divergent/diversity-oriented synthesis[44,45], a combination that could be argued to be obvious in terms of the bond connections desired, but is not obvious in terms of the tools needed to execute those reactions (particularly positionally selective transforms). Whether or not this overall strategy can be extended to the synthesis of other oligomer families remains to be established. What is certain is that the evaluation of that question will offer many opportunities for reagent and reaction discovery, particularly as more complex brominative functionalizations are examined using natural products, compounds pertinent to materials research, and building blocks needed for pharmaceuticals.

METHODS SUMMARY

All reactions were carried out under an argon atmosphere with dry solvents under anhydrous conditions; dry THF, toluene, benzene, diethyl ether, and CH_2Cl_2 were obtained by passing commercially available pre-dried, oxygen-free formulations through activated alumina columns. Yields refer to chromatographically and spectroscopically (^{1}H and ^{13}C NMR) homogeneous materials, unless otherwise stated. Reagents were purchased at the highest commercial quality and used without further purification, unless otherwise stated. Reactions were magnetically stirred and monitored by thin-layer chromatography. For full experimental details, including procedures for all reactions and characterization of all new compounds (^{1}H NMR, ^{13}C NMR, mass spectrometry, infrared, R_f value), see the Supplementary Information.

Received 20 March; accepted 13 May 2011.

1. Fischbach, M. A. & Clardy, J. One pathway, many products. *Nature Chem. Biol.* **3,** 353–355 (2007).
2. Christianson, D. W. Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **106,** 3412–3442 (2006).
3. Chen, K. & Baran, P. S. Total synthesis of eudesmane terpenes by site-selective C–H oxidations. *Nature* **459,** 824–828 (2009).
4. Sotheeswaran, S. & Pasupathy, V. Distribution of resveratrol oligomers in plants. *Phytochemistry* **32,** 1083–1092 (1993).
5. Quideau, S., Deffieux, D., Douat-Casassus, C. & Pouységu, L. Plant polyphenols: chemical properties, biological activities, and synthesis. *Angew. Chem. Int. Edn* **50,** 586–621 (2011).
6. Snyder, S. A., ElSohly, A. M. & Kontes, F. Synthetic approaches to oligomeric natural products. *Nat. Prod. Rep.* **28,** 897–924 (2011).
7. Takaya, Y., Yan, K.-X., Terashima, K., Ito, J. & Niwa, M. Chemical determination of the absolute structures of resveratrol dimers, ampelopsins A, B, D and F. *Tetrahedron* **58,** 7259–7265 (2002).
8. Takaya, Y., Yan, K.-X., Terashima, K., He, Y.-H. & Niwa, M. Biogenetic reactions on stilbene tetramers from Vitaceaeous plants. *Tetrahedron* **58,** 9265–9271 (2002).
9. Wang, S., Ma, D. & Hu, C. Three new compounds from the aerial parts of *Caragana sinica*. *Helv. Chim. Acta* **88,** 2315–2321 (2005).
10. Tanaka, T. *et al.* Six new heterocyclic stilbene oligomers from stem bark of *Shore hemsleyana*. *Heterocycles* **55,** 729–740 (2001).
11. Wang, S., Ma, D., & Hu, C. Two new oligostilbenes from *Caragana sinica*. *J. Asian Nat. Prod. Res.* **6,** 241–248 (2004).
12. Oshima, Y., Ueno, Y., Hisamichi, K. & Takeshita, M. Ampelopsins F and G, novel bridged plant oligostilbenes from *Ampelopsis brevipedunculata* var. *hancei* roots (Vitaceae). *Tetrahedron* **49,** 5801–5804 (1993).
13. Tanaka, T., Ito, T., Nakaya, K., Iinuma, M. & Riswan, S. Oligostilbenoids in stem bark in *Vatica rassak*. *Phytochemistry* **54,** 63–69 (2000).
14. Jang, M. *et al.* Cancer chemopreventive activity of resveratrol, a natural product derived from grapes. *Science* **275,** 218–220 (1997).
15. Milne, J. C. *et al.* Small molecules activators of SIRT1 as therapeutics for the treatment of type 2 diabetes. *Nature* **450,** 712–716 (2007).
16. Tsukamoto, T. *et al.* Vaticanol C, a resveratrol tetramer, activates PPARα and PPARβ/γ *in vitro* and *in vivo*. *Nutr. Metabol.* 7, doi:10.1186/1743–7075–7-46 (2010).
17. Ito, T. *et al.* Antitumor effect of resveratrol oligomers against human cancer cell lines and the molecular mechanism of apoptosis induced by vaticanol C. *Carcinogenesis* **24,** 1489–1497 (2003).
18. Abe, N. *et al.* Resveratrol oligomers from *Vatica albiramis*. *J. Nat. Prod.* **73,** 1499–1506 (2010).
19. Atun, S., Aznam, N., Arianingrum, R., Takaya, Y. & Niwa, M. Resveratrol derivatives from stem bark of Hopea and their biological activity test. *J. Physiol. Sci.* **19,** 7–21 (2008).
20. Yamada, M. *et al.* Stilbenoids of *Kobresia nepalensis* (Cyperaceae) exhibiting DNA topoisomerase II inhibition. *Phytochemistry* **67,** 307–313 (2006).

21. Langcake, P. & Pryce, R. J. Oxidative dimerisation of 4-hydroxystilbenes *in vitro*: production of a grapevine phytoalexin mimic. *J. Chem. Soc. Chem. Commun.* 208–210 (1977).
22. Sako, M., Hosokawa, H., Ito, T. & Iinuma, M. Regioselective oxidative coupling of 4-hydroxystilbenes: synthesis of resveratrol and ε-viniferin (*E*)-dehydrodimers. *J. Org. Chem.* **69,** 2598–2600 (2004).
23. Li, W., Li, H. & Hou, Z. Total synthesis of (±)-quadrangularin A. *Angew. Chem. Int. Edn* **45,** 7609–7611 (2006).
24. Li, W., Li, H., Luo, Y., Yang, Y. & Wang, N. Biosynthesis of resveratrol dimers by regioselective oxidative coupling reaction. *Synlett* 1247–1250 (2010).
25. Velu, S. S. *et al.* Regio- and stereoselective biomimetic synthesis of oligostilbenoid dimers from resveratrol analogues: influence of the solvent, oxidant, and substitution. *Chem. Eur. J.* **14,** 11376–11384 (2008).
26. Takaya, Y. *et al.* Biomimetic transformation of resveratrol. *Tetrahedron* **61,** 10285–10290 (2005).
27. He, Y.-H., Takaya, Y., Terashima, K. & Niwa, M. Determination of absolute structure of (+)-davidiol A. *Heterocycles* **68,** 93–100 (2006).
28. Kim, I. & Choi, J. A versatile approach to oligostilbenoid natural products—synthesis of permethylated analogues of viniferifuran, malibatol A, and shoreaphenol. *Org. Biomol. Chem.* **7,** 2788–2795 (2009).
29. Kraus, G. A. & Gupta, V. A new synthetic strategy for the synthesis of bioactive stilbene dimers. A direct synthesis of amurensin H. *Tetrahedr. Lett.* **50,** 7180–7183 (2009).
30. Jeffrey, J. L. & Sarpong, R. Concise synthesis of paucifloral F using a Larock annulation. *Org. Lett.* **11,** 5450–5453 (2009).
31. Nicolaou, K. C., Kang, Q., Wu, T. R., Lim, C. S. & Chen, D. Y.-K. Total synthesis and biological evaluation of the resveratrol-derived polyphenol natural products hopeanol and hopeahainol A. *J. Am. Chem. Soc.* **132,** 7540–7548 (2010).
32. Snyder, S. A., Zografos, A. L. & Lin, Y. Total synthesis of resveratrol-based natural products: a chemoselective approach. *Angew. Chem. Int. Edn* **46,** 8186–8191 (2007).
33. Snyder, S. A., Breazzano, S. P., Ross, A. G., Lin, Y. & Zografos, A. Total synthesis of diverse carbogenic complexity within the resveratrol class from a common building block. *J. Am. Chem. Soc.* **131,** 1753–1765 (2009).
34. Sculimbrene, B. R., Morgan, A. J. & Miller, S. J. Enantiodivergence in small-molecule catalysis of asymmetric phosphorylation: concise total syntheses of the enantiomeric D-*myo*-inositol-1-phosphate and D-*myo*-inositol-3-phosphate. *J. Am. Chem. Soc.* **124,** 11653–11656 (2002).
35. Lewis, C. A. & Miller, S. J. Site-selective derivatization and remodeling of erythromycin A by using peptide-based chiral catalysts. *Angew. Chem. Int. Edn* **45,** 5616–5619 (2006).
36. Bertolini, F. & Pineschi, M. Recent progress in the synthesis of 2,3-dihydrofurans. *Org. Prep. Proced. Intl* **41,** 385–418 (2009).
37. Corey, E. J. & Chaykovsky, M. Dimethyloxosulfonium methylide and dimethylsulfonium methylide. Formation and application to organic synthesis. *J. Am. Chem. Soc.* **87,** 1353–1364 (1965).
38. Bach, N. J, *et al.* Bicyclic and tricyclic ergoline partial structures. Rigid 3-(2-aminoethyl)pyrroles and 3- and 4-(2-aminoethyl)pyrazoles as dopamine agonists. *J. Med. Chem.* **23,** 481–491 (1980).
39. Baker, R. Cooke, N. G., Humphrey, G. R., Wright, S. H. B. & Hirshfield, J. Stereoselective synthesis of the dihydrobenzo[b]furan segments of the ephedradine alkaloids. *Chem. Commun.* 1102–1004 (1987).
40. Kurosawa, W., Kobayashi, H., Kan, T. & Fukuyama, T. Total synthesis of (−)-ephedradine A: an efficient construction of optically active dihydrobenzofuran-ring via C–H insertion reaction. *Tetrahedron* **60,** 9615–9628 (2004).
41. Snyder, S. A., Treitler, D. S. & Brucks, A. P. Simple reagents for direct halonium-induced polyene cyclization. *J. Am. Chem. Soc.* **132,** 14303–14314 (2010).
42. Gustafson, J., Lim, D. & Miller, S. J. Dynamic kinetic resolution of biaryl atropisomers via peptide-catalyzed asymmetric bromination. *Science* **328,** 1251–1255 (2010).
43. Corey, E. J. & Cheng, X. M. *The Logic of Chemical Synthesis* (Wiley, 1995).
44. Boger, D. L. & Brotherton, C. E. Total synthesis of azafluoranthene alkaloids: rufescine and imeluteine. *J. Org. Chem.* **49,** 4050–4055 (1984).
45. Burke, M. D. & Schreiber, S. L. A planning strategy for diversity-oriented synthesis. *Angew. Chem. Int. Edn* **43,** 46–58 (2004).

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

Acknowledgements We thank G. Sukenick of The Memorial Sloan Kettering Cancer Research Institute and J. Decatur of Columbia University for NMR assistance, Y. Itagaki for mass spectrometric assistance, C. Stathakis for preliminary attempts to form dihydrofuran units on the pallidol core, A. ElSohly for theoretical calculations and discussions, and K. Shaw and J. Boyce for preparing some starting materials. Financial support was provided by Columbia University, the National Institutes of Health (R01-GM84994), Bristol-Myers Squibb, Eli Lilly, the Research Corporation for Science Advancement (Cottrell Scholar Award to S.A.S.), and the Austrian Science Fund (FWF, Schrödinger postdoctoral fellowship J2986-N19 to A.G.).

Author Contributions S.A.S. conceived and directed the research, as well as composed the manuscript. A.G. developed the dihydrofuran synthesis approach and completed compounds **16**, **17**, **30**, and **38**. M.I.C. completed compound **19** as well as the majority of the route towards compound **18** including the BDSB-based functionalization. Both A.G. and M.I.C. worked to complete **18**, and provided commentary and feedback on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.A.S. (sas2197@columbia.edu).

ARTICLE

doi:10.1038/nature10153

Structural basis of steroid hormone perception by the receptor kinase BRI1

Michael Hothorn[1], Youssef Belkhadir[1,2]†, Marlene Dreux[3], Tsegaye Dabi[1,2], Joseph. P. Noel[2,4], Ian A. Wilson[5,6] & Joanne Chory[1,2]

Polyhydroxylated steroids are regulators of body shape and size in higher organisms. In metazoans, intracellular receptors recognize these molecules. Plants, however, perceive steroids at membranes, using the membrane-integral receptor kinase BRASSINOSTEROID INSENSITIVE 1 (BRI1). Here we report the structure of the *Arabidopsis thaliana* BRI1 ligand-binding domain, determined by X-ray diffraction at 2.5 Å resolution. We find a superhelix of 25 twisted leucine-rich repeats (LRRs), an architecture that is strikingly different from the assembly of LRRs in animal Toll-like receptors. A 70-amino-acid island domain between LRRs 21 and 22 folds back into the interior of the superhelix to create a surface pocket for binding the plant hormone brassinolide. Known loss- and gain-of-function mutations map closely to the hormone-binding site. We propose that steroid binding to BRI1 generates a docking platform for a co-receptor that is required for receptor activation. Our findings provide insight into the activation mechanism of this highly expanded family of plant receptors that have essential roles in hormone, developmental and innate immunity signalling.

Signal perception at the cell surface, and transduction of this signal to the cell's interior, are essential to all life forms. Plants have met this challenge in part by evolving membrane-integral receptor kinases. Many of these receptors are composed of an extracellular leucine-rich repeat (LRR) module and a cytoplasmic kinase domain, connected by a single membrane-spanning helix[1]. Receptors with this architecture (LRR-receptor kinase, LRR-RK), for example, regulate plant growth[2], development[3,4] and interactions with the environment[5,6]. Their corresponding ligands range from small molecules[7] and peptides[8] to entire proteins[5].

The LRR-RK BRI1 (refs 2, 9) controls a steroid signalling pathway essential for plant growth[10]. Whereas animal steroid receptors are found predominantly in the nucleus[11], BRI1 is localized at the plasma-membrane and in endosomes[12]. The following model for BRI1 activation has been proposed: in the absence of brassinosteroid, BRI1's kinase domain is kept in a basal state by its auto-inhibitory carboxy-terminal tail[13], as well as by interaction with the inhibitor protein BKI1 (ref. 14). Hormone binding to the extracellular domain of BRI1 (refs 7, 15), in a region that includes a ~70 amino acid 'island' domain[16], causes a change in the receptor (a conformational change in a preformed homodimer[13] or receptor dimerization), leading to autophosphorylation of the BRI1 kinase domain[17], release of its C-terminal tail[13], and trans-phosphorylation of the inhibitor BKI1 (refs 14, 18). BKI1 then dissociates from the membrane, allowing BRI1 to interact with a family of smaller LRR-RKs[19], including the BRI1 ASSOCIATED KINASE 1 (BAK1)[20,21]. The kinase domains of BRI1 and BAK1 trans-phosphorylate each other on multiple sites[22], and the fully activated receptor triggers downstream signalling events[23], resulting in major changes in nuclear gene expression[10].

The architecture of BRI1 is reminiscent of animal Toll-like innate immunity receptors (TLRs), and notably several plant LRR-RKs are immunity receptors[5]. It was, thus, reasonable to assume that the BRI1 ectodomain would form a TLR-like horseshoe structure[24] that binds its ligand along a dimer interface, as observed in several TLRs[25,26]. Here we report the structure of the ligand-binding domain of *Arabidopsis* BRI1 in its free form and bound to the steroid brassinolide, and show that BRI1 folds into a superhelical assembly, whose interior provides the hormone-binding site. Comparison of the free and hormone-bound structures, combined with genetic data, suggests a novel activation mechanism for BRI1 that is distinct from TLRs.

Overall structure of the BRI1 ectodomain

BRI1 was expressed in baculovirus-infected insect cells and the secreted ectodomain was purified by tandem-affinity and size-exclusion chromatography. The crystal structure was solved to 2.5 Å resolution by single isomorphous replacement (see Methods, Supplementary Table 1 and Fig. 1). BRI1 does not adopt the anticipated TLR-horseshoe structure but forms a right-handed superhelix composed of 25 LRRs (Fig. 1a). The helix completes one full turn, with a rise of ~70 Å. The concave surface, which determines the curvature of the solenoid[27], is formed by α- and 3_{10} helices (green in Fig. 1a) that produce inner and outer diameters of ~30 and ~60 Å, respectively. The overall curvature of BRI1 is similar to that of TLR3 (ref. 24; Fig. 1b), but, whereas the TLR3 ectodomain is essentially flat, BRI1 is highly twisted (Fig. 1b). Such twisted assemblies of LRRs have been observed previously with bacterial effector[28] and adhesion proteins[29], and with the plant defence protein PGIP[30] (Supplementary Fig. 2). The twist of PGIP's LRR domain is caused by a non-canonical, second β-sheet that is oriented perpendicular to the central β-sheet forming the inner surface of the solenoid[30]. Additional β-sheets are also present in our structure (blue in Fig. 1a, Supplementary Fig. 3), but in the case of the much larger BRI1 ectodomain result in a superhelical assembly (Fig. 1a). The second β-strand in PGIP and in BRI1 is followed by an Ile-Pro spine that runs along the outer surface of the helix and provides packing interactions between consecutive LRRs (Fig. 2a and ref. 30). Both structural features are directly linked to

[1]Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. [2]Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. [3]Department of Immunology and Microbial Science, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. [4]Jack H. Skirball Center for Chemical Biology and Proteomics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. [5]Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. [6]Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. †Present address: Moroccan Foundation for Advanced Science, Innovation and Research, Biotechnology Development Center, Technopolis Rabatshore, Sala al Jadida, 11000, Morocco.

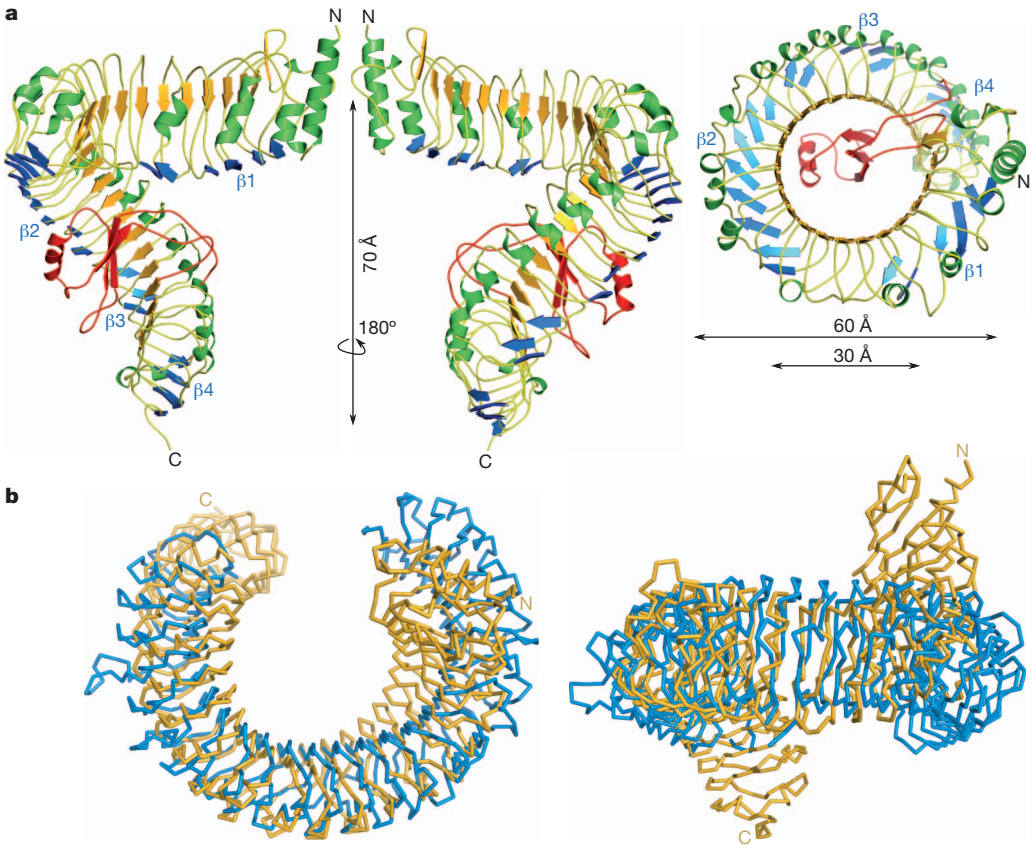


Figure 1 | The BRI1 ectodomain forms a superhelix. **a**, Ribbon diagrams of the ectodomain. Front, back and top views are shown (left, middle and right, respectively). The canonical β-sheet is shown in orange, helices in green, the plant-specific β-sheets in blue, and the island domain is depicted in red. **b**, Structural comparison of BRI1 (C_α trace, in yellow) and TLR3 (in blue; PDB 1ZIW)[24]. The structures superimpose with an root mean square deviation of 4.2 Å between 341 corresponding C_α atoms. Top and side views are shown (left and right, respectively). The island domain has been omitted for clarity.

the Lt/sGxIP consensus sequence of the plant-specific LRR subfamily[31] (Fig. 2c, Supplementary Fig. 4 and Supplementary Table 2). Because this consensus sequence is found in other plant receptor kinases, these receptors may also harbour twisted LRR domains (Fig. 2c), making BRI1 the primary template for the study of diverse signalling pathways in plants[3–6].

Amino- and C-terminal flanking regions that cap the hydrophobic core of the BRI1 solenoid are similar to caps previously described for PGIP[30] (Supplementary Fig. 5). Notably, not only are these caps stabilized by disulphide bridges, but five additional disulphide bonds link consecutive LRR segments in the N-terminal half of the BRI1 ectodomain (Fig. 2b, Supplementary Fig. 4 and Supplementary Table 2).

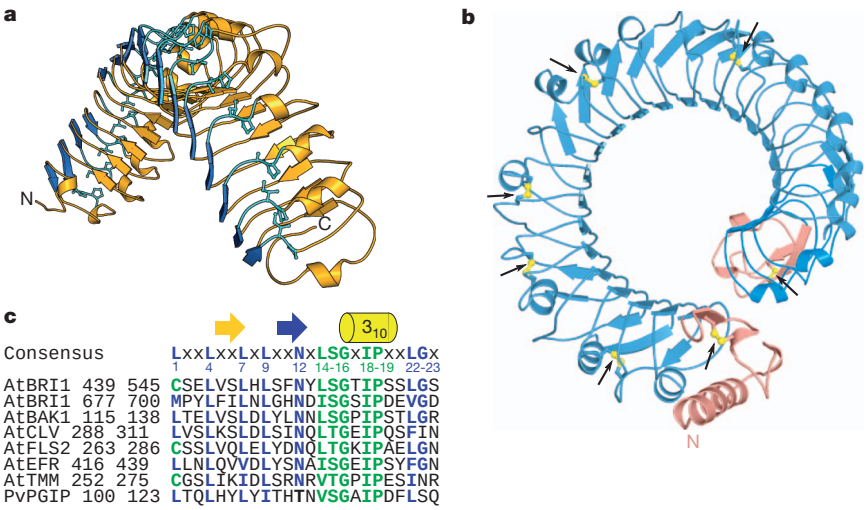


Figure 2 | Plant-specific sequence fingerprints cause the superhelical arrangement. **a**, Ribbon diagram of the convex side of LRRs 9-25. The non-canonical β-strands and the Ile-Pro spine are shown in dark and light blue, respectively. **b**, Top view of the BRI1 ectodomain. Disulphide bridges are in yellow, the N- and C-terminal caps are in pink. **c**, Sequence alignment of LRRs in BRI1, other plant receptor kinases[3–5,20,21] and PGIP[30]. At, *Arabidopsis thaliana*. Pv, *Phaseolus vulgaris*. The canonical LRR consensus sequences are highlighted in blue, plant-specific motifs in green.

The island domain

The island domain in BRI1 corresponds to a large insertion in the regular repeat-structure between LRRs 21 and 22 (residues 584–654; Fig. 1a). The resulting ~70-residue segment forms a small domain that folds back into the interior of the superhelix, where it makes extensive polar and hydrophobic interactions with LRRs 13–25 (Fig. 1a, Supplementary Fig. 6 and Supplementary Table 2). The domain fold is characterized by an anti-parallel β-sheet, which is sandwiched between the LRR core and a 3_{10} helix and stabilized by a disulphide bridge (Fig. 3a, Supplementary Fig. 4). The loss-of-function alleles *bri1-9* (Ser662Phe, weak)[32] and *bri1-113* (Gly611Glu, strong)[2] map to this island domain–LRR interface (Supplementary Fig. 6), and probably interfere with folding of the island domain[33]. Two long loops that connect the island domain to the LRR core appear partially disordered in the unliganded receptor (Supplementary Fig. 7). The insertion of a folded domain into the LRR repeat has not been observed in other LRR receptor structures, and is probably an adaptation to the challenge of sensing a small steroid ligand (rather than larger ligands, such as proteins, nucleic acids, or lipids[25,26]).

We next solved a 2.5 Å co-crystal structure with brassinolide, a potent *Arabidopsis* steroid that binds BRI1 with nanomolar affinity[7,34]. One molecule of brassinolide per BRI1 monomer binds in close proximity to the island domain (Fig. 3a–c), which was previously implicated in steroid binding[7,16]. Our structure reveals that the LRR superhelix and the island domain both extensively contribute to formation of the hormone binding site. The A–D rings of the steroid bind to a hydrophobic surface, which is provided by LRRs 23–25 and that maps to the inner side of the BRI1 superhelix (Fig. 3b, d, Supplementary Fig. 8). The alkyl chain of the hormone fits into a small pocket formed by residues originating from LRRs 21 and 22 (Ile 563, Trp 564, Met 657, Phe 658) and from two loops connecting the island domain with the LRR core (Fig. 3d). The hydrophobic nature and restricted size of this pocket now explain why steroid ligands with bulkier or charged alkyl side chains, such as the arthropod steroid ecdysone (Supplementary Fig. 8), cannot be recognized by BRI1 (ref. 7). A few polar interactions with brassinolide's second diol moiety (Fig. 3d) are established with Tyr 597 and main-chain atoms from His 645 and Ser 647 in the island domain, and are mediated by water molecules (Fig. 3d). Mutation of the neighbouring Gly 644 to Asp may interfere with this hydrogen-bonding network, and explain why this mutation greatly reduces the binding activity of the receptor[7] and causes the loss-of-function phenotype *bri1*-6 (ref. 32; Fig. 3d). No polar contacts are observed with the seven-membered B-ring lactone (Fig. 3d), consistent with B-ring modifications as found in, for example, castasterone (Supplementary Fig. 8) being tolerated by BRI1 (refs 7, 35).

The steroid-complex reveals a hormone-binding site that involves a much larger portion of the LRR domain than previously anticipated[16]. Major interactions between the steroid and the BRI1 ectodomain originate from the very C-terminal LRRs 23–25, which bring the hormone in close proximity to the membrane (Fig. 3a,d). Importantly, while there is a significant hormone–receptor interface (550 $Å^2$) for such a small molecule ligand, large parts of the steroid are exposed to the solvent, including the 2α,3α-diol moiety in brassinolide that is important for biological activity[35]. Thus, protein–protein interactions may be involved in the recognition of the steroid ligand, with the hormone itself providing a docking platform. Importantly, steroid binding induces a conformational rearrangement and fixing of the island domain, which becomes fully ordered and competent to participate in protein–protein interactions that could be critical for receptor activation (Supplementary Fig. 7).

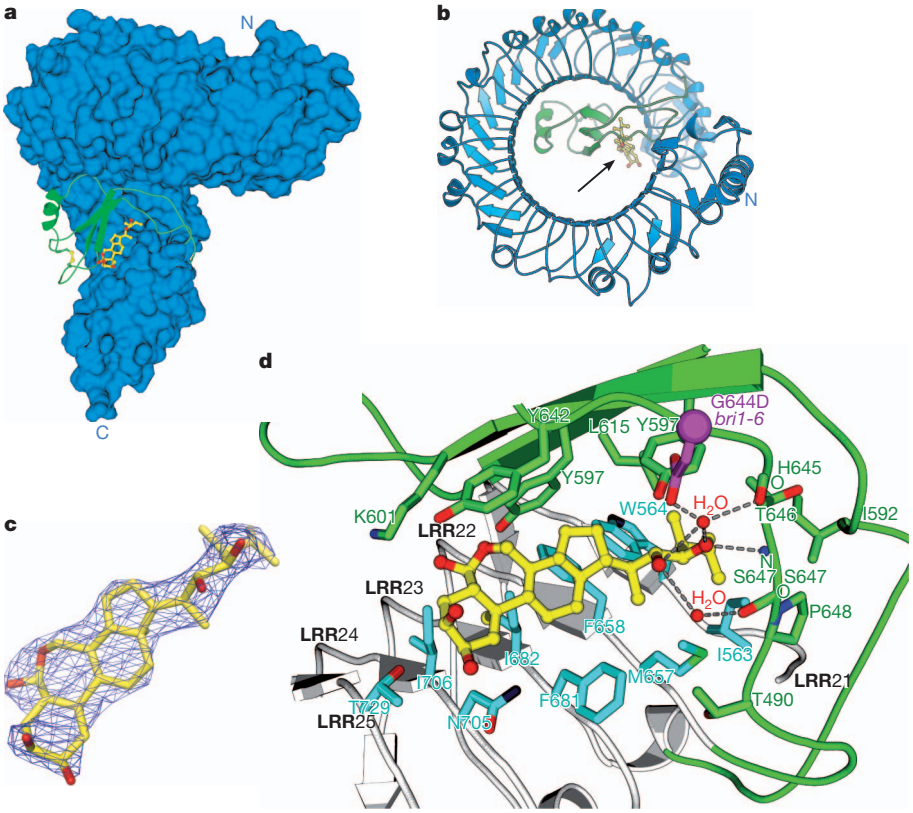


Figure 3 | The steroid hormone binding site maps to the C-terminal inner surface of the superhelix. **a**, Brassinolide (yellow sticks) binds to a surface provided by the LRR domain (in blue) and by parts of the island domain (green ribbon). **b**, Location of the steroid (arrowed) in the centre of the BRI1 superhelix. **c**, Close-up view of brassinolide, including an omit $2F_o - F_c$ electron density map contoured at 1.5 σ. **d**, Protein–hormone interactions in the BRI1 steroid binding site. Ribbon diagram of LRRs 21–25 (in grey) is shown, together with parts of the island domain (in green). Contacting residues are in full side-chain representation, polar interactions are dotted lines, and water molecules are red spheres. *bri1-6* (Gly644Asp) is depicted in magenta.

A protein interaction platform

Four known BRI1 mis-sense alleles map to the inner surface of the last five LRRs (Fig. 4a). This surface is not masked by carbohydrate (Supplementary Fig. 9), and contains both the hormone-binding site and the island domain (Figs 3a, d and 4a). Three mutations cluster in a loop connecting the island domain with LRR 22 (Fig. 4a). This loop is partially disordered in the unliganded structure but is well-defined in the brassinolide complex (Supplementary Fig. 7). We speculate that this loop, when ordered, is engaged in protein–protein interactions that are critical for receptor activation, and that mis-sense alleles in BRI1 modulate these interactions. The gain-of-function allele *sud1* (ref. 36; Gly643Glu) may establish contact with Ser 623 in the island domain, and lead to an ordered loop even in the absence of steroid ligand (Supplementary Fig. 10). Mutation of the neighbouring Gly 644 to Asp causes the loss-of-function phenotype *bri1-6* (ref. 32; see above, and Figs 3d, 4a), and mutation of conserved Thr 649 to Lys inactivates barley BRI1 (ref. 37). These mutations, when modelled *in silico*, induce steric clashes with residues in the island domain and in the underlying LRR domain (Supplementary Fig. 10), and thus may distort the position of the loop. Interestingly, *bri1-102*, a strong loss-of-function mutation (Thr750Ile)[38] that does not affect steroid binding[7], maps to a distinct surface area in LRR 25 (Fig. 4a). Thus protein–protein interactions critical for receptor activation may not be restricted to the island domain, but also involve residues from the LRR core.

Receptor activation

BRI1 has been shown to exist at least partially as a homo-oligomer *in planta*[13,39,40]. Thus steroid binding to the island domain and the concomitant rearrangements of the island domain loop could induce a conformational change in a preformed BRI1 homodimer[13], or allow for ligand-dependent dimerization of the BRI1 ectodomain. However, models of BRI1 dimers that bring the C termini of their ectodomains into close proximity (we note that the cytoplasmic kinase domains of BRI1 can interact[13,18]) and that make use of the interaction surface outlined above, encounter steric clashes with the N-terminal LRRs (Supplementary Fig. 11). Furthermore, in contrast to TLR ectodomains, which in crystals tend to form homodimers even in the absence of ligand[24,25], dimers cannot be seen in BRI1 crystals grown under the same acidic pH conditions that are typically associated with the plant cell wall. The largest interface area between two neighbouring BRI1 molecules amounts to only ~1.5% of the total accessible surface area, consistent with the high solvent content of our crystals (see Supplementary Methods). The main crystal contact involves a head-to-head arrangement of two BRI1 monomers, a configuration that would place the cytoplasmic kinase domains far apart (Supplementary Fig. 12). In size-exclusion chromatography experiments, the recombinant BRI1 ectodomain elutes as a monomer in the absence of steroid ligand, and shows no tendency to dimerize or oligomerize in the presence of a ~4 times molar excess of brassinolide (Fig. 4b).

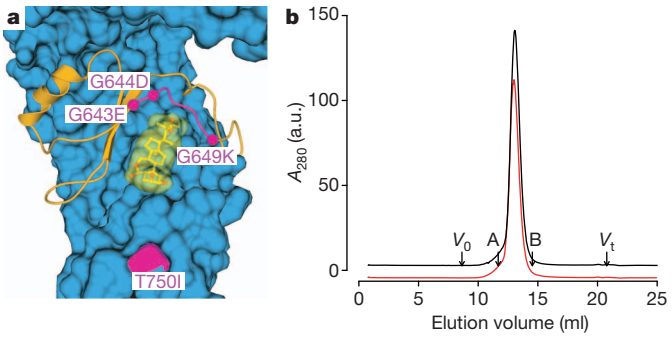


Figure 4 | An accessible membrane-proximal region of BRI1 may provide a protein–protein interaction platform. a, Overview of the C-terminal surface area not masked by carbohydrate. Brassinolide is shown in yellow, the island domain in orange, and genetic alleles in magenta. **b**, Analytical gel-filtration trace (absorbance at 280 nm). The free ectodomain elutes as a monomer (black line), as does a putative complex with brassinolide (red line). Void volume (V_0) and total volume (V_t) are shown, together with elution volumes for molecular mass standards (A, aldolase, molecular mass 158,000 Da; B, conalbumin, molecular mass 75,000 Da). The calculated molecular mass for the monomer peak is ~125 kDa. The molecular mass of purified BRI1 is ~110 kDa.

Our analyses suggest that the superhelical BRI1 LRR domain alone has no tendency to oligomerize, indicating that BRI1 receptor activation may not be mediated by ligand-induced homodimerization of the ectodomain (as described for TLRs[25,26]) or by conformational changes in preformed homodimers[13]. We do not dismiss the facts that the cytoplasmic kinase domain of BRI1 can dimerize[18], or that BRI1 homo-oligomers are present *in vivo*[13,39,40]. However, our structures reinforce the notion that homo-oligomerization of BRI1 may be constitutive and independent of ligand stimulus[39]. The presence of an interaction platform that undergoes conformational changes when steroid binds and that harbours several loss- and gain-of-function alleles suggests that interaction with another protein factor may control BRI1 activation.

Discussion

The structure of the BRI1 ectodomain offers several new insights, and its twisted shape is likely to characterize the architecture of many plant LRR-RKs[1]. The presence of a folded-domain insert appears to be an adaptation to the recognition of a small molecule ligand, a challenge that smaller LRR proteins have met by generating loop insertions into their capping motifs[41]. BRI1's fascinating mode of ligand recognition reveals how steroids can be sensed at membranes and rationalizes a large set of genetic and biochemical findings[2,7,32,38].

Different BRI1 receptor activation mechanisms have been proposed, including ligand-dependent dimerization as seen for TLRs[25,26] and ligand-induced conformational changes in preformed homodimers[13]. Our analyses suggest that the superhelical shape of the BRI1 ectodomain is incompatible with homodimerization, and that the isolated ectodomain behaves as a monomer even in the presence of steroid. These findings leave us with the alternative hypothesis that another protein factor could bind to the interaction platform in BRI1 that would minimally encompass the steroid ligand, LRRs 21–25 and parts of the island domain (Fig. 4a). Although it is possible that an unknown protein fulfils this role and provides a dimerization interface for two BRI1 molecules (as seen, for example, for TLR4; ref. 26), genetic and biochemical screens have not uncovered this protein. It is thus possible that the small receptor kinase BAK1 acts as a direct brassinosteroid co-receptor, as suggested previously[10,20]. It has been demonstrated that BAK1 is a genetic component of the brassinosteroid pathway[20,21], that BRI1 and BAK1 interact in a steroid-dependent manner[22] and that both receptors trans-phosphorylate each other on ligand stimulus[22]. Notably, a homology model of the BAK1 ectodomain (Supplementary Fig. 13) is compatible in size and shape with the interaction platform in BRI1, and the BAK1 *elg* allele, which maps to the BAK1 ectodomain (Supplementary Fig. 14), renders plants hypersensitive to brassinosteroid treatment[42]. We speculate that the *sud1*, *bri1-6*, *bri1-102* and *elg* mutations modulate the interaction between the BRI1 and BAK1 ectodomains in a brassinosteroid-dependent manner (Supplementary Fig. 14). The demonstration that BAK1 is essential for brassinosteroid sensing may have been obscured owing to genetic redundancy[20], with at least two BAK1-like proteins interacting with BRI1 *in vivo*[43,44]. We recently overcame this limitation by showing that the BRI1 inhibitor protein BKI1 blocks the interaction between the BAK1 and BRI1 kinase domains[18]. Importantly, transgenic lines that constitutively deliver BKI1 to the site of BRI1 signalling resemble strong BRI1 loss-of-function mutants, suggesting an important role for receptor–co-receptor association in brassinosteroid signal initiation[18].

Future studies will undoubtedly test this heteromerization model and dissect the relative contributions of the BRI1 and BAK1 ectodomains, their transmembrane segments and their cytoplasmic kinase domains to receptor activation. It will become important then to

understand how BAK1 could serve as co-receptor for other LRR-RK signalling pathways[5,19].

METHODS SUMMARY

The BRI1 ectodomain (residues 29–788) was produced by secreted expression in baculovirus-infected insect cells, harvested 4 d after infection by ultrafiltration and purified by tandem-affinity chromatography, then by gel filtration. BRI1 was concentrated to 15 mg ml^{-1} and crystallized by vapour diffusion using a reservoir solution containing 14% PEG 4,000, 0.2 M $(NH_4)_2SO_4$, 0.1 M citric acid (pH 4.0). The brassinolide complex was obtained by co-crystallization. Diffraction data to 2.5 Å resolution were collected on a rotating anode X-ray generator and at beamline 8.2.1 of the Advanced Light Source (ALS), Berkeley. The structure was solved using the single isomorphous replacement method. Data and refinement statistics are summarized in Supplementary Table 1.

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

Received 21 January; accepted 26 April 2011.
Published online 12 June 2011.

1. Shiu, S. H. & Bleecker, A. B. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc. Natl Acad. Sci. USA* **98,** 10763–10768 (2001).
2. Li, J. & Chory, J. A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* **90,** 929–938 (1997).
3. Clark, S. E., Williams, R. W. & Meyerowitz, E. M. The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis. Cell* **89,** 575–585 (1997).
4. Nadeau, J. A. & Sack, F. D. Control of stomatal distribution on the *Arabidopsis* leaf surface. *Science* **296,** 1697–1700 (2002).
5. Gómez-Gómez, L. & Boller, T. FLS2: an LRR receptor-like kinase involved in the perception of the bacterial elicitor flagellin in *Arabidopsis. Mol. Cell* **5,** 1003–1011 (2000).
6. Nishimura, R. *et al.* HAR1 mediates systemic regulation of symbiotic organ development. *Nature* **420,** 426–429 (2002).
7. Wang, Z. Y., Seto, H., Fujioka, S., Yoshida, S. & Chory, J. BRI1 is a critical component of a plasma-membrane receptor for plant steroids. *Nature* **410,** 380–383 (2001).
8. Ogawa, M., Shinohara, H., Sakagami, Y. & Matsubayashi, Y. *Arabidopsis* CLV3 peptide directly binds CLV1 ectodomain. *Science* **319,** 294 (2008).
9. Belkhadir, Y. & Chory, J. Brassinosteroid signaling: a paradigm for steroid hormone signaling from the cell surface. *Science* **314,** 1410–1411 (2006).
10. Vert, G., Nemhauser, J. L., Geldner, N., Hong, F. & Chory, J. Molecular mechanisms of steroid hormone signaling in plants. *Annu. Rev. Cell Dev. Biol.* **21,** 177–201 (2005).
11. Mangelsdorf, D. J. *et al.* The nuclear receptor superfamily: the second decade. *Cell* **83,** 835–839 (1995).
12. Geldner, N., Hyman, D. L., Wang, X., Schumacher, K. & Chory, J. Endosomal signaling of plant steroid receptor kinase BRI1. *Genes Dev.* **21,** 1598–1602 (2007).
13. Wang, X. *et al.* Autoregulation and homodimerization are involved in the activation of the plant steroid receptor BRI1. *Dev. Cell* **8,** 855–865 (2005).
14. Wang, X. & Chory, J. Brassinosteroids regulate dissociation of BKI1, a negative regulator of BRI1 signaling, from the plasma membrane. *Science* **313,** 1118–1122 (2006).
15. He, Z. *et al.* Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. *Science* **288,** 2360–2363 (2000).
16. Kinoshita, T. *et al.* Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* **433,** 167–171 (2005).
17. Wang, X. *et al.* Identification and functional analysis of *in vivo* phosphorylation sites of the *Arabidopsis* BRASSINOSTEROID-INSENSITIVE1 receptor kinase. *Plant Cell* **17,** 1685–1703 (2005).
18. Jaillais, Y. *et al.* Tyrosine phosphorylation controls brassinosteroid receptor activation by triggering membrane release of its kinase inhibitor. *Genes Dev.* **25,** 232–237 (2011).
19. Chinchilla, D., Shan, L., He, P., de Vries, S. & Kemmerling, B. One for all: the receptor-associated kinase BAK1. *Trends Plant Sci.* **14,** 535–541 (2009).
20. Nam, K. H. & Li, J. BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. *Cell* **110,** 203–212 (2002).
21. Li, J. *et al.* BAK1, an *Arabidopsis* LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell* **110,** 213–222 (2002).
22. Wang, X. *et al.* Sequential transphosphorylation of the BRI1/BAK1 receptor kinase complex impacts early events in brassinosteroid signaling. *Dev. Cell* **15,** 220–235 (2008).
23. Kim, T.-W. & Wang, Z.-Y. Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annu. Rev. Plant Biol.* **61,** 681–704 (2010).
24. Choe, J., Kelker, M. S. & Wilson, I. A. Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. *Science* **309,** 581–585 (2005).
25. Liu, L. *et al.* Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science* **320,** 379–381 (2008).
26. Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* **458,** 1191–1195 (2009).
27. Bella, J., Hindle, K. L., McEwan, P. A. & Lovell, S. C. The leucine-rich repeat structure. *Cell. Mol. Life Sci.* **65,** 2307–2333 (2008).
28. Evdokimov, A. G., Anderson, D. E., Routzahn, K. M. & Waugh, D. S. Unusual molecular architecture of the *Yersinia pestis* cytotoxin YopM: a leucine-rich repeat protein with the shortest repeating unit. *J. Mol. Biol.* **312,** 807–821 (2001).
29. Schubert, W. D. *et al.* Structure of internalin, a major invasion protein of *Listeria monocytogenes*, in complex with its human receptor E-cadherin. *Cell* **111,** 825–836 (2002).
30. Di Matteo, A. *et al.* The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proc. Natl Acad. Sci. USA* **100,** 10124–10128 (2003).
31. Kajava, A. V. Structural diversity of leucine-rich repeat proteins. *J. Mol. Biol.* **277,** 519–527 (1998).
32. Noguchi, T. *et al.* Brassinosteroid-insensitive dwarf mutants of *Arabidopsis* accumulate brassinosteroids. *Plant Physiol.* **121,** 743–752 (1999).
33. Jin, H., Yan, Z., Nam, K. H. & Li, J. Allele-specific suppression of a defective brassinosteroid receptor reveals a physiological role of UGGT in ER quality control. *Mol. Cell* **26,** 821–830 (2007).
34. Fujioka, S. *et al.* The *Arabidopsis* deetiolated2 mutant is blocked early in brassinosteroid biosynthesis. *Plant Cell* **9,** 1951–1962 (1997).
35. Back, T. G. & Pharis, R. P. Structure-activity studies of brassinosteroids and the search for novel analogues and mimetics with improved bioactivity. *J. Plant Growth Regul.* **22,** 350–361 (2003).
36. Diévart, A., Hymes, M. J., Li, J. & Clark, S. E. Brassinosteroid-independent function of BRI1/CLV1 chimeric receptors. *Funct. Plant Biol.* **33,** 723–730 (2006).
37. Gruszka, D., Szarejko, I. & Maluszynski, M. New allele of *HvBRI1* gene encoding brassinosteroid receptor in barley. *J. Appl. Genet.* (published online, doi:10.1007/s13353-011-0031-7 (8 February 2011).
38. Friedrichsen, D. M., Joazeiro, C. A., Li, J., Hunter, T. & Chory, J. Brassinosteroid-insensitive-1 is a ubiquitously expressed leucine-rich repeat receptor serine/threonine kinase. *Plant Physiol.* **123,** 1247–1256 (2000).
39. Hink, M. A., Shah, K., Russinova, E., de Vries, S. C. & Visser, A. J. W. G. Fluorescence fluctuation analysis of *Arabidopsis thaliana* somatic embryogenesis receptor-like kinase and brassinosteroid insensitive 1 receptor oligomerization. *Biophys. J.* **94,** 1052–1062 (2008).
40. Russinova, E. *et al.* Heterodimerization and endocytosis of *Arabidopsis* brassinosteroid receptors BRI1 and AtSERK3 (BAK1). *Plant Cell* **16,** 3216–3229 (2004).
41. Han, B. W., Herrin, B. R., Cooper, M. D. & Wilson, I. A. Antigen recognition by variable lymphocyte receptors. *Science* **321,** 1834–1837 (2008).
42. Whippo, C. W. & Hangarter, R. P. A brassinosteroid-hypersensitive mutant of BAK1 indicates that a convergence of photomorphogenic and hormonal signaling modulates phototropism. *Plant Physiol.* **139,** 448–457 (2005).
43. He, K. *et al.* BAK1 and BKK1 regulate brassinosteroid-dependent growth and brassinosteroid-independent cell-death pathways. *Curr. Biol.* **17,** 1109–1115 (2007).
44. Karlova, R. *et al.* The *Arabidopsis* SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE1 protein complex includes BRASSINOSTEROID-INSENSITIVE1. *Plant Cell* **18,** 626–638 (2006).

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

Acknowledgements We thank J. Vanhnasy and W. Yu for maintaining insect cell stocks, M. Jinek and B. W. Han for advice, W. Kwiatowski for maintenance of the Salk X-ray equipment, Y. Jaillais for discussion, and F. V. Chisari for encouragement and support. This work was supported by the Howard Hughes Medical Institute and a grant from the National Science Foundation (IOS-0649389) to J.C. M.H. was supported by long-term fellowships from the European Molecular Biology Organisation and the International Human Frontier Science Program Organisation. Y.B. was a Howard Hughes Medical Institute fellow of the Life Sciences Research Foundation and also received support from the Philippe Foundation. I.A.W. was supported by NIH grant AI042266 and by the Skaggs Institute for Chemical Biology.

Author Contributions M.H., Y.B., J.P.N. and J.C. designed the project. M.H. expressed the BRI1 ectodomain in the laboratory of I.A.W. with initial help from Y.B. M.H. purified and crystallized the protein, and phased and refined the structures. M.D. determined viral titres and optimized production of viruses. T.D. cloned the modified transfer vector. M.H., I.A.W. and J.C. analysed the data. J.C. supervised the project. M.H. wrote the paper with input from the other authors.

Author Information Atomic coordinates and structure factors for the reported crystal structures have been deposited in the Protein Data Bank under accession numbers 3RIZ for the unliganded BRI1 ectodomain and 3RJ0 for the BRI1–brassinolide complex. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.C. (chory@salk.edu).

METHODS

Protein expression and purification. A synthetic gene comprising the entire BRI1 ectodomain (residues 29–788) and codon optimized for expression in *Trichoplusia ni* was synthesized by Geneart. The gene was cloned into a modified pBAC-6 transfer vector (Novagen), providing a glycoprotein 64 signal peptide and a C-terminal TEV (tobacco etch virus protease) cleavable Strep-9xHis tandem-affinity tag. Recombinant baculoviruses were generated by co-transfecting the transfer vector with linearized baculovirus DNA (ProFold-ER1, AB vector) and amplified in Sf9 cells. The fusion protein was expressed in Hi5 cells using a multiplicity of infection of 5, and harvested from the medium 4 days after infection by tangential flow filtration using a 30 kDa MWCO (molecular weight cut-off) filter membrane (GE Healthcare). BRI1 was purified by sequential Co^{2+} (His select gel, Sigma) and Strep (Strep-Tactin Superflow high-capacity, IBA) affinity chromatography. Next, the tandem-affinity tag was removed by incubating purified BRI1 with recombinant TEV protease in 1:100 molar ratio. The cleaved tag and the protease were separated from BRI1 by size-exclusion chromatography on a Superdex 200 HR10/30 column (GE Healthcare) equilibrated in 20 mM HEPES (pH 7.5), 100 mM NaCl, 1 mM EDTA. Monomeric peak fractions were concentrated to $\sim$15 mg ml^{-1} and snap frozen in liquid nitrogen. About 50–80 μg of purified BRI1 could be obtained from 1 l of insect cell culture.

Crystallization and data collection. Initial crystals of BRI1 appeared in 18% PEG 4,000, 0.8 M KCl using the counter diffusion method. Diffraction quality crystals of about 300 × 80 × 600 μm could be grown after multiple rounds of microseeding at room temperature by vapour diffusion in hanging drops composed of 1.25 μl of protein solution (15 mg ml^{-1}) and 1.25 μl of crystallization buffer (14% PEG 4,000, 0.2 M $(NH_4)_2SO_4$, 0.1 M citric acid pH 4.0) suspended above 1.0 ml of the mother liquor as the reservoir solution. For structure solution crystals were stabilized, derivatized and cryo-protected by serial transfer into 16% PEG 4,000, 1.7 M Na malonate (pH 4.0) and 0.5 M NaI, and cryo-cooled in liquid nitrogen. Single-wavelength anomalous diffraction (SAD) data to 2.9 Å resolution were collected on a Rigaku MicroMax rotating anode equipped with a copper filament (λ = 1.5418 Å), Osmic mirrors and an R-AXIS IV++ detector. Native crystals were transferred to a cryo-protective solution containing 16% PEG 4,000 and 1.7 M Na malonate (pH 4.0) and flash-cooled in liquid nitrogen. An isomorphous native data set to 2.5 Å was collected at beam-line 8.2.1 (λ = 0.9998Å) of the Advanced Light Source (ALS), Berkeley. The hormone-bound structure was obtained by dissolving brassinolide (Chemiclones Inc.) to a concentration of 1 mM in 100% DMSO. This stock solution was diluted to a final concentration of about 50 μM in protein storage buffer (20 mM HEPES pH 7.5, 100 mM NaCl, 1 mM EDTA). Purified BRI1 protein was added to a final concentration of about 12.5 μM (1.5 mg ml^{-1}) and the mixture was incubated at room temperature for 16 h. Next, the complex was re-concentrated to 18 mg ml^{-1}, and immediately used for crystallization. Crystals appeared under similar conditions as established for the unbound form and diffracted again to about 2.5 Å (λ = 1.5418 Å). Data processing and scaling was done with XDS[45] (version: May 2010) (Supplementary Table 1).

Structure solution and refinement. The program XPREP (Bruker AXS) was used to scale native and derivative data for SIRAS (single isomorphous replacement with anomalous scattering) analysis. Using data between 30 and 3.7 Å, SHELXD[46] located 52 iodine sites (CC All/Weak 42.50/19.82). 16 consistent sites were input into the program SHARP[47] for phasing and identification of 10 additional sites at 2.9 Å resolution (Supplementary Fig. 1a). Refined heavy atom sites and phases were input into phenix.resolve[48] for density modification and phase extension to 2.5 Å (final FOM was 0.55). The resulting electron density map was readily interpretable (Supplementary Fig. 1b), and the structure was completed in alternating cycles of model building in COOT[49] and restrained TLS refinement in phenix.refine (http://www.phenix-online.org). Refinement statistics are summarized in Supplementary Table 1. The crystals contain one BRI1 monomer per asymmetric unit with a solvent content of $\sim$60%. The final models comprise residues 29–771, with the C termini (residues 772–788) being completely disordered. The structure contains 25 LRRs as initially proposed[2], and not 24 LRRs as concluded from later modelling studies[10]. Loop residues 590, 637 and 638 in the island domain appear disordered in the unliganded structure. Amino acids whose side chains could not be modelled with confidence were truncated to alanine (2% of all residues). Analysis with MolProbity[50] suggested that both refined models have excellent stereochemistry, with the free form having 93.3% of all residues in the favoured region of the Ramachandran plot, and no outliers (Molprobity score is 2.2 corresponding to the 90th percentile for structures (N = 6,681) at 2.52 ± 0.25 Å resolution). The brassinolide complex structure has 92.7% of all residues in the favoured region of the Ramachandran plot and no outliers (Molprobity score is 2.3 corresponding to the 86th percentile for structures (N = 6,632) at 2.54 ± 0.25 Å resolution).

Size-exclusion chromatography. This was performed using a Superdex 200 HR 10/30 column (GE Healthcare) pre-equilibrated in 25 mM citric acid/sodium citrate buffer (pH 4.5), 100 mM NaCl. 100 μl of sample (5 mg ml^{-1}) was loaded onto the column and elution at 0.6 ml min^{-1} was monitored by ultraviolet absorbance at 280 nm. Incubation with brassinolide was performed as described in the crystallization section.

Homology modelling. Homology modelling of the AtBAK1 ectodomain (residues 27–227; Uniprot http://www.uniprot.org; accession Q94F62) was performed with the program MODELLER (http://www.salilab.org/modeller/) using the BRI1 and PGIP[30] (PDB 1OGQ) structures as template. Structure-based sequence alignments were done using T-COFFEE (http://www.tcoffee.org). BRI1 and BAK1 share $\sim$35%, PGIP and BAK1 share $\sim$31% sequence identity, with the LRR and N-cap consensus sequences being highly conserved.

45. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26,** 795–800 (1993).
46. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64,** 112–122 (2008).
47. 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).
48. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64,** 61–69 (2008).
49. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
50. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–W383 (2007).

ARTICLE

doi:10.1038/nature10178

Structural insight into brassinosteroid perception by BRI1

Ji She[1,2,3]*, Zhifu Han[1]*, Tae-Wuk Kim[4], Jinjing Wang[3], Wei Cheng[3], Junbiao Chang[5], Shuai Shi[5], Jiawei Wang[1], Maojun Yang[1], Zhi-Yong Wang[4] & Jijie Chai[1,3]

Brassinosteroids are essential phytohormones that have crucial roles in plant growth and development. Perception of brassinosteroids requires an active complex of BRASSINOSTEROID-INSENSITIVE 1 (BRI1) and BRI1-ASSOCIATED KINASE 1 (BAK1). Recognized by the extracellular leucine-rich repeat (LRR) domain of BRI1, brassinosteroids induce a phosphorylation-mediated cascade to regulate gene expression. Here we present the crystal structures of BRI1(LRR) in free and brassinolide-bound forms. BRI1(LRR) exists as a monomer in crystals and solution independent of brassinolide. It comprises a helical solenoid structure that accommodates a separate insertion domain at its concave surface. Sandwiched between them, brassinolide binds to a hydrophobicity-dominating surface groove on BRI1(LRR). Brassinolide recognition by BRI1(LRR) is through an induced-fit mechanism involving stabilization of two interdomain loops that creates a pronounced non-polar surface groove for the hormone binding. Together, our results define the molecular mechanisms by which BRI1 recognizes brassinosteroids and provide insight into brassinosteroid-induced BRI1 activation.

Ubiquitously distributed throughout the plant kingdom, brassinosteroids are a class of low-abundance phytohormones that have crucial roles in many aspects of growth and development[1,2]. Studies in *Arabidopsis* have led to the identification of a number of genes involved in brassinosteroid perception and signalling[3–9]. One of them, *BRI1*, when mutated, abolishes brassinosteroid-mediated responses of plants[3]. BRI1 belongs to a large family of plant LRR receptor-like kinases (RLKs) with more than 200 members in *Arabidopsis* and over 300 in rice[10]. LRR RLKs are characterized by an extracellular LRR domain, a single-pass transmembrane segment and a cytoplasmic kinase domain. BRI1 has been established as a bona fide receptor of brassinosteroids by genetic and biochemical investigations[3,11,12]. The most convincing evidence for this comes from a biochemical study showing that BRI1(LRR) is essential and sufficient for recognition of brassinosteroids[12]. A 70-residue island domain (residues 587–656) is indispensable for brassinosteroid recognition. Binding of brassinosteroids to BRI1 initiates a phosphorylation-mediated cascade, transducing the extracellular steroid signal to transcriptional programs[13,14].

A second gene named *BAK1* involved in brassinosteroid signalling also encodes an LRR RLK, albeit with only five extracellular LRRs[4,5]. Although not involved in brassinosteroid binding, BAK1 promotes brassinosteroid-induced signalling by physically associating with BRI1 (refs 4, 5, 15). Several models[2,4,5,8,16–19] of brassinosteroid-induced BRI1 activation converge at transphosphorylation within the BRI1–BAK1 complex, which involves brassinosteroid-enhanced BRI1 homodimerization[16], release of BRI1 KINASE INHIBITOR 1 (BKI1)[19] from BRI1, and BRI1–BAK1 heterodimerization[17–19]. It is still not fully understood how brassinosteroids are perceived by BRI1 to give rise to these events.

In the present study, we report the crystal structures of free BRI1(LRR) and its complex with brassinolide (the most active form of brassinosteroid) (Supplementary Table 1). The structures not only reveal the molecular mechanisms underlying brassinosteroid recognition by BRI1 and provide insight into brassinosteroid-induced BRI1 activation, but also explain the structure–activity relationship of brassinosteroids and serve as a foundation for the rational design of nonsteroidal mimetics.

A helical solenoid structure of BRI1(LRR)

The three-dimensional structure of brassinolide-free BRI1(LRR) (residues 30–589, 596–643 and 647–772) contains 25 LRRs as predicted[3], including 24 regular ones and an irregular one abutting the amino-terminal side. The LRRs packed in tandem assemble into a highly curved solenoid structure (Fig. 1a), with the overall rotation angle about the central axis approximately 360 degrees (Fig. 1b). Compared to the regular horseshoe-shaped structures of other LRR proteins, one notably distinct feature of the BRI1(LRR) structure is that it is exceptionally twisted, resulting in formation of a whole turn of right-handed helix with an inner diameter of about 30 Å. Relative to LRR1, LRR25 shifts about 60 Å along the central axis of the helix (Fig. 1a). An unprecedented structural feature of BRI1(LRR) is that it has an insertion domain anchored to the inner surface of the solenoid and running across six LRRs (Fig. 1a). The N-terminal capping domain (residues 30–70), consisting of one β strand (β1) and two α helices, is integrated into the LRR structure by forming an anti-parallel β sheet with the β strand from the irregular repeat, whereas the carboxy-terminal capping domain (residues 752–772) uses two short helices to tightly pack against the last repeat (Fig. 1a). Eight potential glycosylation sites (Supplementary Fig. 2) as defined by sufficient electron density were found in BRI1(LRR). An assignment of functions to them awaits further investigations.

Like canonical LRR proteins, the concave surface of the BRI1(LRR) solenoid is composed of a parallel β sheet comprising 25 continuously running parallel β strands. Compared to many other LRR proteins, however, BRI1(LRR) also possesses parallel but distorted β sheets following the inner β structure in most of the repeats (Fig. 1a, b). This is probably due to the plant-specific consensus sequence L/fXGxI/vP (X and x stand for polar and any amino acid, respectively)[20] of BRI1(LRR)s (Supplementary Fig. 2) found in many LRR RLKs, as a similar structural

[1]Key Laboratory for Protein Sciences of Ministry of Education School of Life Sciences, Tsinghua University, Beijing 100084, China. [2]College of Biological Sciences, Peking University, Beijing 100084, China. [3]National Institute of Biological Sciences, No. 7 Science Park Road, Beijing 102206, China. [4]Department of Plant Biology, Carnegie Institution for Science, Stanford, California 94305, USA. [5]Department of Chemistry, Zhengzhou University, Zhengzhou 450001, China.
*These authors contributed equally to this work.

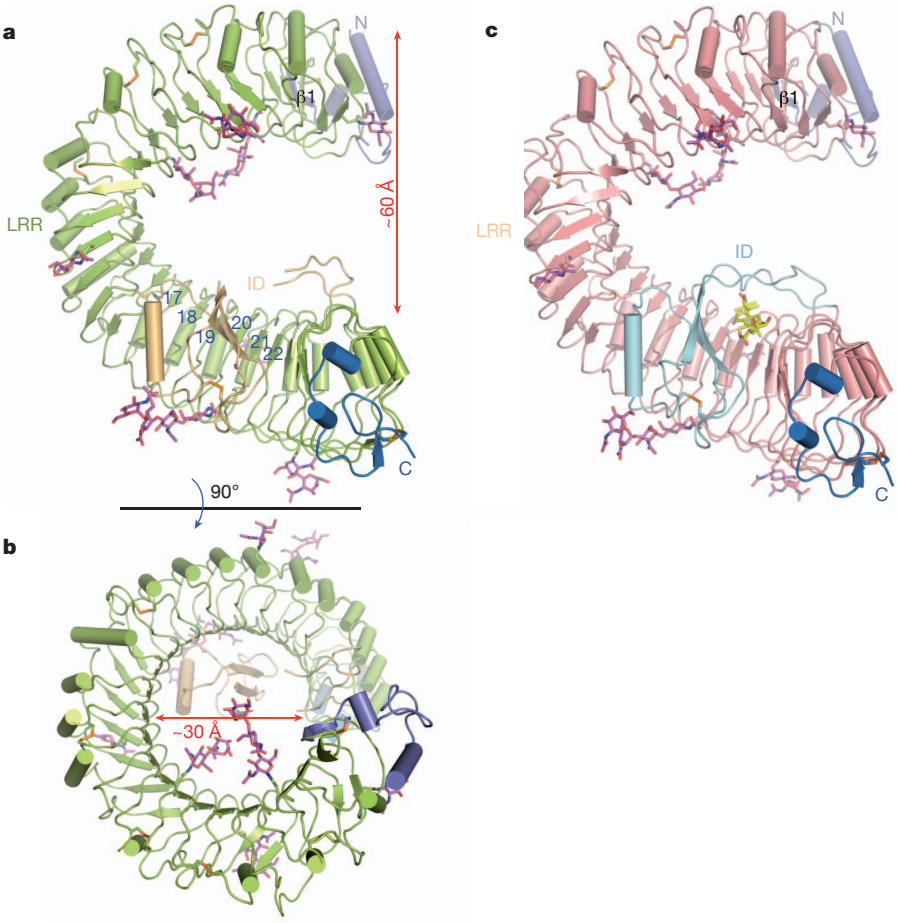


Figure 1 | BRI1(LRR) has a helical solenoid structure. **a**, **b**, Overall structures of brassinolide-free BRI1(LRR) shown in two different orientations. The N-linked sugars (N-acetylglucosamines) are shown in magenta stick representation. Coloured in orange are disulphide bonds. ID, insertion domain. The N- and C-terminal cap is shown in slate grey and marine blue, respectively. The blue numbers in **a** indicate the positions of LRRs. The β strand (β1) from the N-terminal cap is labelled. **c**, Overall structure of brassinolide-bound BRI1(LRR) with the same orientation as **a**. Brassinolide molecule is shown in stick representation and coloured in yellow.

feature also exists in the plant LRR protein PGIP2 (ref. 21). In contrast with the concave side, the convex outer surface consists of varied second structure elements, including 3_{10} helices, α helices and different length of loops (Fig. 1a, b). Some of the loops are stabilized through disulphide bonds formed between two consecutive repeats (LRR2–LRR3, LRR5–LRR6, LRR7–LRR8, LRR10–LRR11 and LRR14–LRR15) (Fig. 1a and Supplementary Fig. 1). The overall structure of BRI1(LRR) (residues 30–775) remains nearly unchanged on brassinolide binding, with a root mean squared deviation (r.m.s.d.) of 0.70 Å over 734 Cα atoms. The brassinolide molecule is sandwiched between the insertion domain and the concave surface of the solenoid structure (Fig. 1c).

Interruption of BRI1(LRR)s by a folded domain

Two short segments (residues 590–595 and 644–646) from the insertion domain do not have interpretable electron density in the structure of brassinolide-free BRI1(LRR) but become well defined after brassinolide binding (Supplementary Fig. 3). Residues 596–643 form a separate folded domain that consists of one three-stranded antiparallel β sheet and one 3_{10} helix (Fig. 2a). In addition, this domain contains one disulphide bridge (Cys 609–Cys 635) that can further contribute to its structural integrity. A database search using the Dali server (http://ekhidna.biocenter.helsinki.fi/dali_server) did not reveal known structures that share significant homology with the insertion domain, indicating that it represents a novel fold.

Extensive non-covalent interactions including van der Waals and polar interactions exist between the insertion domain and the solenoid structure. While making few contacts with the remaining parts of the insertion domain, the 3_{10} helix packs tightly against Phe 449 and Leu 473 from underneath and forms a large network of hydrogen bonding interactions with those from one flanking side (Fig. 2b). In contrast with the 3_{10} helix, β36 primarily binds the concave surface (Fig. 2a). Around this binding site, Trp 516, Ile 540, Trp 564 and Phe 658 establish close contacts with one side of β36, whereas Trp 472 and Phe 497 wedge between β36 and the 3_{10} helix (Fig. 2b). Hydrogen bonds dominate the interactions of the loop linking the 3_{10} helix and β36 with the concave surface. Additionally, packing of the carbonhydrate moiety of the glycosylated Asn 545 against the loop connecting the 3_{10} helix and β37 seems to have a role in positioning the insertion domain in the concave surface (Fig. 2a). Two mutations, G613S and S662F, generated weak hormone-insensitive phenotypes[2]. Our structure indicates that they can perturb local conformations and consequently generate a deleterious effect on BRI1 recognition of brassinosteroids. Substitution of Gly 613 with serine would produce steric clash with the carbonyl oxygen of Ile 600 or the benzene ring of Tyr 599 and consequently generate a damaging effect on the β sheet of the insertion domain (Supplementary Fig. 4). Ser 662 is limited within a small pocket and its mutation to the bulky residue phenylalanine is expected to generate serious clash with its neighbouring residues, in particular the Gly 611 that is N terminal to β36.

Brassinolide binds to a surface groove on BRI1(LRR)

We used ^{3}H-labelled brassinolide to test if the purified recombinant BRI1(LRR) was active in binding brassinolide. As shown in Fig. 3a, the protein showed a comparable brassinolide-binding activity to that

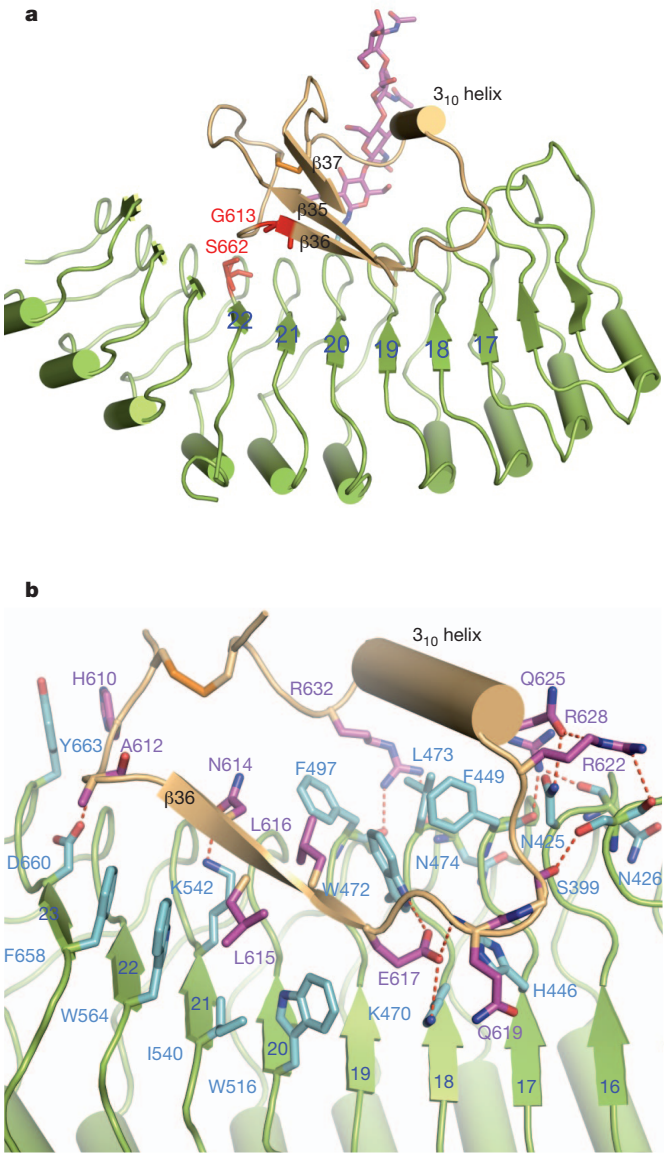


Figure 2 | Interaction of the insertion domain with LRRs. **a**, Cartoon representation of BRI1(LRR) highlighting interaction of the insertion domain with the concave surface. The positions of two genetic mutants are shown in red and stick representation. **b**, Detailed interactions of the insertion domain with LRRs. The side chains from the insertion domain are shown and labelled in purple, and those from LRRs in cyan. Red dashed lines represent hydrogen bonds.

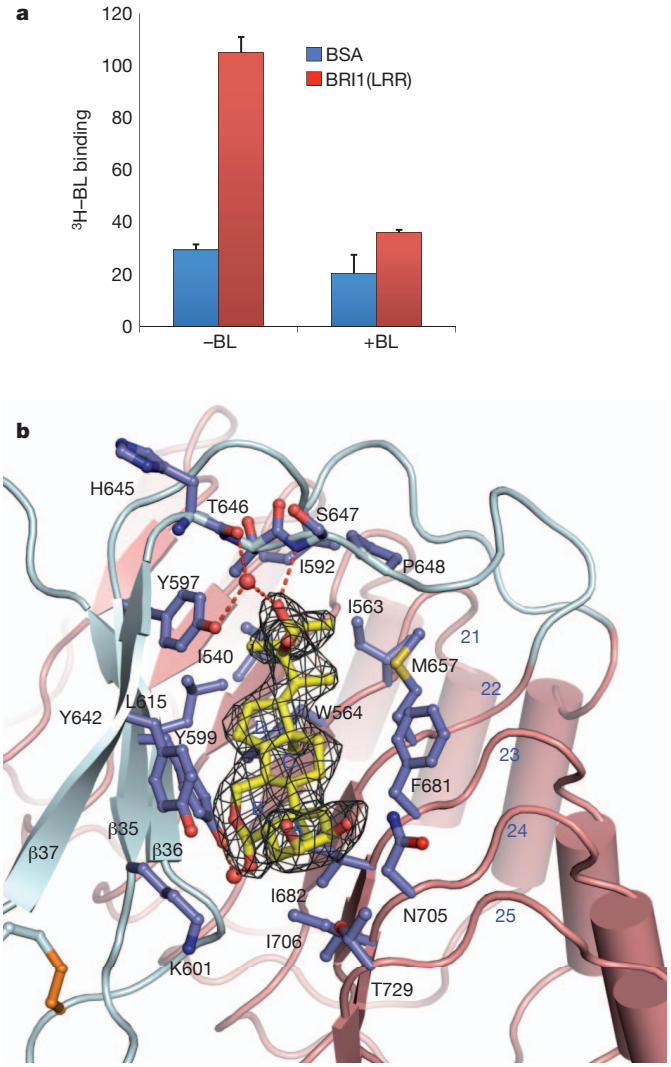


Figure 3 | Brassinolide binds a hydrophobic groove between the insertion domain and the inner surface of LRRs. **a**, ^{3}H-brassinolide (^{3}H-BL) binding activity of BRI1(LRR). About 1 mg ml^{-1} BRI1(LRR)–His (red bar) or bovine serum albumin (BSA) as control (blue bar) was incubated with 20 nM ^{3}H-brassinolide in the absence (−BL) or presence (+BL) of 20 μM unlabelled brassinolide. BRI1-bound ^{3}H-brassinolide was recovered using nickel beads and quantified by scintillation counting. Data represent the average of triplicate assays and error bars are standard deviations. **b**, Detailed interactions between brassinolide and BRI1(LRR). Shown in mesh is omit electron density (2.5δ) around brassinolide. The insertion domain and LRRs are coloured in slate grey and salmon pink, respectively. The side chains from both the insertion domain and LRRs are shown in slate grey. Red spheres represent oxygen atoms of water molecules. The three β strands from the insertion domain are labelled.

previously reported[12]. In the brassinolide-bound BRI1(LRR) structure, the electron density unambiguously defines the brassinolide molecule that binds to a pronounced hydrophobic groove between the insertion domain and the concave side of the solenoid (Fig. 3b). Most of the residues lining the surface groove are hydrophobic and highly conserved (Supplementary Fig. 1). The fused ring moiety of brassinolide (Supplementary Fig. 5) occupies most of the surface groove (Fig. 3b). The A ring (see Supplementary Fig. 5) makes marginal contacts with the concave surface, whereas the B ring tightly stacks Tyr 642 from one lateral side. The other two rings are sandwiched between Phe 681 and Tyr 599. Despite a smaller size, the two methyl groups C18 and C19 form much denser hydrophobic interactions with BRI1 by docking into two cavities at the bottom. In contrast, the other side of the fused ring is nearly solvent exposed (Fig. 3b). Although extensive shape complementarities exist between brassinolide and the base of the surface groove, a water molecule fills the cavity that is sealed from the solvent region by Lys 601, bridging a hydrogen bond between the carbonyl oxygen at C6 and the hydroxyl group of Tyr 599. Standing in contrast with the fused ring, the distal carbon atoms C24–C28 of the side chain are completely buried into a hydrophobic pocket (Fig. 3b). Further reinforcing the interactions around this contact interface, the hydroxyl group at C23 establishes a hydrogen bond with the backbone nitrogen of Ser 647 and two water-mediated hydrogen bonds with the carbonyl oxygen atoms of His 645 and the hydroxyl group of Tyr 597.

Induced-fit brassinolide recognition by BRI1(LRR)

The crystals of both free and brassinolide-bound BRI1(LRR)s contain one protein molecule per asymmetric unit and have nearly identical packing. This indicates that homodimers formed by two crystallographic symmetry-related molecules (Supplementary Fig. 6) result from crystal packing rather than brassinolide binding. Consistently,

gel filtration assays showed that BRI1(LRR) was monomeric in solution and its apparent molecular weight was not affected by the presence of brassinolide (Fig. 4a). The latter is consistent with the observation[22] that brassinolide had no effect on BRI1 homodimerization in protoplasts. Structural comparison revealed that upon brassinolide binding marked local structural rearrangement occurs to the two loops linking the insertion domain and the LRR structure (referred to as interdomain loops) (Fig. 4b). Brassinolide-induced stabilization of the interdomain loops results in formation of a more marked surface groove where brassinolide binds (Fig. 4c), showing that brassinolide recognition by BRI1(LRR) is a process of induced fit.

Discussion

The hydrophobicity-dominating brassinolide-binding surface groove (Fig. 4c, right) indicates that BRI1(LRR) may be accommodating in ligand binding. This would afford an explanation for why BRI1 is able to respond to a variety of brassinolide derivatives[23]. BRI1 ligand selectivity, on the other hand, can be conferred by the distal polar groups involved in hydrogen bonding interactions as well as the actual shape of the hormone-binding groove. The structural mechanism of brassinolide recognition by BRI1(LRR) rationalizes (Supplementary Fig. 7) the data on the structure–activity relationship of brassinosteroids accumulated during the past decades[23], thus opening new perspectives for designing and developing nonsteroidal mimetic of brassinosteroids with low cost that can be widely applied in agricultural practice.

Our structural analyses reveal that, although the overall structure of BRI1(LRR) remains nearly unchanged upon brassinolide binding, significant structural rearrangement occurs to the interdomain loops around the brassinolide-binding site (collectively referred to as the brassinolide-created surface), which should be associated with brassinosteroid-induced signalling. It is not clear how this region contributes to brassinosteroid-initiated signalling, but it is probably involved in stabilization[16] of the kinase-domain-mediated BRI1 homodimers[19]. The absence of an effect of brassinolide on BRI1 homodimerization in the current study may be due to the isolated ectodomain used for assays or requirement of another protein or peptide, as indicated by the suppression of *bri1* mutants by a secreted peptidase[24]. If this is the case, BRI1 would be different from Toll-like receptors, in which ligand-induced homodimerization for activation can be recapitulated through their isolated extracellular domains[25,26]. Stabilization of BRI1 homodimers by the brassinolide-created surface is expected to initially activate BRI1 kinase to a basal level, resulting in tyrosine phosphorylation and disassociation of BKI1 (ref. 19); dissociation of BKI1 allows stable association and sequential reciprocal phosphorylation between the kinase domains of BRI1 and BAK1, fully activating the BRI1 kinase as proposed by the sequential phosphorylation model[18]. Alternatively, brassinolide binding may initiate signalling by altering an interaction between the extracellular domains of BRI1 and BAK1. A role of the extracellular domain of BAK1 in brassinosteroid signalling has been indicated by the observations that mutations in this domain of BAK1 affect brassinosteroid sensitivity[27]

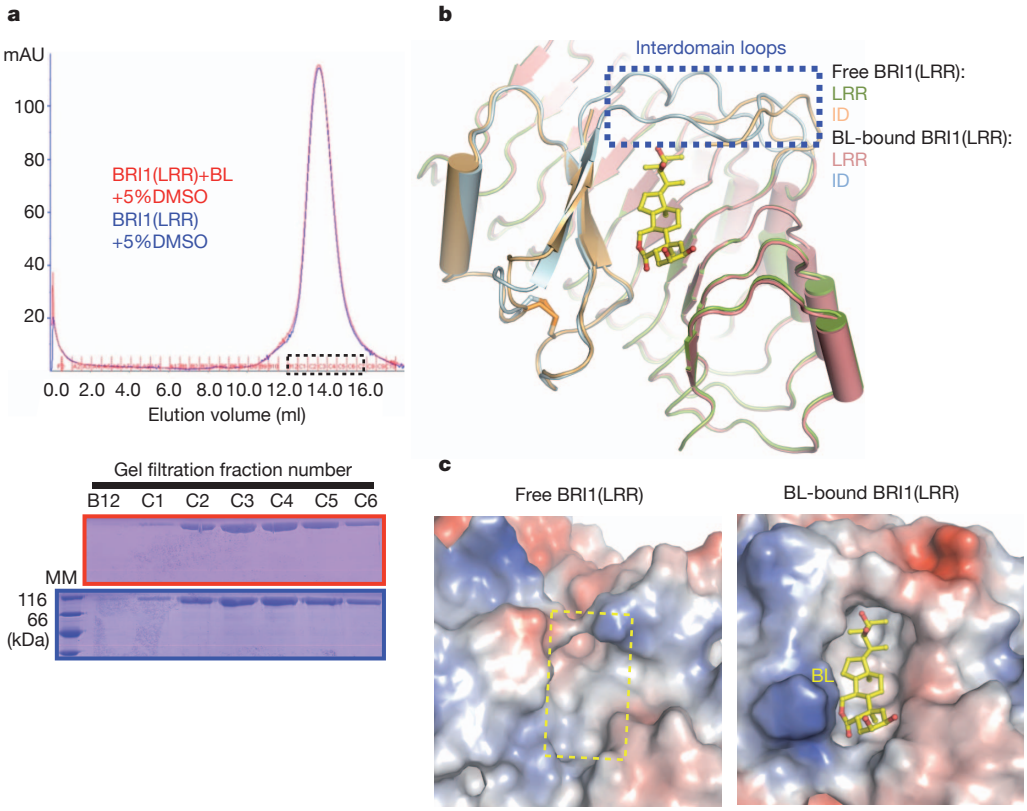


Figure 4 | Brassinolide induces stabilization of two interdomain loops but no dimerization of BRI1(LRR). **a**, Brassinolide (BL) has no effect on the oligomeric status of BRI1(LRR) in solution. Top, superposition of the gel filtration chromatograms of BRI1(LRR) in the absence (blue) and presence (red) of brassinolide. The vertical and horizontal axes represent ultraviolet absorbance ($\lambda = 280$ nm) and elution volume, respectively. Peak fractions are highlighted within the dashed square. The apparent molecular weight of BRI1(LRR) was 109.4 kDa in the presence or absence of brassinolide, higher than the theoretical BRI1(LRR) monomer (83 kDa), probably owing to the existence of multiple glycosylation sites in BRI1(LRR). DMSO, dimethylsulphoxide. Bottom, Coomassie blue staining of the peak fractions shown in the top of the panel after SDS–PAGE. MM, molecular mass marker. **b**, Brassinolide binding induces stabilization of two interdomain loops. Structural superimposition of the free and brassinolide-bound BRI1(LRR) around the brassinolide-binding site. **c**, Brassinolide binding generates a marked hydrophobic surface groove on BRI1(LRR). Shown on the left and right are the electrostatic surfaces of free BRI1(LRR) and brassinolide-bound BRI1(LRR) (shown in the same orientation) around the brassinolide-binding site, respectively. The area highlighted with the yellow dashed square on the left panel is the brassinolide-binding site. White, blue and red indicate neutral, positive and negative surfaces, respectively.

and interaction with BRI1[28]. Although stable BRI1–BAK1 association is a consequence of initial BRI1 activation[19], a possible role of BAK1 and its homologues in initial BRI1 activation has not been excluded by genetic studies; for example, it remains unclear whether mutations of BAK1 and its homologues affect brassinolide-induced BKI1 phosphorylation by BRI1. It is interesting to note that BAK1 has five LRRs and that the new brassinolide-created surface is also located about five LRRs from the membrane surface. Thus, it is also possible that the brassinolide-created surface is involved in interaction with BAK1(LRR), which triggers BRI1 phosphorylation of BKI1 through unknown mechanisms, allowing formation of more stable BRI1–BAK1 receptor complexes. Although the brassinolide-created surface is proposed to be involved in BRI1(LRR) homodimerization or heterodimerization with BAK1(LRR), further studies are needed to determine if brassinosteroids initiate signalling through promoting protein–protein interactions, as many other plant hormones do[29,30].

METHODS SUMMARY

The extracellular domain (residues 24–784) of *Arabidopsis* BRI1 fused with 6×His at the C terminus was expressed with high five cells and a baculovirus expression system. Protein was harvested from the media and purified using Ni-NTA followed by size-exclusion chromatography (Superdex200 10/300 GL column; GE Healthcare). Crystals of free BRI1 and brassinolide-bound BRI1 were grown using the hanging-drop vapour-diffusion method and their structures were determined using molecular replacement. ^{3}H-labelled brassinolide binding assay was performed as previously described[12].

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

Received 10 February; accepted 6 May 2011.
Published online 12 June 2011.

1. Clouse, S. D. & Sasse, J. M. BRASSINOSTEROIDS: essential regulators of plant growth and development. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **49,** 427–451 (1998).
2. Vert, G., Nemhauser, J. L., Geldner, N., Hong, F. & Chory, J. Molecular mechanisms of steroid hormone signaling in plants. *Annu. Rev. Cell Dev. Biol.* **21,** 177–201 (2005).
3. Li, J. & Chory, J. A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. *Cell* **90,** 929–938 (1997).
4. Li, J. *et al.* BAK1, an *Arabidopsis* LRR receptor-like protein kinase, interacts with BRI1 and modulates brassinosteroid signaling. *Cell* **110,** 213–222 (2002).
5. Nam, K. H. & Li, J. BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. *Cell* **110,** 203–212 (2002).
6. Li, J. & Nam, K. H. Regulation of brassinosteroid signaling by a GSK3/SHAGGY-like kinase. *Science* **295,** 1299–1301 (2002).
7. Sun, Y. *et al.* Integration of brassinosteroid signal transduction with the transcription network for plant growth regulation in *Arabidopsis*. *Dev. Cell* **19,** 765–777 (2010).
8. Wang, X. & Chory, J. Brassinosteroids regulate dissociation of BKI1, a negative regulator of BRI1 signaling, from the plasma membrane. *Science* **313,** 1118–1122 (2006).
9. Tang, W. *et al.* BSKs mediate signal transduction from the receptor kinase BRI1 in *Arabidopsis*. *Science* **321,** 557–560 (2008).
10. Shiu, S. H. *et al.* Comparative analysis of the receptor-like kinase family in *Arabidopsis* and rice. *Plant Cell* **16,** 1220–1234 (2004).
11. He, Z. *et al.* Perception of brassinosteroids by the extracellular domain of the receptor kinase BRI1. *Science* **288,** 2360–2363 (2000).
12. Kinoshita, T. *et al.* Binding of brassinosteroids to the extracellular domain of plant receptor kinase BRI1. *Nature* **433,** 167–171 (2005).
13. Li, J. Brassinosteroid signaling: from receptor kinases to transcription factors. *Curr. Opin. Plant Biol.* **8,** 526–531 (2005).
14. Kim, T. W. & Wang, Z. Y. Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annu. Rev. Plant Biol.* **61,** 681–704 (2010).
15. Russinova, E. *et al.* Heterodimerization and endocytosis of *Arabidopsis* brassinosteroid receptors BRI1 and AtSERK3 (BAK1). *Plant Cell* **16,** 3216–3229 (2004).
16. Wang, X. *et al.* Autoregulation and homodimerization are involved in the activation of the plant steroid receptor BRI1. *Dev. Cell* **8,** 855–865 (2005).
17. Wang, X. *et al.* Identification and functional analysis of *in vivo* phosphorylation sites of the *Arabidopsis* BRASSINOSTEROID-INSENSITIVE1 receptor kinase. *Plant Cell* **17,** 1685–1703 (2005).
18. Wang, X. *et al.* Sequential transphosphorylation of the BRI1/BAK1 receptor kinase complex impacts early events in brassinosteroid signaling. *Dev. Cell* **15,** 220–235 (2008).
19. J aillais, Y. *et al.* Tyrosine phosphorylation controls brassinosteroid receptor activation by triggering membrane release of its kinase inhibitor. *Genes Dev.* **25,** 232–237 (2011).
20. Kajava, A. V. Structural diversity of leucine-rich repeat proteins. *J. Mol. Biol.* **277,** 519–527 (1998).
21. Di Matteo, A. *et al.* The crystal structure of polygalacturonase-inhibiting protein (PGIP), a leucine-rich repeat protein involved in plant defense. *Proc. Natl Acad. Sci. USA* **100,** 10124–10128 (2003).
22. Hink, M. A., Shah, K., Russinova, E., de Vries, S. C. & Visser, A. J. Fluorescence fluctuation analysis of *Arabidopsis thaliana* somatic embryogenesis receptor-like kinase and brassinosteroid insensitive 1 receptor oligomerization. *Biophys. J.* **94,** 1052–1062 (2008).
23. Zullo, M. T. A. & Adam, G. Brassinosteroid phytohormones—structure, bioactivity and applications. *Braz. J. Plant Physiol.* **14,** 143–181 (2002).
24. Li, J., Lease, K. A., Tax, F. E. & Walker, J. C. BRS1, a serine carboxypeptidase, regulates BRI1 signaling in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **98,** 5916–5921 (2001).
25. Liu, L. *et al.* Structural basis of Toll-like receptor 3 signaling with double-stranded RNA. *Science* **320,** 379–381 (2008).
26. Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* **458,** 1191–1195 (2009).
27. Whippo, C. W. & Hangarter, R. P. A brassinosteroid-hypersensitive mutant of BAK1 indicates that a convergence of photomorphogenic and hormonal signaling modulates phototropism. *Plant Physiol.* **139,** 448–457 (2005).
28. Yun, H. S. *et al.* Analysis of phosphorylation of the BRI1/BAK1 complex in *Arabidopsis* reveals amino acid residues critical for receptor formation and activation of BR signaling. *Mol. Cells* **27,** 183–190 (2009).
29. Sheard, L. B. & Zheng, N. Plant biology: signal advance for abscisic acid. *Nature* **462,** 575–576 (2009).
30. Sheard, L. B. *et al.* Jasmonate perception by inositol-phosphate-potentiated COI1–JAZ co-receptor. *Nature* **468,** 400–405 (2010).

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

Acknowledgements We thank S. Huang and J. He at Shanghai Synchrotron Radiation Facility (SSRF) for assistance with data collection; J. Chory from the Salk Institute for providing the tritium-labelled brassinolide. This research was funded by the National Outstanding Young Scholar Science Foundation of National Natural Science Foundation of China grant no. 20101331722 to J. Chai and NIH R01GM066258 to Z.-Y.W.

Author Contributions J. Chai, Z.H., J.S. and Z.-Y.W. designed the experiments. The binding assay was performed by T.-W.K. and the other assays by J.S., Z.H., Jinjing W., W.C., Jiawei W., M.Y., S.S. and J. Chang. Data were analysed by J. Chai, Z.H., J.S. and Z.-Y.W. J. Chai, Z.-Y.W. and Z.H. wrote the paper.

Author Information The atomic coordinates and structure factors of BRI1(LRR) and the BRI1(LRR)–brassinolide complex have been deposited in the Protein Data Bank under the accession codes 3RGX and 3RGZ, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J. Chai (chaijj@mail.tsinghua.edu.cn).

METHODS

Protein expression and purification. The LRR domain of BRI1 (residues 24–784) from *Arabidopsis* with an engineered C-terminal 6×His tag was generated by standard PCR-based cloning strategy and its identity was confirmed by sequencing. The protein was expressed in high five cells using the vector pFastBac 1 (Invitrogen) with a modified N-terminal Hemolin peptide. One litre of cells (2.0×10^6 cells ml^{-1}) was infected with 20 ml baculovirus using a multiplicity of infection of 4 at 22 °C, and protein was harvested from the media after 48 h. The protein was purified using Ni-NTA (Novagen) and size-exclusion chromatography (Superdex 200, Pharmacia) in buffer (10 mM Tris, pH 8.0, 100 mM NaCl). Samples from relevant fractions were applied to SDS–PAGE and visualized by Coomassie blue staining. Protein purification was performed at 4 °C. For crystallization of BRI1(LRR), the purified protein was concentrated to about 3.0 mg ml^{-1} in buffer containing 10 mM Tris, pH 8.0, 100 mM NaCl.

Crystallization, data collection, structure determination and refinement. Crystals of BRI1(LRR) were generated by mixing the protein with an equal amount of well solution (1 µl) by the hanging-drop vapour-diffusion method. A mixture of BRI1(LRR) and brassinolide with a molar ratio of 1:10 was used for generating crystals of their complex. The initial buffer producing crystals of both free BRI1(LRR) and brassinolide-bound BRI1(LRR) contained 0.2 M Na_2SO_4 and 20% (w/v) polyethylene glycol (PEG) 3,350, which was further optimized by adding 10 mM trimethylamine-HCl. Crystals grew to their maximum size ($0.1 \times 0.1 \times 0.1$ mm^3) within 10 days at room temperature (20 °C).

All the diffraction data sets were collected at the Shanghai Synchrotron Radiation Facility (SSRF) at beam line BL17U1 using a CCD detector. Crystals of both brassinolide-free and brassinolide-bound BRI1(LRR) belong to space group C2 with one protein molecule per asymmetric unit. For data collection, the crystals were equilibrated in a cryoprotectant buffer containing reservoir buffer plus 20.0% (v/v) glycerol. The data were processed using HKL2000 (ref. 31). Molecular replacement (MR) with the program PHASER[32] was used to solve the crystal structures of both brassinolide-free and brassinolide-bound BRI1(LRR). The atomic coordinates of PGIP2 (ref. 21) (PDB code 1OGQ) and InlA[33] (PDB code 1O6V) were used as the initial searching model. The model from MR was built with the program COOT[34] and subsequently subjected to refinement by the program PHENIX[35]. The electron density for brassinolide in BRI1(LRR)–brassinolide complex crystal became apparent after refinement of BRI1(LRR). The final refined model contains residues 30–775 of BRI1(LRR) and one brassinolide molecule in brassinolide-bound BRI1(LRR), and residues 30–772 in brassinolide-free BRI1(LRR) except the two segments containing residues 590–595 and 644–646 that have no clear electron density and are presumed to be disordered in solution. The structure figures were prepared using PyMOL[36].

Gel filtration assay. BRI1(LRR) protein purified as described earlier was subjected to gel filtration analysis (Superdex200 10/300 GL column; GE Healthcare) in the presence and absence of brassinolide. Buffer containing 50 mM NaH_2PO_4/Na_2HPO_4 pH 7.4, 100 mM NaCl, 2.0 mM $MgCl_2$ and 5% (v/v) DMSO was used for the assay. A mixture of BRI1(LRR) and brassinolide with a molar ratio of about 1:10 was used to test the effect of brassinolide on the apparent molecular weight of BRI1(LRR). The assays were performed with a flow rate of 0.5 ml min^{-1} and an injection volume of 0.5 ml buffer containing BRI1(LRR) (0.83 mg ml^{-1}) with or without brassinolide at 20 °C. The column was calibrated using five standard proteins, β-amylase 200 kDa, alcohol dehydrogenase 150 kDa, albumin 66 kDa, carbonic anhydrase 29 kDa and cytochrome C 12.4 kDa, which had elution volumes (V_e) of 12.53 ml, 13.25 ml, 14.25 ml, 16.18 ml and 16.97 ml, respectively, under the conditions of assay. The void volume (V_o) for the column used was determined to be 8.15 ml using fresh blue dextran solution (1.0 mg ml^{-1}). The calibration curve of the gel-phase distribution coefficient (K_{av}) versus log molecular weight (M) was obtained using Excel. $K_{av} = (V_e - V_o)/(V_c - V_o)$, where V_e is the elution volume, V_o is the void volume, V_c is the geometric column volume, 24 ml for the column used. The elution volume of BRI1(LRR) in the presence and absence of brassinolide was 13.63 ml under the conditions of assay, corresponding to an apparent molecular weight of 109.4 kDa based on the calibration equation $K_{av} = -0.2366 \lg M + 1.538$, $R^2 = 0.9814$.

^{3}H-labelled brassinolide-binding assay. His-tagged BRI1(LRR), or BSA as control, was incubated with 20 nM ^{3}H-brassinolide, or 20 nM ^{3}H-brassinolide mixed with 20 µM non-radiolabelled brassinolide as competitor, and nickel beads in the brassinolide-binding buffer (0.25 M mannitol, 10 mM Tris-2-[N-morpholino] ethanesulphonic acid (MES), pH 5.7, 5 mM $MgCl_2$, 0.1 mM $CaCl_2$) for 30 min. The beads were washed three times, and His–BRI1(LRR) and bound [^{3}H]-brassinolide were eluted by using 1 M imidazole. The eluted radioactivity was measured by scintillation counter.

31. Otwinowski, Z., &. Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326 (1997).
32. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
33. Schubert, W. D. *et al.* Structure of internalin, a major invasion protein of *Listeria monocytogenes*, in complex with its human receptor E-cadherin. *Cell* **111,** 825–836 (2002).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60,** 2126–2132 (2004).
35. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58,** 1948–1954 (2002).
36. DeLano, W. L. PyMOL molecular viewer ⟨http://www.pymol.org⟩ (2002).

ARTICLE

doi:10.1038/nature10179

Condensin association with histone H2A shapes mitotic chromosomes

Kenji Tada[1,2], Hiroaki Susumu[1,2], Takeshi Sakuno[1,3] & Yoshinori Watanabe[1,2]

Chromosome structure is dynamically regulated during cell division, and this regulation is dependent, in part, on condensin. The localization of condensin at chromosome arms is crucial for chromosome partitioning during anaphase. Condensin is also enriched at kinetochores but its precise role and loading machinery remain unclear. Here we show that fission yeast (*Schizosaccharomyces pombe*) kinetochore proteins Pcs1 and Mde4—homologues of budding yeast (*Saccharomyces cerevisiae*) monopolin subunits and known to prevent merotelic kinetochore orientation—act as a condensin 'recruiter' at kinetochores, and that condensin itself may act to clamp microtubule binding sites during metaphase. In addition to the regional recruitment factors, overall condensin association with chromatin is governed by the chromosomal passenger kinase Aurora B. Aurora-B-dependent phosphorylation of condensin promotes its association with histone H2A and H2A.Z, which we identify as conserved chromatin 'receptors' of condensin. Condensin phosphorylation and its deposition onto chromosome arms reach a peak during anaphase, when Aurora B kinase relocates from centromeres to the spindle midzone, where the separating chromosome arms are positioned. Our results elucidate the molecular basis for the spatiotemporal regulation of mitotic chromosome architecture, which is crucial for chromosome partitioning.

Chromosomes carry genetic information and change their shape during the cell cycle. During mitosis, chromosomes are packaged into rod-shaped structures because the interphase (transcriptionally active) chromosomes are too diffuse to be separated along the spindle at anaphase. Condensin, a major protein component of mitotic chromosomes, is required for the assembly of rigid chromosomes and their separation[1–3]. Condensation of centromeric chromatin may also be important because depletion of condensin impairs the stiffness of centromeres and causes dysfunction of kinetochores[4–8]. However, this explanation was obtained by depleting whole cellular condensin rather than the centromere-specific pool, leaving the precise role of condensin at centromeres elusive.

Unlike cohesin, which embraces DNA topologically to establish sister-chromatid cohesion[9], condensin may regulate chromosome architecture through a dynamic (non-topological) interaction with chromatin[10,11]. Studies in budding yeast and fission yeast identified the transcription factor of polymerase III (Pol III) as a potential condensin loader[12,13], whereas in budding yeast condensin loading in rDNA regions depends on monopolin[14], a factor also required for mono-orientation of sister chromatids at meiosis I (ref. 15). In vertebrates, the non-catalytic action of PP2A and the methylation of histone H4 lysine 20 are implicated in the chromosomal association of non-canonical condensin (condensin II)[16,17], whereas these regulations are not applicable to canonical condensin (condensin I), which is conserved in all eukaryotic organisms.

Since the coincidence of histone phosphorylation and chromosome condensation during the cell cycle was recognized, it has been suspected that these two events are linked[1,18]. The conserved mitotic kinase Aurora B, which also phosphorylates histone H3S10, regulates several crucial mitotic events, including chromosome orientation and cytokinesis[19]. Aurora B is important for chromosome condensation or the association of condensin with chromatin in a wide range of species (including vertebrate condensin I)[8,20–27], whereas H3S10 phosphorylation might not be responsible for this function[28,29]. Thus, because the key target of Aurora B kinase in chromosome condensation has not been identified, the link between Aurora B and chromosome architecture remains a crucial issue[1,2,18].

Condensin may clamp microtubule binding sites

To address mechanisms of condensin loading to chromatin, we first dissected the functions of condensin at kinetochores and chromosome arm regions in fission yeast. By inactivating only the kinetochore pool of condensin in fission yeast, we demonstrate that condensin has a specific role at kinetochores in preventing merotelic attachment, most probably by assembling rigid kinetochores (Supplementary Fig. 1). The fission yeast kinetochore proteins Pcs1 and Mde4, which are homologues of budding yeast monopolin subunits Csm1 and Lrs4 but are dispensable for mono-orientation at meiosis I, are reportedly required for preventing merotelic attachment[15,30,31]. Thus, the Pcs1–Mde4 complex and condensin, both enriched at the kinetochore[8,31], might cooperate in a similar function. Supporting this assumption, we detected a physical interaction between the Pcs1–Mde4 complex and condensin (Supplementary Fig. 2) as well as synthetic growth defects in *pcs1Δ* and the condensin mutation *cut3-477* (Fig. 1a). To examine the localization interdependency, we performed a chromatin immunoprecipitation (ChIP) assay using prometaphase-arrested cells. Condensin is enriched at the kinetochore and rDNA region, and moderately along chromosome arms, as reported previously[8,12,13]. Notably, kinetochore condensin localization, but not pericentric localization, is diminished in *pcs1*Δ cells (Fig. 1b and Supplementary Fig. 3), whereas monopolin localization is intact in *cut3-477* cells (Supplementary Fig. 4a). Although condensin at the rDNA region is also diminished in *pcs1*Δ cells (Fig. 1b), no selective defects in anaphase separation of the rDNA region were observed in *pcs1*Δ or *cut3-477* cells (Supplementary Fig. 5). We confirmed that Aurora B localization, as well as its activity at centromeres, is intact in *pcs1*Δ cells (Supplementary Fig. 6), precluding

[1]Laboratory of Chromosome Dynamics, Institute of Molecular and Cellular Biosciences, University of Tokyo, Yayoi, Tokyo 113-0032, Japan. [2]Graduate Program in Biophysics and Biochemistry, Graduate School of Science, University of Tokyo, Yayoi, Tokyo 113-0032, Japan. [3]Promotion of Independence for Young Investigators, University of Tokyo, Yayoi, Tokyo 113-0032, Japan.

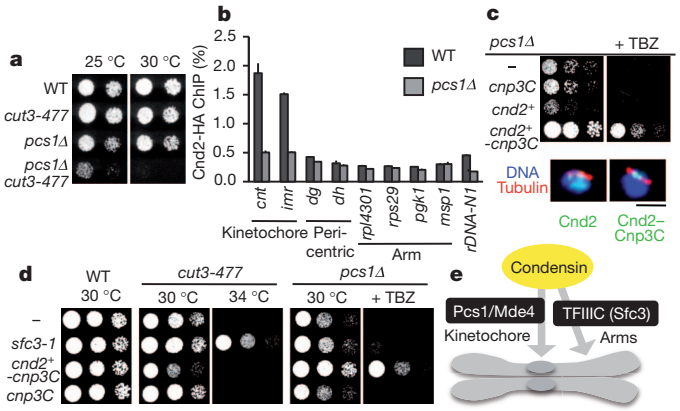


Figure 1 | Dissection of condensin function at the kinetochore and chromosome arm. **a**, Serial dilutions growth assay on YE plates. **b**, A ChIP assay was used to measure Cnd2-HA throughout kinetochore (*cnt* and *imr*), pericentric (*dg* and *dh*), arm (*rpl4301*, *rps29*, *pgk1* and *msp1*) and rDNA (*rDNA-N1*) regions in the indicated strains arrested at prometaphase by *nda3-KM311* inactivation (also see Supplementary Fig. 3). Error bars represent s.d. ($n = 3$ PCR amplifications). **c**, Serial dilutions growth assay on YE plates containing 0 or 10 μg ml^{-1} TBZ. Photos show the representative localization of Cnd2–tdTomato (left) and Cnd2–tdTomato–Cnp3C (right) at metaphase. Scale bar, 2 μm. **d**, Serial dilutions growth assay on YE plates. **e**, Schematic depiction illustrating distinct condensin recruitment at kinetochores and chromosome arms (bottom).

the possibility that merotelic attachment is caused by Aurora B dysfunction at centromeres.

To validate further the importance of condensin at kinetochores in the context of Pcs1–Mde4, we engineered a Cnd2–Cnp3C fusion protein, which targets kinetochores through the Cnp3C domain[32], even in *pcs1*Δ cells. The expression of this fusion protein largely suppresses the growth defect, sensitivity to thiabendazole (TBZ, a microtubule destabilizing drug) and the incidence of lagging chromosomes in *pcs1*Δ cells, whereas neither Cnd2 nor Cnp3C protein alone suppresses these phenotypes (Fig. 1c and Supplementary Fig. 7). These results suggest that condensin itself may have a role in clamping microtubule attachment sites to prevent merotelic attachment, the function previously attributed to monopolin[30,31].

Dissection of condensin function

The condensin mutant *cut3-477*, which impairs, at least, condensin localization along whole chromosome regions (Supplementary Fig. 4a), shows a defect mainly in anaphase separation of the chromosome arms[33] (Supplementary Fig. 5). Indeed, the growth defect in *cut3-477* cells is not suppressed by centromere tethering of condensin (Fig. 1d), indicating that condensin at the kinetochore is not rate-limiting for the viability of *cut3-477* cells. Recent studies suggest that condensin loading along the chromosome arms depends on the Pol III transcription factor TFIIIC, the targets of which are distributed along the chromosome arms including pericentric tRNA clusters[12,13,34]. As reported previously[12], the *cut3-477* mutant is suppressed by *sfc3-1*, a 'gain-of-function' mutation of TFIIIC, by which condensin localization is elevated along chromosome arms but not at the kinetochores and rDNA regions (Fig. 1d and Supplementary Fig. 8). In contrast, the growth defect and TBZ sensitivity of *pcs1*Δ cells are not suppressed by the *sfc3-1* mutation, whereas they are suppressed by an increase in kinetochore condensin (Fig. 1d). Taken together, these results indicate that the pools of condensin at kinetochores and along the chromosome arms, which are recruited respectively by the Pcs1–Mde4 complex and TFIIIC, are functionally separable (Fig. 1e).

Aurora-B-dependent condensin localization

Given that the mitotic kinase Aurora B may regulate condensin localization to chromatin in fission yeast[8,23], we re-investigated condensin

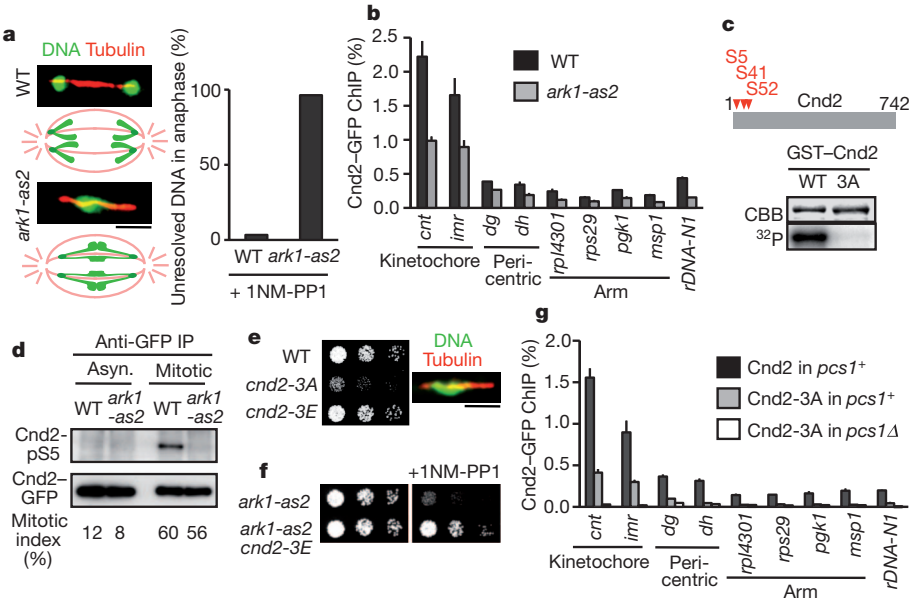


Figure 2 | Phosphorylation of condensin by Aurora B is required for chromatin association. **a**, Frequencies of unresolved DNA at anaphase ($n > 100$ cells) were examined in the indicated cells. **b**, Wild-type (WT) and *ark1-as2* cells expressing Cnd2–GFP were arrested at metaphase by the overexpression of *mad2*$^{+}$. For the last 1 h, cells were cultured in the presence of 1NM-PP1, then fixed and examined by ChIP assay. Error bars represent s.d. ($n = 3$ PCR amplifications). **c**, Schematic diagram of fission yeast Cnd2. Aurora B phosphorylation sites are indicated by arrowheads. GST–Cnd2 and GST–Cnd2-3A (alanine substitution at S5, S41 and S52) were phosphorylated by Ark1 and analysed for phosphate incorporation (^{32}P) and protein levels (Coomassie brilliant blue; CBB). **d**, Cnd2–GFP proteins were precipitated from asynchronous and mitotic (arrested by the overexpression of *mad2*$^{+}$) extracts of wild-type and *ark1-as2* cells, and analysed by immunoblot using anti-Cnd2-pS5 or anti-GFP antibodies. For the last 30 min of mitotic arrest, cells were cultured in the presence of 5 μM 1NM-PP1. The mitotic index was determined by counting nuclear Cnd2–GFP-positive cells. **e**, Serial dilution growth assay on YE plates. Photo shows a representative *cnd2-3A* cell at anaphase. **f**, Serial dilution growth assay on YE plates in the presence or absence of 0.2 μM 1NM-PP1. **g**, The indicated cells ectopically expressing Cnd2–GFP or Cnd2-3A–GFP were arrested at prometaphase by *nda3-KM311* inactivation and examined by ChIP assay. Error bars represent s.d. ($n = 3$ PCR amplifications). Scale bars, 2 μm (**a**, **e**).

loading using the ATP-analogue (1NM-PP1)-sensitive Aurora B mutant *ark1-as2* (ref. 35), which showed an extensive defect in chromosome separation at anaphase only when treated with 1NM-PP1 (Fig. 2a). A ChIP assay indicated that condensin localization decreases in *ark1-as2* cells in both the kinetochore and arm regions, whereas Mde4 localization is preserved intact (Fig. 2b and Supplementary Fig. 4b). Thus, Aurora B might act on condensin itself rather than the region-specific recruiter, Pcs1–Mde4 or TFIIIC. By performing *in vitro* phosphorylation of all the condensin subunits purified from bacteria, we identified the amino-terminal fragment of kleisin subunit Cnd2 (Cnd2(N)), which carries several Aurora B consensus sites[36] (Supplementary Fig. 9a). *In vitro* phosphorylation assays using mutant Cnd2 proteins indicate that S5, S41 and S52 are the major Aurora B phosphorylation sites (Fig. 2c and Supplementary Fig. 9b). A phospho-specific antibody raised against one of these sites, Cnd2-pS5 (Supplementary Fig. 10), recognized Cnd2 prepared from mitotic cell extract, and this phosphorylation was abolished by inactivation of Ark1 (Fig. 2d). We conclude that Cnd2 is phosphorylated depending on the mitotic kinase Aurora B *in vivo*.

To examine the functional significance of Cnd2 phosphorylation, we replaced the endogenous $cnd2^{+}$ with the non-phosphorylatable mutant *cnd2-3A*. The *cnd2-3A* mutant showed a severe growth defect, unresolved DNA signal during anaphase and a high incidence of lagging chromosomes, hallmarks of condensin defects (Fig. 2e and Supplementary Fig. 11). A ChIP assay indicates that the chromatin localization of Cnd2-3A is largely reduced at all tested chromosome loci (Fig. 2g). This defect is not due to the inability of the Cnd2-3A protein to form the condensin complex or to localize to the nucleus (Supplementary Fig. 12). Notably, the phospho-mimetic mutant *cnd2-3E* suppresses the growth defect of *ark1-as2* cells (Fig. 2f), indicating that the phosphorylation of condensin is the important essential target of Aurora B in fission yeast.

In human cells, chromatin association of condensin I is regulated by Aurora B (refs 25, 26) (Supplementary Fig. 13). Indeed, bacterially purified CAP-H (kleisin of condensin I), but not CAP-H2 (kleisin of condensin II), is phosphorylated by Aurora B *in vitro*; this phosphorylation is largely attributable to the serine 70 residue of CAP-H (Fig. 3a and Supplementary Fig. 14). To examine *in vivo* phosphorylation, we raised a phospho-specific antibody against CAP-H-pS70 (Supplementary Fig. 15). Immunoblot assays revealed that the phosphorylation of CAP-H at serine 70 is mitosis-specific and is Aurora-B-dependent (Fig. 3b), and that phosphorylated CAP-H is selectively enriched in the chromatin of mitotic cells (Fig. 3c). To examine the significance of Aurora-B-dependent phosphorylation of CAP-H for chromatin binding, we expressed an EGFP-tagged non-phosphorylatable CAP-H-S70A protein and compared it with its wild-type counterpart in HeLa cells (Fig. 3d). Whereas both the wild-type and mutant CAP-H proteins could form intact condensin complexes *in vivo* (Fig. 3e), the CAP-H-S70A protein showed a defect in chromatin localization. Treatment of mitotic cells with the Aurora B inhibitor caused a similar

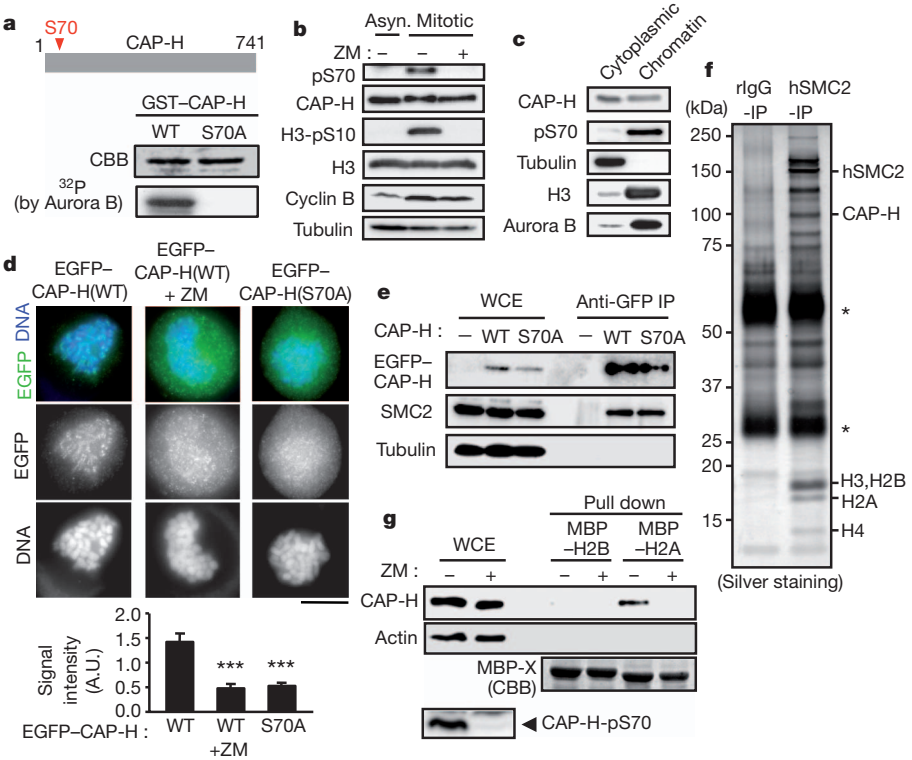


Figure 3 | Chromatin association of human condensin I is also regulated by Aurora-B-dependent phosphorylation. **a**, GST–CAP-H wild-type and GST–CAP-H-S70A were incubated with GST–Aurora B, and analysed for phosphate incorporation (^{32}P) and protein levels (CBB). **b**, Whole-cell extracts were obtained from HeLa cells treated with or without the Aurora inhibitor ZM447439 (ZM) at prometaphase or asynchronous cells, and analysed by immunoblot using the indicated antibodies. **c**, Cell extracts prepared from nocodazole-arrested cells were fractionated into chromatin and cytoplasmic fractions, and analysed by immunoblot. **d**, Synchronized mitotic HeLa cells expressing GFP–CAP-H were treated with nocodazole in the absence or presence of ZM447439 for 2 h. Signals of GFP–CAP-H at prometaphase were examined by immunostaining. Cells expressing GFP–CAP-H-S70A were similarly analysed without ZM447439 treatment. The ratios of nucleus and cytoplasmic signals were quantified in the indicated cells. Error bars represent s.e.m. ($n = 15$ cells). ***, $P < 0.001$. Scale bar, 5 μm. **e**, Co-immunoprecipitation of GFP–CAP-H and SMC2. Whole-cell extracts (WCE) were prepared from nocodazole-arrested HeLa cells expressing GFP–CAP-H wild type or GFP–CAP-H-S70A. GFP–CAP-H was immunoprecipitated (IP) with anti-GFP antibodies to examine the co-precipitation of SMC2. **f**, Chromatin fractions were prepared from nocodazole-arrested cells, precipitated with control or anti-SMC2 antibodies, and analysed by silver staining. Asterisks indicate IgG. **g**, Whole-cell extracts were prepared from mitotic HeLa cells treated with or without ZM447439, pulled down with MBP (maltose-binding protein)–H2A or MBP–H2B, and analysed by immunoblot using antibodies against CAP-H, CAP-H-pS70 or actin.

defect even in the wild-type CAP-H protein (Fig. 3d). We conclude that the association of condensin I with chromatin largely relies on the Aurora-B-dependent phosphorylation of the kleisin subunit in human cells, as in fission yeast.

Condensin associates with histone H2A

The foregoing results indicate that the Aurora-B-dependent phosphorylation of Cnd2/CAP-H might promote its interaction with one or more chromatin factors that locate commonly at the kinetochores and arm regions. To identify such factors, we immunoprecipitated condensin from mitotic HeLa cell extracts using anti-Smc2 antibodies. Silver staining of the immunoprecipitates detected most condensin subunits and several additional bands in the 10–20 kDa range, which correspond to histones (Fig. 3f and Supplementary Fig. 16). To test the possibility that a histone acts as the chromatin receptor of CAP-H, we performed a pull-down assay using recombinant CAP-H and histones. Remarkably, CAP-H, but not CAP-H2, interacted specifically with histone H2A and its variant H2A.Z *in vitro* (Supplementary Fig. 17a). Furthermore, histone H2A pulled down CAP-H from mitotic cell extracts, but failed to do so when cell extracts were prepared after the inhibition of Aurora B (Fig. 3g), indicating that CAP-H associates with H2A in an Aurora-B-dependent manner. Although condensin may associate with DNA[1,2], the detected interaction between H2A and condensin is not mediated by DNA (Supplementary Fig. 18). Although the depletion of H2A from cells is technically difficult, we were able to deplete H2A.Z by means of an RNA interference (RNAi) method reported previously[37]. Consequently, condensin localization was significantly decreased in H2A.Z RNAi prometaphase cells, although a substantial amount of condensin still localized on the chromatin, presumably because of intact H2A (Supplementary Fig. 17b). This result fits with the recent finding that the depletion of H2A.Z partially impairs condensin localization in fission yeast[38].

To delineate the significance of the interaction between the condensin kleisin subunit and H2A or H2A.Z, we performed comparable analyses in fission yeast, a genetically tractable organism. Fission yeast Cnd2 interacted with H2A and H2A.Z *in vitro*, and these interactions were enhanced when Cnd2 was phosphorylated by Aurora B (Fig. 4a, b and Supplementary Fig. 19a). The interaction with Cnd2 is mediated by

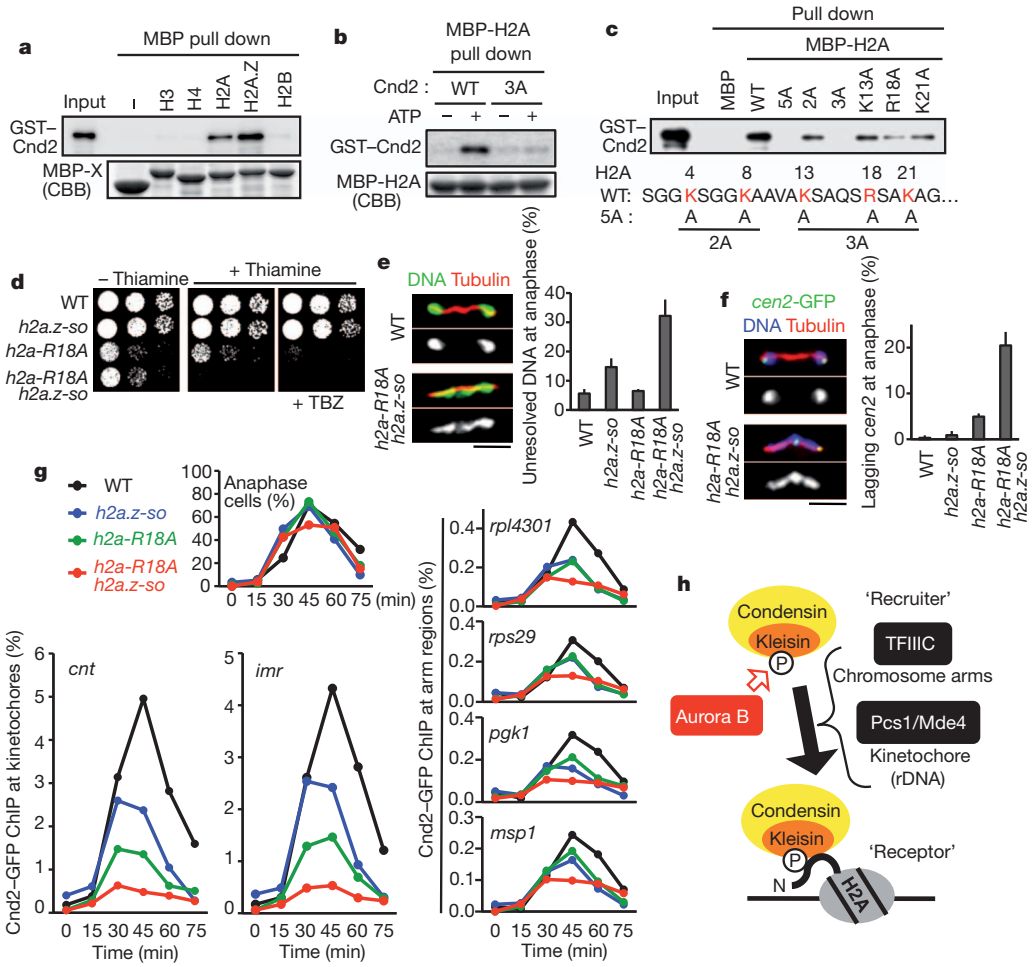


Figure 4 | H2A and H2A.Z act as a chromatin receptor of condensin. **a**, GST–Cnd2 phosphorylated by Ark1 was pulled down with the indicated MBP-fused proteins and analysed by immunoblot using anti-GST antibody. **b**, GST–Cnd2 wild type or GST–Cnd2-3A reacted with Ark1 in the absence or presence of ATP was pulled down with MBP–H2A. **c**, GST–Cnd2 phosphorylated by Ark1 was pulled down with the indicated MBP-fused proteins. Mutated residues within the N-terminal tail of H2A are shown in the diagram. **d**, Serial dilution growth assay on SSA plates containing 0 or 8 μM thiamine and 0 or 10 μg ml^{-1} TBZ. **e**, The indicated cells were cultured in SSA medium (−thiamine), transferred to YE medium (+thiamine) and incubated for 36 h at 30 °C to suppress the expression of H2A.Z. Frequencies of unresolved DNA at anaphase ($n > 100$ cells) were examined. Error bars represent s.d. ($n = 3$ experiments). **f**, The indicated cells cultured in YE medium for 24 h at 30 °C were shifted to 18 °C for 6 h. Frequencies of lagging chromosomes of *cen2*-GFP at anaphase ($n > 100$ cells) were examined. Error bars represent s.d. ($n = 3$ experiments). **g**, The indicated cells were arrested at G2 phase by *cdc25-22* inactivation and released into mitosis. Cells collected at each time point were examined by ChIP assay (average of two PCR amplifications). Anaphase cells were monitored by DAPI staining ($n > 100$ cells). **h**, Schematic depiction of the pathway that regulates condensin localization on chromatin. Scale bars, 2 μm (**e**, **f**).

the N-terminal tail of H2A or H2A.Z, which is enriched in basic amino acid residues (Fig. 4c and Supplementary Fig. 19b). Accordingly, an H2A-3A mutant protein in which three lysine/arginine residues (K13, R18 and K21) were replaced with alanines lost the affinity for phosphorylated Cnd2 (Fig. 4c). Among these three residues, arginine 18 seemed to be most important for the interaction with Cnd2.

H2A and H2A.Z act as condensin receptors

To explore the significance of the binding between Cnd2 and H2A *in vivo*, we tried to replace two fission yeast H2A genes (*hta1*$^+$ and *hta2*$^+$) with mutant alleles. The H2A-3A protein, which entirely abolishes the *in vitro* binding with Cnd2, causes cell lethality (not shown). Instead, we used mutant cells in which both histone H2A genes were replaced with the H2A-R18A mutant allele (called *h2a-R18A*, hereafter). The *h2a-R18A* cells showed a growth defect as well as sensitivity to TBZ (Fig. 4d). In contrast, depletion of the H2A.Z gene (*pht1*$^+$) resulted in no TBZ sensitivity, whereas this mutation showed synthetic lethality with the *h2a-R18A* mutation (not shown). To explore the synthetic defects, we replaced the *pht1*$^+$ promoter with a conditionally repressible promoter to shut off (so) the expression of H2A.Z (*h2a.z-so*), and crossed it with *h2a-R18A*. The double mutant showed lethality only under the H2A.Z shut-off condition (Fig. 4d), and exhibited unresolved DNA as well as lagging of *cen2*-GFP signals at anaphase, a typical phenotype of the condensin mutant (Fig. 4e, f). Similar, if not identical, defects were observed when *h2a.z-so* was combined with *h2a-R18Q*, another mutation erasing the positive charge at R18, but not with *h2a-R18K* (Supplementary Fig. 20). Taken together, these results indicate that H2A and H2A.Z act redundantly in condensin function by directly interacting with condensin.

To explore condensin loading in the histone mutants, we arrested cells at the G2/M boundary, and then monitored condensin occupancy kinetics at the kinetochore and chromosome arm regions through synchronous mitosis. In wild-type cells condensin loading took place through prophase to metaphase and culminated in anaphase (Fig. 4g). In *h2a-R18A* cells, however, condensin loading was significantly decreased in all tested regions throughout mitosis (Fig. 4g). A similar defect was observed in *h2a.z-so* cells, which showed the condensin loading defect particularly in anaphase, as previously reported in H2A.Z-depleted cells[38]. The decrease in condensin localization was much more marked in the *h2a-R18A h2a.z-so* double mutant than in either single mutant alone (Fig. 4g), indicating the redundancy of histone H2A and H2A.Z function in localizing condensin to chromatin. Notably, *h2a-R18A* cells reduced condensin and its function more at the kinetochores than the arm regions as compared with *h2a.z-so* cells (Fig. 4e–g). This can be explained by the fact that H2A.Z is less abundant around the centromeres[39] and so the contribution of H2A might be relatively larger in this region. We confirmed that the condensin loading defect in the *h2a-R18A* and *h2a.z-so* mutants is not due to decreased expression or defective complex formation of condensin (Supplementary Fig. 21). In contrast to condensin localization, the localization of the condensin recruiter Mde4 or cohesin is intact in *h2a-R18A* cells, indicative of the specificity (Supplementary Fig. 22). Taken together, these results support the notion that H2A and H2A.Z function as the chromatin receptor of condensin (Fig. 4h).

Spatiotemporal localization of condensin

The overall chromatin localization of condensin, which peaks in anaphase rather than metaphase (Fig. 4g), may reflect Aurora B activity towards condensin. Accordingly, phosphorylation of condensin by Aurora B culminates in anaphase (Fig. 5a). As a chromosomal passenger component, Aurora B localizes to centromeres in metaphase, and subsequently transfers to the spindle midzone in anaphase[19,23]. Because the redistributed Aurora B forms a special phosphorylation gradient centred at the spindle midzone[40], this gradient might efficiently phosphorylate chromatin-free condensin, especially around the separating chromosome arms. Indeed, if condensin association with chromatin is compared between metaphase and anaphase, the increase during anaphase is more pronounced at the arm regions than at the kinetochores (Supplementary Fig. 23).

The spatiotemporal condensin localization model implies that if chromosome separation is impeded along the chromosome arms at anaphase, condensin accumulation will be increased especially at the arm regions. To test this prediction, we used cells depleted for Taz1, a telomere binding protein, because at lower temperature this mutant shows entanglement at telomeres and interferes with chromosome separation in anaphase[41]. Indeed, *taz1*Δ cells show a transient anaphase bridging of chromosomes, although most chromosomes finally managed to separate presumably by pulling off the entanglement (Fig. 5b). Remarkably, condensin localization at anaphase increases along the chromosome arms in *taz1*Δ cells compared with wild-type cells, whereas the condensin profile at the kinetochores is invariable (Fig. 5b). Because accumulation of condensin I during anaphase is also reported in human cells[10,26], we compared the distribution of chromosomes and condensin I in mitotic HeLa cells. Immunostaining of early anaphase cells revealed that condensin I deposition onto chromatin is enhanced near the Aurora B sites (Supplementary Fig. 24). These results support the notion that Aurora B promotes chromosome separation near the spindle midzone at anaphase by efficiently localizing condensin to the chromosome arms (Fig. 5c).

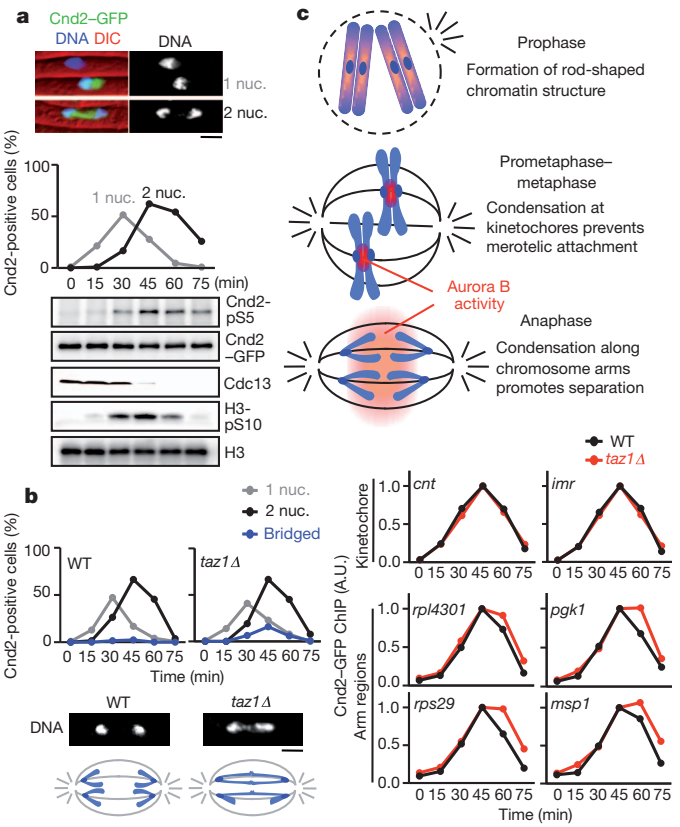


Figure 5 | Spatiotemporal regulation of condensin localization by Aurora B. **a**, Cells expressing Cnd2–GFP were arrested at G2 phase by *cdc25-22* inactivation and released into mitosis. Cell extracts prepared from cells collected at each time point were analysed by immunoblot using the indicated antibodies. Mitotic nuclear division was monitored by nuclear accumulation of Cnd2–GFP and DAPI staining ($n > 100$ cells) (top panels). DIC, differential interference contrast. **b**, Wild-type and *taz1*Δ cells expressing Cnd2–GFP were arrested at G2 phase by *cdc25-22* inactivation, released into mitosis and examined by ChIP assay. A representative anaphase bridge in a *taz1*Δ cell is shown compared with normal anaphase in a wild-type cell. **c**, Schematic depiction of the spatiotemporal regulation of condensin localization by Aurora B. Scale bars, 2 μm (**a**, **b**).

Discussion

Our results, together with previous findings[12], indicate that fission yeast condensin is enriched at kinetochores and along chromosome arms depending on region-specific recruiters: monopolin homologue Pcs1–Mde4 and Pol III transcription factor TFIIIC, respectively. Although recent studies have suggested that the Pcs1–Mde4/Csm1–Lrs1 complex itself functions in crosslinking kinetochore components[30,31], our results demonstrate that condensin can replace the complex in this role, at least in fission yeast, indicating that condensin itself may clamp kinetochore chromatin. Budding yeast monopolin, which is required for mono-orientation, may recruit casein kinase 1, rather than condensin, to kinetochores at meiosis I (refs 42, 43), indicating some functional divergence of this complex.

Our study further uncovers that condensin association with chromatin depends largely on the interaction with histone H2A and its variant H2A.Z, which we here define as the chromatin receptor of condensin. The interaction between condensin and the receptor is promoted when Aurora B phosphorylates the condensin subunit Cnd2. The Pcs1–Mde4 pathway and Cnd2 phosphorylation act synergistically to localize condensin at the kinetochores (Fig. 2g). Consistently, although the Pcs1–Mde4 function can be substituted by artificial tethering of Cnd2 to kinetochores, this tethering does not work with the Cnd2-3A protein, which lacks the ability to interact with H2A and H2A.Z (Supplementary Fig. 25). These results support the notion that condensin localization on chromosomes is controlled by two distinct layers of regulation: recruiting and receiving (Fig. 4h). H2A.Z is involved in diverse chromosomal processes such as transcriptional regulation and DNA repair[44]. Similarly, H2A R18, a crucial amino acid for binding with condensin, is implicated in transcriptional silencing and DNA repair[45], indicating that condensin might influence these processes[46].

Although previous studies have established that centromeric Aurora B destabilizes erroneous kinetochore–microtubule attachment[47], our current study suggests that centromeric Aurora B may also promote clamping kinetochore chromatin; both actions thus cooperatively prevent merotelic kinetochore orientation. Our results further indicate that the redistribution of Aurora B at anaphase may promote the phosphorylation of chromatin-unbound condensin around the spindle midzone, which will deliver condensin efficiently to chromosome arms, thereby facilitating their separation (Fig. 5c). These results reaffirm the crucial role of condensin in arm separation at anaphase[33,48] and explain why longer chromosomes condense more than shorter ones during anaphase[49].

METHODS SUMMARY

All *Schizosaccharomyces pombe* strains used in this study are listed in Supplementary Table 1. Methods to generate the strains are described in Methods. ChIP assays were conducted as described in Methods. For M phase arrest, we used the *nda3-KM311* mutation or *mad2*$^{+}$ overexpression. For G2/M arrest and release, we used the *cdc25-22* mutant cells cultured at 25 °C, shifted to 36 °C for 4 h to arrest them at G2/M, and released them into mitosis by shifting back to 25 °C. For inactivation of *ark1-as2*, cells were cultured in the presence of 5 μM 1NM-PP1 (1-naphthylmethyl-4-amino-1-tert-butyl-3-(*p*-methylphenyl)pyrazolo[3,4-d]pyrimidine) at 30 °C. HeLa cells were synchronized by double thymidine block and release, and incubated in the medium containing 330 nM nocodazole. Immunoprecipitation, immunofluorescence and immunostaining were performed as described in Methods. Statistical analysis was done with a two-tailed unpaired *t*-test.

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

Received 28 December 2010; accepted 11 May 2011.
Published online 1 June 2011.

1. Hirano, T. Condensins: organizing and segregating the genome. *Curr. Biol.* **15,** R265–R275 (2005).
2. Hudson, D. F., Marshall, K. M. & Earnshaw, W. C. Condensin: Architect of mitotic chromosomes. *Chromosome Res.* **17,** 131–144 (2009).
3. Koshland, D. & Strunnikov, A. Mitotic chromosome condensation. *Annu. Rev. Cell Dev. Biol.* **12,** 305–333 (1996).
4. Ono, T., Fang, Y., Spector, D. L. & Hirano, T. Spatial and temporal regulation of condensins I and II in mitotic chromosome assembly in human cells. *Mol. Biol. Cell* **15,** 3296–3308 (2004).
5. Oliveira, R. A., Coelho, P. A. & Sunkel, C. E. The condensin I subunit Barren/CAP-H is essential for the structural integrity of centromeric heterochromatin during mitosis. *Mol. Cell. Biol.* **25,** 8971–8984 (2005).
6. Samoshkin, A. *et al.* Human condensin function is essential for centromeric chromatin assembly and proper sister kinetochore orientation. *PLoS ONE* **4,** e6831 (2009).
7. Ribeiro, S. A. *et al.* Condensin regulates the stiffness of vertebrate centromeres. *Mol. Biol. Cell* **20,** 2371–2380 (2009).
8. Nakazawa, N. *et al.* Dissection of the essential steps for condensin accumulation at kinetochores and rDNAs during fission yeast mitosis. *J. Cell Biol.* **180,** 1115–1131 (2008).
9. Nasmyth, K. & Haering, C. H. Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* **43,** 525–558 (2009).
10. Gerlich, D., Hirota, T., Koch, B., Peters, J. M. & Ellenberg, J. Condensin I stabilizes chromosomes mechanically through a dynamic interaction in live cells. *Curr. Biol.* **16,** 333–344 (2006).
11. Hudson, D. F. *et al.* Molecular and genetic analysis of condensin function in vertebrate cells. *Mol. Biol. Cell* **19,** 3070–3079 (2008).
12. Iwasaki, O., Tanaka, A., Tanizawa, H., Grewal, S. I. & Noma, K. Centromeric localization of dispersed Pol III genes in fission yeast. *Mol. Biol. Cell* **21,** 254–265 (2010).
13. D'Ambrosio, C. *et al.* Identification of cis-acting sites for condensin loading onto budding yeast chromosomes. *Genes Dev.* **22,** 2215–2227 (2008).
14. Johzuka, K. & Horiuchi, T. The cis element and factors required for condensin recruitment to chromosomes. *Mol. Cell* **34,** 26–35 (2009).
15. Rabitsch, K. P. *et al.* Kinetochore recruitment of two nucleolar proteins is required for homolog segregation in meiosis I. *Dev. Cell* **4,** 535–548 (2003).
16. Takemoto, A. *et al.* The chromosomal association of condensin II is regulated by a noncatalytic function of PP2A. *Nature Struct. Mol. Biol.* **16,** 1302–1308 (2009).
17. Liu, W. *et al.* PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* **466,** 508–512 (2010).
18. Bazile, F., St-Pierre, J. & D'Amours, D. Three-step model for condensin activation during mitotic chromosome condensation. *Cell Cycle* **9,** 3243–3255 (2010).
19. Ruchaud, S., Carmena, M. & Earnshaw, W. C. Chromosomal passengers: conducting cell division. *Nature Rev. Mol. Cell Biol.* **8,** 798–812 (2007).
20. Giet, R. & Glover, D. M. *Drosophila* aurora B kinase is required for histone H3 phosphorylation and condensin recruitment during chromosome condensation and to organize the central spindle during cytokinesis. *J. Cell Biol.* **152,** 669–682 (2001).
21. Hagstrom, K. A., Holmes, V. F., Cozzarelli, N. R. & Meyer, B. J. C. *elegans* condensin promotes mitotic chromosome architecture, centromere organization, and sister chromatid segregation during mitosis and meiosis. *Genes Dev.* **16,** 729–742 (2002).
22. Kaitna, S., Pasierbek, P., Jantsch, M., Loidl, J. & Glotzer, M. The Aurora B kinase AIR-2 regulates kinetochores during mitosis and is required for separation of homologous chromosomes during meiosis. *Curr. Biol.* **12,** 798–812 (2002).
23. Petersen, J. & Hagan, I. M. S. *pombe* Aurora kinase/survivin is required for chromosome condensation and the spindle checkpoint attachment response. *Curr. Biol.* **13,** 590–597 (2003).
24. Lavoie, B. D., Hogan, E. & Koshland, D. *In vivo* requirements for rDNA chromosome condensation reveal two cell-cycle-regulated pathways for mitotic chromosome folding. *Genes Dev.* **18,** 76–87 (2004).
25. Takemoto, A. *et al.* Analysis of the role of Aurora B on the chromosomal targeting of condensin I. *Nucleic Acids Res.* **35,** 2403–2412 (2007).
26. Lipp, J. J., Hirota, T., Poser, I. & Peters, J. M. Aurora B controls the association of condensin I but not condensin II with mitotic chromosomes. *J. Cell Sci.* **120,** 1245–1255 (2007).
27. Mora-Bermúdez, F., Gerlich, D. & Ellenberg, J. Maximal chromosome compaction occurs by axial shortening in anaphase and depends on Aurora kinase. *Nature Cell Biol.* **9,** 822–831 (2007).
28. Hsu, J. Y. *et al.* Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* **102,** 279–291 (2000).
29. Murnion, M. E. *et al.* Chromatin-associated protein phosphatase 1 regulates aurora-B and histone H3 phosphorylation. *J. Biol. Chem.* **276,** 26656–26665 (2001).
30. Corbett, K. D. *et al.* The monopolin complex crosslinks kinetochore components to regulate chromosome-microtubule attachments. *Cell* **142,** 556–567 (2010).
31. Gregan, J. *et al.* The kinetochore proteins Pcs1 and Mde4 and heterochromatin are required to prevent merotelic orientation. *Curr. Biol.* **17,** 1190–1200 (2007).
32. Yokobayashi, S. & Watanabe, Y. The kinetochore protein Moa1 enables cohesion-mediated monopolar attachment at meiosis I. *Cell* **123,** 803–817 (2005).
33. Saka, Y. *et al.* Fission yeast cut3 and cut14, members of a ubiquitous protein family, are required for chromosome condensation and segregation in mitosis. *EMBO J.* **13,** 4938–4952 (1994).
34. Haeusler, R. A., Pratt-Hyatt, M., Good, P. D., Gipson, T. A. & Engelke, D. R. Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes. *Genes Dev.* **22,** 2204–2214 (2008).
35. Hauf, S. *et al.* Aurora controls sister kinetochore mono-orientation and homolog bi-orientation in meiosis-I. *EMBO J.* **26,** 4475–4486 (2007).
36. Koch, A., Krug, K., Pengelley, S., Macek, B. & Hauf, S. Mitotic substrates of the kinase Aurora with roles in chromatin regulation identified through quantitative phosphoproteomics of fission yeast. *Sci. Signal.* doi:10.1126/scisignal.2001588 (in the press).

37. Rangasamy, D., Greaves, I. & Tremethick, D. J. RNA interference demonstrates a novel role for H2A.Z in chromosome segregation. *Nature Struct. Mol. Biol.* **11,** 650–655 (2004).
38. Kim, H. S. *et al.* An acetylated form of histone H2A.Z regulates chromosome architecture in *Schizosaccharomyces pombe*. *Nature Struct. Mol. Biol.* **16,** 1286–1293 (2009).
39. Buchanan, L. *et al.* The *Schizosaccharomyces pombe* JmjC-protein, Msc1, prevents H2A.Z localization in centromeric and subtelomeric chromatin domains. *PLoS Genet.* **5,** e1000726 (2009).
40. Fuller, B. G. *et al.* Midzone activation of aurora B in anaphase produces an intracellular phosphorylation gradient. *Nature* **453,** 1132–1136 (2008).
41. Miller, K. M. & Cooper, J. P. The telomere protein Taz1 is required to prevent and repair genomic DNA breaks. *Mol. Cell* **11,** 303–313 (2003).
42. Brito, I. L., Yu, H. G. & Amon, A. Condensins promote coorientation of sister chromatids during meiosis I in budding yeast. *Genetics* **185,** 55–64 (2010).
43. Petronczki, M. *et al.* Monopolar attachment of sister kinetochores at meiosis I requires casein kinase 1. *Cell* **126,** 1049–1064 (2006).
44. Talbert, P. B. & Henikoff, S. Histone variants—ancient wrap artists of the epigenome. *Nature Rev. Mol. Cell Biol.* **11,** 264–275 (2010).
45. Parra, M. A. & Wyrick, J. J. Regulation of gene transcription by the histone H2A N-terminal domain. *Mol. Cell. Biol.* **27,** 7641–7648 (2007).
46. Aono, N., Sutani, T., Tomonaga, T., Mochida, S. & Yanagida, M. Cnd2 has dual roles in mitotic condensation and interphase. *Nature* **417,** 197–202 (2002).
47. Tanaka, T. U. Kinetochore-microtubule interactions: steps towards bi-orientation. *EMBO J.* **29,** 4070–4082 (2010).
48. Renshaw, M. J. *et al.* Condensins promote chromosome recoiling during early anaphase to complete sister chromatid separation. *Dev. Cell* **19,** 232–244 (2010).
49. Neurohr, G. *et al.* A midzone-based ruler adjusts chromosome compaction to anaphase spindle length. *Science* **332,** 465–468 (2011).

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

Acknowledgements We thank S. Hauf for critically reading the manuscript and K.-i. Noma, T. Hirano and the Yeast Genetic Resource Center (YGRC) for yeast strains or reagents. We also thank S. Hauf and M. Yanagida for communicating unpublished results, and all the members of our laboratory for their support and discussion. This work was supported in part by Global COE Program (Integrative Life Science Based on the Study of Biosignaling Mechanisms) (to K.T. and H.S.), Special Coordination Funds for Promoting Science and Technology (to T.S.) and a Grant-in-Aid for Specially Promoted Research, MEXT, Japan (to Y.W.).

Author Contributions K.T., supported by T.S., performed most of the experiments using fission yeast cells and proteins. H.S. performed all experiments using human cells or proteins. Experimental design and interpretation of data were conducted by all authors. Y.W. supervised the project and K.T. and Y.W. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.W. (ywatanab@iam.u-tokyo.ac.jp).

METHODS

***Schizosaccharomyces pombe* strain.** For the construction of TEV protease-cleavable Cnd2, the target site for TEV protease (Glu-Asn-Leu-Tyr-Phe-Gln-Gly(-Ala-Ser)[50]) was inserted between Val (amino acid 376) and Val (amino acid 377) in Cnd2-6HA (six copies of HA, abbreviated as HA hereafter) by site-directed mutagenesis. The resulting *cnd2(TEV)-HA* or control *cnd2-HA* construct carrying the 3′-untranslated region and the *kan*^{r} marker within it was integrated into the chromosome, replacing the endogenous *cnd2*$^+$ gene. To express Cnp3C-CFP-TEV, CFP-TEV, or Cnp3C-CFP protein[32], the sequence encoding them was cloned under the promoter P*adh21* (a weak version of the *adh1*$^+$ promoter) and integrated into the locus adjacent to the *zfs1*$^+$ gene of chromosome 2 (denoted by the *Z* locus) using the *nat*^{r} marker. Deletion of *pcs1*$^+$, *cnd2*$^+$, *hta1*$^+$, *hta2*$^+$, *taz1*$^+$ and *sgo2*$^+$, tagging of endogenous *mde4*$^+$, *cnd2*$^+$, *ark1*$^+$ and *cut3*$^+$ by Myc, GFP, or HA, and promoter exchange of endogenous *pht1*$^+$ by P*nmt81* (a weak version of the *nmt1*$^+$ promoter) were performed using the PCR-based gene targeting method for *S. pombe*[51]. To express Cnd2–tdTomato–Cnp3C, a sequence encoding tdTomato and Cnp3C (amino acids 384–642) was fused to the carboxy terminus of Cnd2 and cloned under the promoter P*adh31* (a weak version of the *adh1*$^+$ promoter). The resulting plasmid was linearized and integrated into the locus adjacent to the SPAC26F1.12c gene of chromosome 1 (denoted by the *C* locus) using the *hyg*^{r} marker. To generate the *cnd2-3A* strain, S5, S41 and S52 were changed to alanines using Kunkel methods. The genomic *cnd2* fragments carrying the mutations were then transformed into *cnd2::ura4*$^+$ cells (additional copy of *cnd2*$^+$ under the promoter P*adh15* was integrated at the *C* locus with *hyg*^{r} marker) and the integration at *cnd2::ura4*$^+$ was selected by 5-fluoroorotic acid (5-FOA) resistance and confirmed by PCR. The *cnd2-3E* strain was similarly constructed. For the analysis of Cnd2-3A localization, wild-type *cnd2* or *cnd2-3A*, fused with GFP, was cloned under the promoter P*adh21* and integrated into the *Z* locus using the *nat*^{r} marker. For generating the *h2a-R18A* strain, R18 of *hta1*$^+$ and *hta2*$^+$ were changed to alanines, using the PrimeSTAR Mutagenesis Basal Kit (Takara). The genomic *hta1* and *hta2* fragments carrying the mutations were then transformed into *hta1::ura4*$^+$ and *hta2::ura4*$^+$ cells, respectively, and integrants were selected for 5-FOA resistance and confirmed by PCR. After crossing *hta1-R18A* and *hta2-R18A* cells, double mutants (*h2a-R18A*) were selected. The *h2a-R18Q* or *h2a-R18K* strain was similarly constructed. To visualize tubulin, a sequence encoding mCherry or CFP was fused to the N terminus of *atb2*$^+$, cloned under the promoter P*adh13* or P*adh15* (a weak version of the *adh1*$^+$ promoter), and integrated into the *Z* or *C* locus using *nat*^{r} or *hyg*^{r} marker. To visualize nucleolus, Gar1–CFP (ref. 52) was expressed from the P*adh21* promoter (integrated at *C* locus using *hyg*^{r} marker). All strains used are listed in Supplementary Table 1.

Introduction of a *lacO* or *tetO* array into various chromosomal regions. A *lacO* array engineered from pCT31 (a gift from A. Yamamoto) or a *tetO* array engineered from p306*tetO*$_2$x112 (a gift from K. Nasmyth) was used in this study. For the visualization of *cnt1* and *dh1* (ref. 53), *lacO* arrays were introduced into these two regions simultaneously. To obtain *tel-lacO* and *arm-lacO* (centromeric side of the rDNA locus), a fragment containing a *lacO* array with the *kan*^{r} cassette, flanked by the regions adjacent to the telomeric region at the left arm of chromosome 1 (9282–9805 and 9806–10254 of chromosome 1) or to the centromeric side of the rDNA locus at right arm of chromosome 3 (2428525–2429283 and 2429284–2430083 of chromosome 3), was used for transformation, respectively. To obtain *imr1-tetO* and *imr3-tetO*, fragments containing *tetO* with the *ura4*$^+$ cassette, flanked by the *imr1* regions (3764007–3764631 and 3764632–3765171 of chromosome 1) or the *imr3* regions (1093757–1094520 and 1094521–1095451 of chromosome 3), were used for transformation, respectively. Correct integration was confirmed by PCR. Among symmetric sequences of *dh1*, *imr1*, or *imr3*, the integrated locus (left or right) was also determined by PCR.

Two-hybrid assay. The constructs of Pcs1 (amino acids 1–222) and Mde4 (amino acids 1–421) were amplified by PCR and cloned into pGBKT7 vectors and used as bait. The constructs of Cut14 (1–1172), Cut3(N) (1–477), Cut3(M) (468–882), Cut3(C) (873–1324), Cnd2(N) (1–250), Cnd2(M) (241–500), Cnd2(C) (491–742), Cnd1(N) (1–600), Cnd1(C) (591–1158) and Cnd3 (1–875) were amplified by PCR and cloned into pGADT7 vectors and used as prey. These plasmids were transformed into *Saccharomyces cerevisiae* AH109 strain. Plates lacking histidine and adenine and containing 2 mM 3AT (3-aminotriazole) were used as selective media.

Culture of fission yeast. For prometaphase arrest, we used the *nda3-KM311* mutation[54] and cultured cells at 17 °C for 13 h. For metaphase arrest, we used the pREP1(*mad2*$^+$) plasmids expressed by the *nmt1*$^+$ promoter (gift from T. Matsumoto) and cultured cells in the minimal medium at 30 °C for 14–16 h. For inactivation of *ark1-as2*, cells were cultured in the presence of 5 μM 1NM-PP1 (1-naphthylmethyl-4-amino-1-tert-butyl-3-(*p*-methylphenyl)pyrazolo[3,4-d]pyrimidine) at 30 °C. For the inactivation of *sfc3-1*, cells were cultured at 34 °C for the last 2 h of arrest. For the depletion of H2A.Z by the P*nmt81* promoter shut off, we cultured cells in YE medium (+thiamine) for 24 h at 25 °C. For G2/M arrest and release, we used the *cdc25-22* mutant cells cultured at 25 °C, shifted to 36 °C for 4 h to arrest them at G2/M, and released them into mitosis by shifting back to 25 °C. For the induction of telomere entanglement in *taz1Δ* cells, cells were cultured at 20 °C for 5 h just before shifting to 36 °C.

Preparation of phospho-specific antibodies. To generate phospho-specific antibodies against Cnd2-pS5, phosphorylated peptides (MKRA(pS)LGGH) were used to immunize rabbits (Operon). Antibodies were purified from anti-serum with phosphorylated peptide-conjugated CNBr-activated sepharose and dialysed against PBS. Similarly, for the production of phospho-specific antibodies against hCAP-H-pS70, phosphorylated peptides (LQRRR(pS)RVFDL) were used to immunize rabbits (Operon).

Chromatin immunoprecipitation assay. The procedure was carried out essentially as described previously[52]. Anti-HA antibodies (12CA5, Roche), anti-GFP antibodies (Living Colours Full-length A.v. Polyclonal Antibody, Clontech), anti-Myc antibodies (9E10, Santa Cruz), anti-H3-pS10 antibodies (Upstate), anti-Cnp1 antibodies[55] and anti-Psc3 antibodies[56] were used for immunoprecipitation. DNA prepared from whole-cell extracts or immunoprecipitated fractions was analysed by quantitative PCR with the ABI PRISM7000 system (Applied Biosystems) using SYBR Premix Ex Taq (Perfect Real Time) (Takara). The primers used for PCR were described previously[32,52,57] except those for *dg* (5′-CACATCATCGTCGTACTACAT-3′ and 5′-GATATCATCTATATTTAATGACTACT-3′), *dh* (5′-GAAAACACATCGTTGTCTTCAGAG-3′ and 5′-CGTCTTGTAGCTGCATGTGAA-3′), *rpl4301* (5′-GTATGGAAAAGCAAGGTACT-3′ and 5′-GATGGTGTATAAACTTTAGT-3′), *rps29* (5′-CGTAACTTGGTATGAAAGCA-3′ and 5′-TTAATTACCAGTAATCACAC-3′), *pgk1* (5′-TCAATAAGTTGGATCTTGTC-3′ and 5′-CTTATGAAGAAGGAGCTCGA-3′) and *rDNA-N1* (5′-GAAAAAGTCGAGCGAGTCG-3′ and 5′-TCAAAAGTGCATTACCCTTAC-3′). We included control IgG or untagged immunoprecipitation in each experiment to account for non-specific binding in the ChIP fractions. We normalized the immunoprecipitation efficiency among samples by performing Cnp1 immunoprecipitation. Error bars indicate s.d. of three PCR amplifications. We confirmed that all ChIP data are reproducible in separate experiments.

Co-immunoprecipitation from fission yeast extracts. Cultured cells were harvested, suspended in HB buffer (25 mM MOPS (pH 7.2), 15 mM $MgCl_2$, 15 mM EGTA, 60 mM β-glycerophosphate, 0.1 mM Na-orthovanadate, 0.1 mM NaF, 15 mM *p*-nitrophenylphosphate, 1% Triton X-100, 1 mM dithiothreitol, 1 mM PMSF, Complete protease inhibitor (Roche)), and disrupted using the Multi-bead shocker (Y. Kikai). Then the supernatants, collected after centrifugation, were incubated with anti-Myc antibody (9E10, Santa Cruz) or anti-GFP polyclonal antibodies (Living Colours Full-length A.v. Polyclonal Antibody, Clontech) for 1.5 h at 4 °C. Protein A beads (Amersham) were added and incubation was continued for 1.5 h at 4 °C. After washing with HB buffer, we analysed the immunoprecipitates by SDS–PAGE and western blotting with anti-HA (1:1,000, 12CA5, Roche), anti-GFP (1:1,000, Roche), anti-Myc (1:1,000, 9E10, Santa Cruz), anti-Cnd2-pS5 (1:1,000), anti-Cdc13 (1:1,000), anti-H3 (1:1,000, Abcam), anti-H3-pS10 (1:1,000, Upstate), anti-Psc3 (ref. 56) (1:1,000) and TAT1 (1:5000) antibodies.

***In vitro* kinase and pull-down assays of fission yeast proteins.** The constructs of Ark1 (amino acids 1–355), Pic1IN-box (914–1018), Cut14(N) (1–620), Cut14(C) (611–1172), Cut3(N) (1–477), Cut3(M) (468–882), Cut3(C) (873–1324), Cnd2(N) (1–250), Cnd2(M) (241–500), Cnd2(C) (491–742), Cnd1(N) (1–600), Cnd1(C) (591–1158) and Cnd3 (1–875) were amplified by PCR and cloned into pGEX4T-1, pGEX4T-2, or pGEX4T-3 vectors. For kinase assay, kinases and substrates were incubated with kinase buffer (50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 1% Triton X-100) at 30 °C for 30 min in the presence of [γ-^{32}P]ATP. Incorporation of the radioactive phosphate group was visualized by autoradiography and protein loading was analysed by staining with Coomassie blue (CBB). For pull-down assay, MBP and MBP-fused proteins were purified from *E. coli* and bound to amylose beads (NEB). GST-fused proteins, pre-incubated with GST–Ark1 and GST–Pic1IN in kinase buffer (50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 0.1% Triton X-100) at 30 °C for 30 min with or without ATP, were mixed with beads and incubated in binding buffer (20 mM Tris-HCl (pH 7.5), 1 mM EDTA, 200 mM NaCl, 10% glycerol, 0.8% NP-40, 1 mg ml^{-1} BSA, Complete protease inhibitor (Roche)) at 4 °C for 1.5 h. After washing with binding buffer, the bound proteins were analysed by western blotting with anti-GST antibodies (1:500, Santa Cruz). Protein loading was analysed by staining with CBB.

Time-lapse imaging. For the measurement of the distance between *imr3*-tdTomato signals during anaphase (Supplementary Fig. 5), wild-type, *cut3-477* or *pcs1*Δ cells were cultured in minimal medium at 30 °C for 4 h, then live cell recordings were performed on a DeltaVision P system (Applied Precision) in an

air-conditioned chamber maintained at 30 °C. A glass-bottomed dish (Matsunami) coated with 0.2% concanavalin A (Sigma) was used to mount cells. During observation, the cells transferred to the glass-bottomed dish were supplied with minimal medium. Images were acquired by *Z*-sectioning with 45-s intervals, and deconvoluted in the softWoRx software (Applied Precision). To measure the distance of *imr3*-tdTomato signals, softWoRx software (Applied Precision) was used.

Quantification of fluorescent signals in fission yeast. To quantify the fluorescent signals at centromeres, in-focus images of Ark1–GFP cells were taken with Axio Vision imaging software (Carl Zeiss). We measured the average intensity of the centromeric dots and subtracted the average background intensity.

Culture, synchronization and RNAi of HeLa cells. HeLa cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, 0.03% L-glutamine, 100 units ml^{-1} penicillin, and 100 mg ml^{-1} streptomycin in a moist atmosphere. Nocodazole (Sigma) and thymidine (Sigma) were added at 330 nM and 2 mM, respectively, for synchronization. RNAi was performed as described[58]. Sequences of siRNA against *hH2A.Z*, 5′-GAAGAAAGGACAACA GAAGdTdT-3′ and siRNA against *BubR1*, 5′-CGGGCAUUUGAAUAUGAA AdTdT-3′.

Expression of GFP–hCAP-H. Full-length hCAP-H cDNA was cloned from HeLa RNA by PCR. Full-length wild type and hCAP-H-S70A tagged with GFP at their N termini were inserted into pcPURO vectors derived from Y. Goto's laboratory for transient expression. Vectors were transfected into HeLa cells using Lipofectamine LTX (Invitrogen).

***In vitro* kinase and pull-down assays of human proteins.** GST-fused and MBP-fused proteins were bacterially expressed and purified with glutathione-sepharose beads (GE) and amylase resin (NEB), respectively. For kinase assays, reactions were carried out in 30 μl of kinase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 0.5% Triton X-100 supplemented with Complete protease inhibitor (Roche), 1 mM dithiothreitol, 1 mM PMSF) in the presence of [γ-^{32}P]ATP. Incorporation of radioactive phosphate group was visualized by autoradiography and protein level was analysed by staining with CBB. For pull-down assay, purified proteins were phosphorylated with GST–Aurora B in kinase buffer in the presence of 0.25 mM ATP at 30 °C for 45 min. Phosphorylated proteins were mixed with MBP-tagged proteins beads and incubated in binding buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 10% glycerol, 0.8% NP40 supplemented with Complete protease inhibitor, 1 mM dithiothreitol, 1 mg ml^{-1} BSA) at 4 °C for 3 h. The beads were then washed five times with binding buffer. The bead-bound proteins were subjected to SDS–PAGE and transferred onto the Immunobilon PVDF membrane (Millipore) and detected by immunoblot using anti-GST (1:500, Santa Cruz). Protein loading was analysed by staining with CBB.

Preparation of HeLa cell extracts and immunoprecipitation. HeLa cells were synchronized by single thymidine block and release, and incubated in the medium containing nocodazole for 16 h. Mitotic cells were collected by mitotic shake-off and washed with PBS once, then re-suspended in CSK buffer (10 mM PIPES pH 6.8, 10% glycerol, 3 mM $MgCl_2$, supplemented with Complete protease inhibitor, PhosSTOP phosphatase inhibitor mixture (Roche), 1 mM dithiothreitol, 0.25 mM PMSF) containing 100 mM NaCl. After incubation at 4 °C for 30 min, supernatants (chromatin-unbound) were extracted and the insoluble chromatin pellet was collected by centrifugation at 16,100*g* for 10 min, washed once with PBS and re-suspended in CSK buffer containing 250 mM NaCl, 1 mM $CaCl_2$ and digested with micrococcal nuclease (Sigma) at 30 °C for 15 min. After centrifugation at 16,100*g* for 10 min, supernatants (chromatin-bound) were prepared. Whole-cell extracts (WCE)—the mixture of chromatin-bound and chromatin-unbound fractions—were immunoblotted with the following antibodies: anti-hCAP-H-pS70 (1:3000), anti-H3-pS10 (1:1,000, Cell Signaling), anti-cyclin B (1:1,000, Santa Cruz), anti-tubulin (1:5,000, DM1A; Sigma), anti-Aurora B (1:500, BD Bioscience), anti-GFP (1:1,000, Roche). The chromatin-bound fraction was rotated with anti-SMC2 (Abcam) or rabbit control IgG for 1 h at 4 °C, followed by 1.5 h rotation with protein A sepharose beads (GE) at 4 °C. The beads were washed five times with CSK buffer containing 250 mM NaCl. We analysed the immunoprecipitates by silver staining using the SilverQuest Staining kit (Invitrogen) or by immunoblot with anti-SMC2 (1:5,000), anti-CAP-H (1:2,000, Bethyl), anti-H2A (1:1,000, Abcam), anti-H2B (1:1,000, Millipore), anti-H3 (1:1,000, Abcam), anti-H2A.Z (1:1,000, Cell Signaling) and anti-actin (1:1,000, Santa Cruz) antibodies.

Preparation of ZM447439-treated cell extracts. HeLa cells were synchronized by double thymidine block and release, and incubated in medium containing 1,000 nM nocodazole for 14 h in the presence of 4 μM ZM447439 during the last 2 h. Mitotic cells were collected by mitotic shake-off.

Immunostaining of HeLa cells. Immunofluorescence staining was performed as described[58] with anti-GFP (1:1,000, Invitrogen), anti-H3-pS10 (1:1,000), anti-CAP-H (1:500), biotinylated anti-CAP-G (1:500, a gift from T. Hirano), anti-H2A.Z (1:200) antibodies. DNA was stained with 3 mg ml^{-1} Hoechst 33342 (Wako). For pre-extraction, the cells were permeabilized in 0.1% Triton X-100 in XBE2 (10 mM HEPES pH 7.7, 2 mM $MgCl_2$, 100 mM KCL and 5 mM EGTA) at room temperature for 1 min and then fixed with 2% paraformaldehyde in XBE2 for 15 min. Immunostaining images were obtained with the use of DeltaVision softWoRx software (Applied Precision). Image segmentation into chromatin and cytoplasm was done using the Hoechst signal. Average cytoplasm signals were subtracted from chromatin signal and normalized to the Hoechst signal as reported[26]. Signal intensities of Aurora B, CAP-H and DNA were measured using the ImageJ program.

50. Dougherty, W. G., Cary, S. M. & Parks, T. D. Molecular genetic analysis of a plant virus polyprotein cleavage site: a model. *Virology* **171,** 356–364 (1989).
51. Bähler, J. *et al.* Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14,** 943–951 (1998).
52. Yokobayashi, S., Yamamoto, M. & Watanabe, Y. Cohesins determine the attachment manner of kinetochores to spindle microtubules at meiosis I in fission yeast. *Mol. Cell. Biol.* **23,** 3965–3973 (2003).
53. Sakuno, T., Tada, K. & Watanabe, Y. Kinetochore geometry defined by cohesion within the centromere. *Nature* **458,** 852–858 (2009).
54. Hiraoka, Y., Toda, T. & Yanagida, M. The *NDA3* gene of fission yeast encodes β-tubulin: a cold-sensitive *nda3* mutation reversibly blocks spindle formation and chromosome movement in mitosis. *Cell* **39,** 349–358 (1984).
55. Takayama, Y. *et al.* Biphasic incorporation of centromeric histone CENP-A in fission yeast. *Mol. Biol. Cell* **19,** 682–690 (2008).
56. Nonaka, N. *et al.* Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast. *Nature Cell Biol.* **4,** 89–93 (2002).
57. Kawashima, S. A. *et al.* Shugoshin enables tension-generating attachment of kinetochores by loading Aurora to centromeres. *Genes Dev.* **21,** 420–435 (2007).
58. Kitajima, T. S. *et al.* Shugoshin collaborates with protein phosphatase 2A to protect cohesin. *Nature* **441,** 46–52 (2006).

LETTER

doi:10.1038/nature10090

X-ray illumination of the ejecta of supernova 1987A

J. Larsson[1], C. Fransson[1], G. Östlin[1], P. Gröningsson[1], A. Jerkstrand[1], C. Kozma[1], J. Sollerman[1], P. Challis[2], R. P. Kirshner[2], R. A. Chevalier[3], K. Heng[4], R. McCray[5], N. B. Suntzeff[6], P. Bouchet[7], A. Crotts[8], J. Danziger[9], E. Dwek[10], K. France[11], P. M. Garnavich[12], S. S. Lawrence[13], B. Leibundgut[14], P. Lundqvist[1], N. Panagia[15,16,17], C. S. J. Pun[18], N. Smith[19], G. Sonneborn[10], L. Wang[20] & J. C. Wheeler[21]

When a massive star explodes as a supernova, substantial amounts of radioactive elements—primarily ^{56}Ni, ^{57}Ni and ^{44}Ti—are produced[1]. After the initial flash of light from shock heating, the fading light emitted by the supernova is due to the decay of these elements[2]. However, after decades, the energy powering a supernova remnant comes from the shock interaction between the ejecta and the surrounding medium[3]. The transition to this phase has hitherto not been observed: supernovae occur too infrequently in the Milky Way to provide a young example, and extragalactic supernovae are generally too faint and too small. Here we report observations that show this transition in the supernova SN 1987A in the Large Magellanic Cloud. From 1994 to 2001, the ejecta faded owing to radioactive decay of ^{44}Ti as predicted. Then the flux started to increase, more than doubling by the end of 2009. We show that this increase is the result of heat deposited by X-rays produced as the ejecta interacts with the surrounding material. In time, the X-rays will penetrate farther into the ejecta, enabling us to analyse the structure and chemistry of the vanished star.

Owing to the proximity of SN 1987A (located only 160,000 light yr away), we can study the evolution of the supernova in great detail. The central ejecta are surrounded by a ring of circumstellar material (Fig. 1) that was shed from the star 20,000 yr before the explosion in 1987[4]. Since the explosion, the ejecta have been expanding, and now the outer parts of the ejecta are colliding with the ring, causing it to brighten at all wavelengths[5–8]. The dense, central part of the ejecta contains most of the mass from the disrupted star and acts as a calorimeter for the energy input to the supernova. We have determined the energy input by tracking the energy output with the NASA Hubble Space Telescope (HST).

Because the ejecta are roughly elliptical in projection on the sky, we used an elliptical aperture to measure the brightness. To monitor a constant mass of expanding material, we allowed the measuring aperture to expand linearly with time. The axes of the aperture were therefore three times larger in 2009 than in 1994 (Fig. 1). Using this aperture, we determined the R- and B-band light curves of the ejecta, as shown in Fig. 2 (see Supplementary Table 1 and Supplementary Information, section 1, for further details of the observations and light curves). Our measurements show that the flux from the ejecta decays during the first ~5,000 d after the explosion, as expected from radioactive input, but then starts to increase, reaching a level that is two to three times higher at around day 8,000 (end of 2009). A new energy source must be present in addition to radioactive decay. Below, we consider a model for the declining phase and then discuss the new energy source that is responsible for the observed increase in flux.

The energy input to the declining phase of the light curve after ~1,500 d is expected to come from positrons produced in the decay of ^{44}Ti (refs 2, 9, 10). To test this, we use a model[2] with abundances taken from the 14E1 explosion model[11] and a ^{44}Ti mass of $1.4 \times 10^{-4} M_{\odot}$ (ref. 12; Supplementary Information, section 3), where $M_{\odot}$ is the solar mass. The model is shown in Fig. 3 together with the observed broadband luminosities. The good agreement with the observations up to day 5,000 confirms that the ^{44}Ti positrons provide the energy input up to this point. However, after day 5,000 the model fails to describe the light curve; radioactive decay cannot explain the increase in flux that we observe.

One possible origin for the flux increase is the reverse shock that results from the interaction between the ejecta and the H II region inside the ring[13–16]. The reverse shock produces strong Lyα and Hα emission, which increased by a factor of ~1.7 between 2004 and 2010[16]. Although most of this emission originates close to the ring, there is also a component of projected high-velocity Hα emission that can be traced to the central parts of the ejecta[16] and which would therefore contribute to the flux we measure. To determine the contribution of the reverse shock to our light curves, we have examined HST Space Telescope Imaging Spectrograph spectra from 2004 and 2010 (Supplementary Information, section 2, and Supplementary Fig. 5). The reverse shock can be isolated in the spectra because of its boxy line profile, allowing us to place a limit on its contribution at $\lesssim$20%. Furthermore, this changes only marginally between 2004 and 2010, as the expanding measuring aperture remains well inside the area where most of the shock emission is seen. Importantly, an increase in flux is also seen in the [Ca II] doublet lines at rest wavelengths 7,292 Å and 7,324 Å between 2000 and 2010 (determined from Ultraviolet and Visual Echelle Spectrograph observations at the Very Large Telescope; Fig. 2). These lines have speeds of $\lesssim$5,000 km s^{-1}, implying that they originate in the inner ejecta (the projected ejecta speed near the edge of the ring is $\gtrsim$7,000 km s^{-1} at the present time). We conclude that the increase in flux occurs primarily in the inner ejecta and cannot be explained by emission from the shock region.

We believe that the strong X-ray flux produced in the ring collision is the dominant source of energy input to the ejecta. The X-ray flux from the ring increased by a factor of ~3 in the 0.5–10-keV band between day 6,000 and day 8,000[6], similar to what we find for the optical emission from the ejecta. To investigate this, we calculated the fraction of X-rays absorbed by the ejecta from a point source located at the ring, using the partially mixed 14E1 explosion model[17]. As shown in Supplementary Fig. 6, most of the observed X-ray flux is

[1]Department of Astronomy, The Oskar Klein Centre, Stockholm University, 106 91 Stockholm, Sweden. [2]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, MS-19, Cambridge, Massachusetts 02138, USA. [3]Department of Astronomy, University of Virginia, Charlottesville, Virginia 22903, USA. [4]Eidgenössische Technische Hochschule Zürich, Institute for Astronomy, Wolfgang-Pauli-Strasse 27, CH-8093 Zürich, Switzerland. [5]JILA, University of Colorado, Boulder, Colorado 803090-0440, USA. [6]George P. and Cynthia Woods Mitchell Institute for Fundamental Physics and Astronomy, Texas A&M University, Department of Physics and Astronomy, College Station, Texas 77843, USA. [7]DSM/IRFU/Service d'Astrophysique Commissariat á l'Energie Atomique et aux Energies Alternatives, Saclay, Orme des Merisiers, FR 91191 Gif-sur-Yvette, France. [8]Department of Astronomy, Mail Code 5240, Columbia University, 550 West 120th Street, New York, New York 10027, USA. [9]Osservatorio Astronomico di Trieste, Via Tiepolo 11, Trieste 34131, Italy. [10]NASA Goddard Space Flight Center, Code 665, Greenbelt, Maryland 20771, USA. [11]Center for Astrophysics and Space Astronomy, University of Colorado, Boulder, Colorado 80309, USA. [12]225 Nieuwland Science, University of Notre Dame, Notre Dame, Indiana 46556-5670, USA. [13]Department of Physics and Astronomy, Hofstra University, Hempstead, New York 11549, USA. [14]ESO, Karl-Schwarzschild-Strasse 2, 85748 Garching, Germany. [15]Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. [16]INAF-CT, Osservatorio Astrofisico di Catania, Via S. Sofia 78, I-95123 Catania, Italy. [17]Supernova Limited, OYV #131, Northsound Road, Virgin Gorda, British Virgin Islands. [18]Department of Physics, University of Hong Kong, Pok Fu Lam Road, Hong Kong, China. [19]Steward Observatory, University of Arizona, 933 North Cherry Avenue, Tucson, Arizona 85721, USA. [20]Department of Physics and Astronomy, Texas A&M University, College Station, Texas 77843-4242, USA. [21]Department of Astronomy, University of Texas, Austin, Texas 78712-0259, USA.

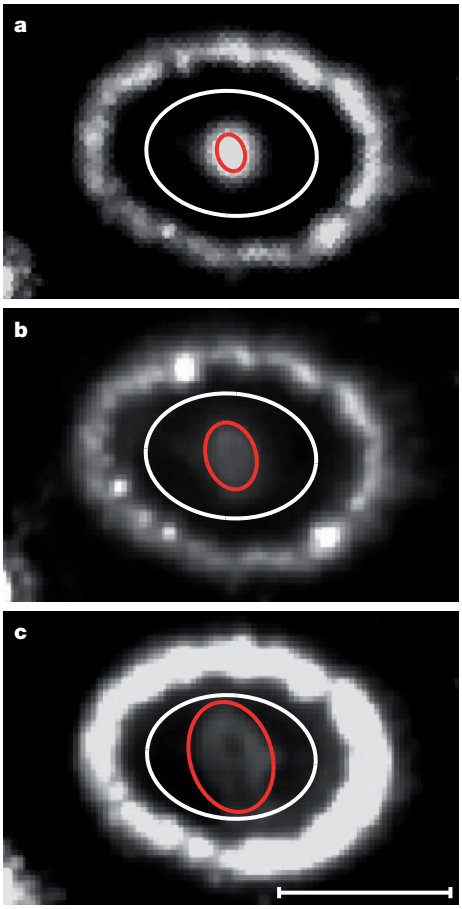


Figure 1 | HST R-band images. The observing dates are 1994 September 24 (**a**), 2000 November 13 (**b**) and 2009 April 29 (**c**), which correspond to 2,770, 5,012 and 8,101 d after the explosion, respectively. The scale bar in **c** represents 1″. The circumstellar ring is inclined at an angle of 45° with respect to the line of sight and is approximately 1.3 light yr across. The red ellipse shows the expanding aperture used for the light curve in Fig. 2. By using an initial semi-major axis of 0.11″ for the observation in 1994, we always follow the bright, central part of the ejecta without being significantly affected by emission from the circumstellar ring. The white ellipse shows the fixed aperture used for one of the light curves in Supplementary Fig. 2. The R-band emission from the ejecta is dominated by Hα emission with a small contribution from [Ca I] and [Ca II] lines, whereas the B band (Supplementary Fig. 1) is dominated by H I, Fe I and Fe II lines[12,22]. Only the densest, central parts of the ejecta are visible, owing to the low surface brightness of the outer parts. In reality, the ejecta extend to the ring, as is evident from the strong interaction with the ring.

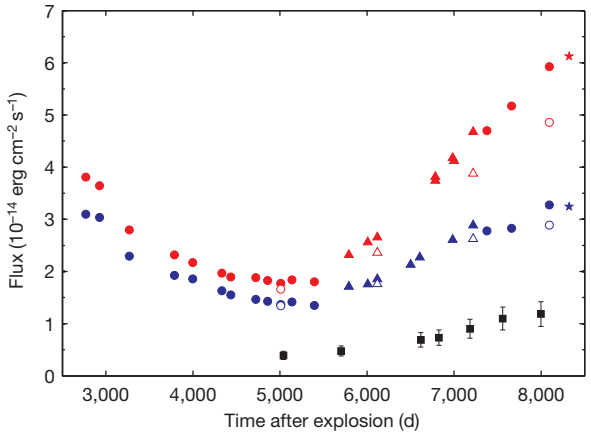


Figure 2 | Light curves of the ejecta in different wavebands. Data points from the HST Wide Field Planetary Camera 2 (WFPC2), Advanced Camera for Surveys (ACS) and Wide Field Camera 3 (WFC3) are shown as dots, triangles and stars, respectively. Red and blue symbols correspond to the R and B bands, respectively. A correction factor has been applied to the ACS and WFC3 fluxes to account for the differences between these instruments and WFPC2. To quantify the contamination from the brightening circumstellar ring, we created detailed models of the ring for a number of different epochs (Supplementary Information, section 1, and Supplementary Figs 3 and 4). The open symbols show the ejecta fluxes after the contribution from the ring has been removed. Although the contamination from the ring increases with time, it never exceeds ~20% of the flux from the ejecta. The statistical errors in the HST fluxes are smaller than the data points, and systematic errors are discussed in the Supplementary Information, section 1. The black squares show the flux of the [Ca II] doublet lines at rest wavelengths 7,291.5 Å and 7,323.9 Å, measured using the Ultraviolet and Visual Echelle Spectrograph of the Very Large Telescope at the European Southern Observatory (error bars, s.d.). These lines are free from contamination from the circumstellar ring and the reverse shock. We note the decreasing flux during the first ~5,000 d and the increase thereafter, indicating an extra source of energy. A further indication that the energy source has changed is that the colour, determined from the flux ratio between the B and R bands, changes from a level of ~0.8 up to day 5,000 to a value close to ~0.6 on day 8,000.

absorbed in the core region of the ejecta (corresponding to speeds less than 5,000 km s^{-1}), where most of the heavy elements reside. At an energy of ~0.35 keV, which corresponds to the temperature of the dominant component in the X-ray spectrum[18], the fraction of flux absorbed by the ejecta at t_{yr} years can be approximated by $1.6 \times 10^{-3} t_{yr}^{1.67}$ (the increase with time is mainly due to the increasing solid angle of the expanding ejecta, assumed to be spherical, as seen from the ring). This gives a present-day absorbed X-ray luminosity of $\sim 5.0 \times 10^{35}$ erg s^{-1}. In this calculation, we have neglected the weaker, highest-energy component that contributes to the X-ray spectrum[18]. We note that this does not significantly affect the estimate of the absorbed flux, although the hard X-rays may be important owing to their greater penetrating power.

To model the ejecta light curve produced by input from the X-rays, we scaled the observed X-ray flux[6] by the fraction absorbed at 0.35 keV, multiplied the resulting flux by a constant (corresponding to the conversion efficiency from X-rays to optical emission) and added this to the radioactive energy input. Figure 3 shows the scaled X-ray flux together with the observed light curves. This model follows the general trend of the observed fluxes in both bands, although we note that a more accurate model would need to take into account the detailed shape of the X-ray spectrum and the reprocessing of the X-rays into optical emission. The required conversion efficiency from X-rays to optical emission in our model is 5.0% in the R band and 3.1% in the B band.

The conversion of X-rays to optical/infrared emission is similar to that of the ^{44}Ti positrons. Both involve degradation of non-thermal electrons into heating, ionization and excitation. For a typical ionization fraction of 10^{-3}–10^{-2}, the expected efficiency of conversion from X-rays to Hα emission (the dominant line in the R band) is ~5% (Supplementary Information, section 3). This conversion factor is consistent with the scaling factor we used to model the light curve. Similar arguments apply to the B band. Furthermore, the density in the core is high enough for the timescale of recombination to be shorter than the expansion timescale, ensuring a balance between the energy input and output.

Other possible explanations for the increase in flux include input from a central pulsar[19], a transition from optically thick to optically thin dust and positron leakage from the iron-rich regions. We find that input from a pulsar is unlikely for several reasons. In particular, it would be a strange coincidence for the emission from the pulsar to mimic the increasing X-ray flux from the ring interaction. Also, we expect the energy input from a pulsar to be concentrated towards the low-velocity material at the centre of the ejecta, but observations of the Hα and [Ca II] lines show that the increase occurs for speeds up to ~5,000 km s^{-1}. We also note that constraints on a point source at the centre of SN 1987A have already been obtained using HST data

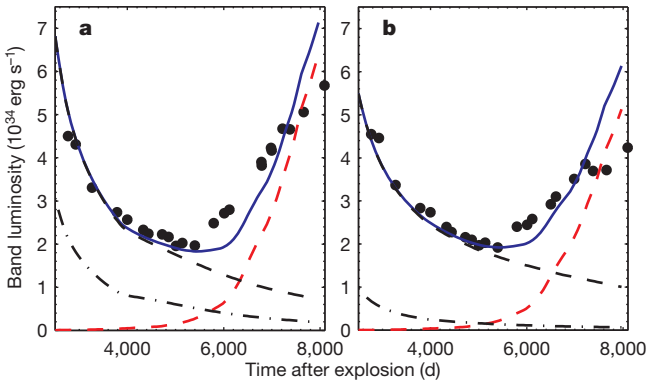


Figure 3 | Evolution of the luminosity from the ejecta in the R and B bands. **a**, R band; **b**, B band. The black dashed lines show a model with only radioactive input, mainly from ^{44}Ti. The ^{44}Ti mass used for the model is $1.4 \times 10^{-4} M_{\odot}$ (ref. 12), as determined from a detailed analysis that takes into account the effects of scattering and fluorescence of the ultraviolet flux, as well as a correction for internal dust[23,24] (here taken to have a 50% covering factor of the core inside the radius where the ejecta speed is 2,000 km s^{-1}). The dot–dash lines show the light curves with no ^{44}Ti, illustrating the necessity of including this isotope. The red dashed lines show a model based on a constant fraction of the observed X-ray flux[6], corrected for the fraction of flux absorbed by the ejecta (Supplementary Fig. 7). The blue solid line shows the sum of both models (black and red dashed). The R and B bands contain 5% and 3% of the total bolometric flux, respectively, and we expect these fractions to remain roughly constant in time. This is because the relative amount of energy resulting in heating, ionization and excitation will remain nearly constant as long as the ionization by the X-rays is $\lesssim 10^{-2}$. It is clear from this figure that there is a transition at ~5,000 d from a radioactivity-dominated phase to a phase dominated by the X-ray input from the collision with the ring.

taken near the minimum of the ejecta light curve[20]. A change in the properties of the dust or a transition in the positron deposition process is also unable to explain the observed increase in flux quantitatively (Supplementary Information, section 4).

We conclude that SN 1987A has made a transition from a radioactively dominated phase to a phase dominated by the X-ray input from the ring collision. This conclusion has interesting implications for the observed morphology. In particular, most of the X-rays are likely to be absorbed at the boundary of the ejecta core, where the density increases rapidly. This may lead to the light from the ejecta being emitted in a ring that is concentrated in the plane of the circumstellar ring. The 'hole' in the ejecta (Fig. 1), which has become more pronounced since about 2001, may in fact be a result of this rather than reflecting the true density distribution or dust obscuration. The asymmetric morphology seen at speeds of $\lesssim$3,000 km s^{-1} in the near-infrared [Si I] and [Fe II] lines[21] is, however, likely to be intrinsic to the metal core. By studying future changes in the morphology of the ejecta, we will be able to understand the origin of this asymmetry

In the future, the density of the ejecta will decrease and the fraction of X-rays absorbed will grow (Supplementary Fig. 7). As a result, the ionization will increase and a smaller fraction of the X-ray flux will produce line excitation. A larger fraction will go into heating, leading to an increase in the mid-infrared flux and a flattening of the optical light curves. In time, the X-rays will also penetrate deeper layers of the ejecta, thereby allowing us to probe the chemical structure of the innermost ejecta. This will be a novel form of X-ray tomography.

Received 31 January; accepted 31 March 2011.
Published online 8 June 2011.

1. Woosley, S. E., Heger, A. & Weaver, T. A. The evolution and explosion of massive stars. *Rev. Mod. Phys.* **74,** 1015–1071 (2002).
2. Fransson, C. & Kozma, C. Radioactivities and nucleosynthesis in SN 1987A. *N. Astron. Rev.* **46,** 487–492 (2002).
3. McKee, C. F. in *Young Supernova Remnants* (eds Holt, S. S. & Hwang, U.) 17–28 (Am. Inst. Phys. Conf. Proc. 565, Springer, 2001).
4. Morris, T. & Podsiadlowski, P. The triple-ring nebula around SN 1987A: fingerprint of a binary merger. *Science* **315,** 1103–1105 (2007).
5. Gröningsson, P. *et al.* Time evolution of the line emission from the inner circumstellar ring of SN 1987A and its hot spots. *Astron. Astrophys.* **492,** 481–491 (2008).
6. Racusin, J. L. *et al.* X-ray evolution of SNR 1987A: the radial expansion. *Astrophys. J.* **703,** 1752–1759 (2009).
7. Zanardo, G. *et al.* Multifrequency radio measurements of supernova 1987A over 22 years. *Astrophys. J.* **710,** 1515–1529 (2010).
8. Dwek, E. *et al.* Five years of mid-infrared evolution of the remnant of SN 1987A: the encounter between the blast wave and the dusty equatorial ring. *Astrophys. J.* **722,** 425–434 (2010).
9. Timmes, F. X., Woosley, S. E., Hartmann, D. H. & Hoffman, R. D. The production of ^{44}Ti and ^{60}Co in supernovae. *Astrophys. J.* **464,** 332–341 (1996).
10. Diehl, R. & Timmes, F. X. Gamma-ray line emission from radioactive isotopes in stars and galaxies. *Publ. Astron. Soc. Pacif.* **110,** 637–659 (1998).
11. Shigeyama, T. & Nomoto, K. Theoretical light curve of SN 1987A and mixing of hydrogen and nickel in the ejecta. *Astrophys. J.* **360,** 242–256 (1990).
12. Jerkstrand, A., Fransson, C. & Kozma, C. The ^{44}Ti-powered spectrum of SN 1987A. *Astron. Astrophys.* (in the press); preprint at ⟨http://arxiv.org/abs/1103.3653⟩ (2011).
13. Michael, E. *et al.* Hubble Space Telescope observations of high-velocity Lyα and Hα emission from supernova remnant 1987A: the structure and development of the reverse shock. *Astrophys. J.* **593,** 809–830 (2003).
14. Smith, N. *et al.* The reverse shock of SNR 1987A at 18 years after outburst. *Astrophys. J.* **635,** L41–L44 (2005).
15. Heng, K. *et al.* Evolution of the reverse shock emission from SNR 1987A. *Astrophys. J.* **644,** 959–970 (2006).
16. France, K. *et al.* Observing supernova 1987A with the refurbished Hubble Space Telescope. *Science* **329,** 1624–1627 (2010).
17. Blinnikov, S., Lundqvist, P., Bartunov, O., Nomoto, K. & Iwamoto, K. Radiation hydrodynamics of SN 1987A. I. Global analysis of the light curve for the first 4 months. *Astrophys. J.* **532,** 1132–1149 (2000).
18. Zhekov, S. A., Park, S., McCray, R., Racusin, J. L. & Burrows, D. N. Evolution of the Chandra CCD spectra of SNR 1987A: probing the reflected-shock picture. *Mon. Not. R. Astron. Soc.* **721,** 518–529 (2010).
19. Woosley, S. E., Hartmann, D. & Pinto, P. A. Hard emission at late times from SN 1987A. *Astrophys. J.* **346,** 395–404 (1989).
20. Graves, G. J. M. *et al.* Limits from the Hubble Space Telescope on a point source in SN 1987A. *Astrophys. J.* **629,** 944–959 (2005).
21. Kjær, K., Leibundgut, B., Fransson, C., Jerkstrand, A. & Spyromilio, J. The 3-D structure of SN 1987A's inner ejecta. *Astron. Astrophys.* **517,** A51–A60 (2010).
22. Chugai, N. N., Chevalier, R. A., Kirshner, R. P. & Challis, P. M. Hubble Space Telescope spectrum of SN 1987A at an age of 8 years: radioactive luminescence of cool gas. *Astrophys. J.* **483,** 925–940 (1997).
23. Lucy, L. B., Danziger, I. J., Gouiffes, C. & Bouchet, P. in *Supernovae* (ed. Woosley, S. E.) 82–94 (Springer, 1991).
24. Wooden, D. H. *et al.* Airborne spectrophotometry of SN 1987A from 1.7 to 12.6 microns: time history of the dust continuum and line emission. *Astrophys. J.* **88** (suppl.), 477–507 (1993).

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

Acknowledgements This work was supported by the Swedish Research Council and the Swedish National Space Board. Support for the HST observing programme was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc.

Author Contributions J.L. carried out the data reduction and analysis together with G.Ö., P.G., B.L., J.S. and P.C.; C.F. performed the theoretical modelling together with A.J. and C.K.; and J.L. and C.F. wrote the paper. R.P.K. is the principal investigator for the HST/SAINTS collaboration. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.L. (josefin.larsson@astro.su.se) or C.F. (claes@astro.su.se).

LETTER

doi:10.1038/nature10095

Hydrogen-poor superluminous stellar explosions

R. M. Quimby[1], S. R. Kulkarni[1], M. M. Kasliwal[1], A. Gal-Yam[2], I. Arcavi[2], M. Sullivan[3], P. Nugent[4], R. Thomas[4], D. A. Howell[5,6], E. Nakar[7], L. Bildsten[5,8], C. Theissen[9], N. M. Law[1,10], R. Dekany[11], G. Rahmer[11], D. Hale[11], R. Smith[11], E. O. Ofek[11], J. Zolkower[11], V. Velur[11], R. Walters[11], J. Henning[11], K. Bui[11], D. McKenna[11], D. Poznanski[4,12], S. B. Cenko[12] & D. Levitan[13]

Supernovae are stellar explosions driven by gravitational or thermonuclear energy that is observed as electromagnetic radiation emitted over weeks or more[1]. In all known supernovae, this radiation comes from internal energy deposited in the outflowing ejecta by one or more of the following processes: radioactive decay of freshly synthesized elements[2] (typically ^{56}Ni), the explosion shock in the envelope of a supergiant star[3], and interaction between the debris and slowly moving, hydrogen-rich circumstellar material[4]. Here we report observations of a class of luminous supernovae whose properties cannot be explained by any of these processes. The class includes four new supernovae that we have discovered and two previously unexplained events[5,6] (SN 2005ap and SCP 06F6) that we can now identify as members of the same class. These supernovae are all about ten times brighter than most type Ia supernova, do not show any trace of hydrogen, emit significant ultraviolet flux for extended periods of time and have late-time decay rates that are inconsistent with radioactivity. Our data require that the observed radiation be emitted by hydrogen-free material distributed over a large radius ($\sim 10^{15}$ centimetres) and expanding at high speeds ($>10^4$ kilometres per second). These long-lived, ultraviolet-luminous events can be observed out to redshifts $z > 4$.

The Palomar Transient Factory[7,8] (PTF) is a project dedicated to finding explosive events and has so far identified over one thousand supernovae. PTF09atu, PTF09cnd and PTF09cwl (also known as SN 2009jh[9]) were detected using the Palomar Observatory's 1.2-m Samuel Oschin Telescope during commissioning of the PTF system, in 2009, and PTF10cwr[10–12] (SN 2010gx[13]) was detected the following year (Fig. 1; see Supplementary Information, section 1). As with other supernova candidates, optical spectra for classification were obtained using the W. M. Keck Observatory's 10-m Keck I telescope, Palomar Observatory's 5.1-m Hale Telescope, and the 4.2-m William Herschel Telescope. The spectra (Fig. 2) show broad absorption dips at short wavelengths and mostly smooth continua at longer wavelengths. We further identify narrow absorption features in the PTF spectra from the Mg II doublet at rest wavelengths 2,796 Å and 2,803 Å, and measure redshifts of $z = 0.501$, 0.258, 0.349 and 0.230 for PTF09atu, PTF09cnd, PTF09cwl and

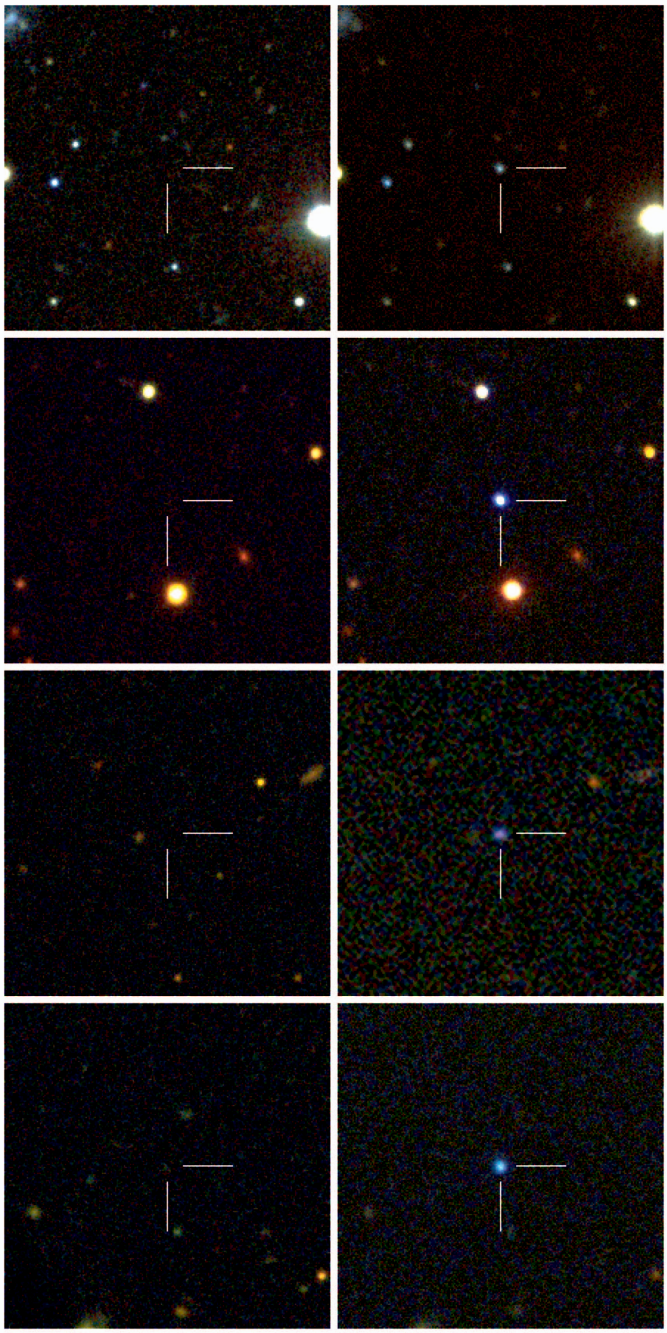


Figure 1 | Ultraviolet-luminous transients discovered by the PTF. Left: before explosion; right: after explosion; top to bottom: PTF09atu, PTF09cnd, PTF09cwl and PTF10cwr. Each tile shows a false-colour image constructed by assigning image data from three separate band passes to red, blue and green (g, r and i bands, respectively, for PTF09atu; u, g or V, and r bands for PTF09cnd, PTF09cwl and PTF10cwr). In each case, Sloan Digital Sky Survey reference data form the pre-explosion image. The post-explosion images are composed from observations made with the Palomar Observatory's 1.5-m telescope, the Wise Observatory's 1.0-m telescope and the Ultraviolet/Optical Telescope on board NASA's Swift satellite.

[1]Cahill Center for Astrophysics 249-17, California Institute of Technology, Pasadena, California 91125, USA. [2]Benoziyo Center for Astrophysics, Faculty of Physics, Weizmann Institute of Science, 76100 Rehovot, Israel. [3]Department of Physics (Astrophysics), University of Oxford, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH, UK. [4]Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA. [5]Las Cumbres Observatory Global Telescope Network, 6740 Cortona Drive, Suite 102, Goleta, California 93117, USA. [6]Department of Physics, University of California, Santa Barbara, Broida Hall, Santa Barbara, California 93106, USA. [7]Raymond and Beverly Sackler School of Physics & Astronomy, Tel Aviv University, Tel Aviv 69978, Israel. [8]Kavli Institute for Theoretical Physics, Kohn Hall, University of California, Santa Barbara, California 93106, USA. [9]University of California, San Diego, Department of Physics, 9500 Gilman Drive, La Jolla, California 92093, USA. [10]Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St George Street, Toronto, Ontario M5S 3H4, Canada. [11]Caltech Optical Observatories, California Institute of Technology, Pasadena, California 91125, USA. [12]Astronomy Department, University of California, Berkeley, 601 Campbell Hall, Berkeley, California 94720, USA. [13]Department of Physics, California Institute of Technology, Pasadena, California 91125, USA.

PTF10cwr, respectively. After combining the three available spectra of the SCP 06F6 transient, we find that the data correlate to the PTF sample and may also show narrow Mg II absorption with redshift $z = 1.189$ (Supplementary Information, section 4). Similarly to all PTF events, SN 2005ap ($z = 0.283$) shows a distinct W-shaped absorption feature near rest wavelength 4,300 Å. Although the broad spectral features of SN 2005ap are systematically shifted to higher speeds, the overall resemblance to the other PTF events is striking. The PTF discoveries bridge the redshift gap between SCP 06F6 and SN 2005ap and link these once disparate events, thus unifying them all into a single class.

With the redshifts above and a standard flat cosmology with Hubble parameter $H_0 = 71$ and matter energy density $\Omega_m = 0.27$, the peak absolute u-band AB magnitudes[14] for the PTF transients in the rest frame are near −22 mag and that for SCP 06F6 is near −22.3 mag (Fig. 3). The ~50-d rise of SCP 06F6 to maximum in the rest frame is compatible with the PTF sample, although there seems to be some diversity in the rise and decline timescales. To power these high peak magnitudes with radioactivity, several solar masses ($M_\odot$) of ^{56}Ni are needed ($>10M_\odot$, following ref. 15), and yet in the rest frame V band, the post-maximum decline rates of the PTF events are all >0.03 mag d^{-1}, which is a few times higher than the decay rate of ^{56}Co (the long-lived daughter nucleus of ^{56}Ni). These are therefore not radioactively powered events.

Next we check whether the observed photons could have been deposited by the explosion shock as it traversed the progenitor star. The photospheric radius, R_{ph}, that we infer for PTF09cnd at peak luminosity, on the basis of the observed temperature and assuming blackbody emission, is $R_{ph} \approx 5 \times 10^{15}$ cm (Supplementary Fig. 1). If the radiated photons were generated during the star explosion, then adiabatic losses would result in only a fraction R_*/R_{ph} of the energy remaining in the radiation at any given time, where R_* is the initial stellar radius. Given that the energy radiated around the peak is $\sim 10^{51}$ erg and that $R_*/R_{ph} < 10^{-3}$ for virtually any hydrogen-stripped progenitor, this model requires an unrealistic total explosion energy of $>10^{54}$ erg. In fact, the large radius and the duration of PTF09cnd leave almost no place for adiabatic losses (Supplementary Information, section 5), implying that the internal energy must have been deposited at a radius that is not much smaller than R_{ph}.

Integrating the rest frame g-band light curve and assuming no bolometric correction, we find that PTF09cnd radiated $\sim 1.2 \times 10^{51}$ erg. A similar analysis of the SCP 06F6 data gives a radiated energy of $\sim 1.7 \times 10^{51}$ erg. We also fit Planck functions to the ultraviolet and optical observations of PTF09cnd (Supplementary Fig. 1) and find an approximate bolometric output of $\sim 1.7 \times 10^{51}$ erg. The derived blackbody radii indicate a photospheric expansion speed of $v_{ph} \approx 14{,}000$ km s^{-1}. If the main source of luminance were the conversion of kinetic energy, then the bolometric energy would require $\sim 1M_\odot$ of material at this speed, assuming a conversion efficiency of 100%. A more realistic efficiency factor would make the minimum mass a few times larger. Because no traces of hydrogen are seen in any of the spectra (Supplementary Information, section 3), interaction with ordinary hydrogen-rich circumstellar material is ruled out. We thus conclude that these events cannot be powered by any of the commonly invoked processes driving known supernova classes.

The early spectra presented here are dominated by oxygen lines and do not show calcium lines, iron lines or other features commonly seen in ordinary core-collapse supernovae. The lack of metals is particularly noticeable in the ultraviolet flux, which is typically depleted by absorption. These events are hosted by low-luminosity galaxies that may provide a subsolar progenitor environment (Supplementary Information, section 6). The new class of events we have identified is thus observationally characterized by extreme peak luminosities, short decay times inconsistent with radioactivity, and very hot early spectra with significant ultraviolet flux and lacking absorption lines from heavy elements such calcium and iron, which are commonly seen in all other types of supernova.

These observations require a late deposition of a large amount of energy ($>10^{51}$ erg) into hydrogen-poor, rapidly expanding material (slow-moving material would produce narrow spectroscopic features, which are not observed). We point out two possible physical processes that can perhaps power these superluminous sources. One is a strong interaction with a massive, rapidly expanding, hydrogen-free shell.

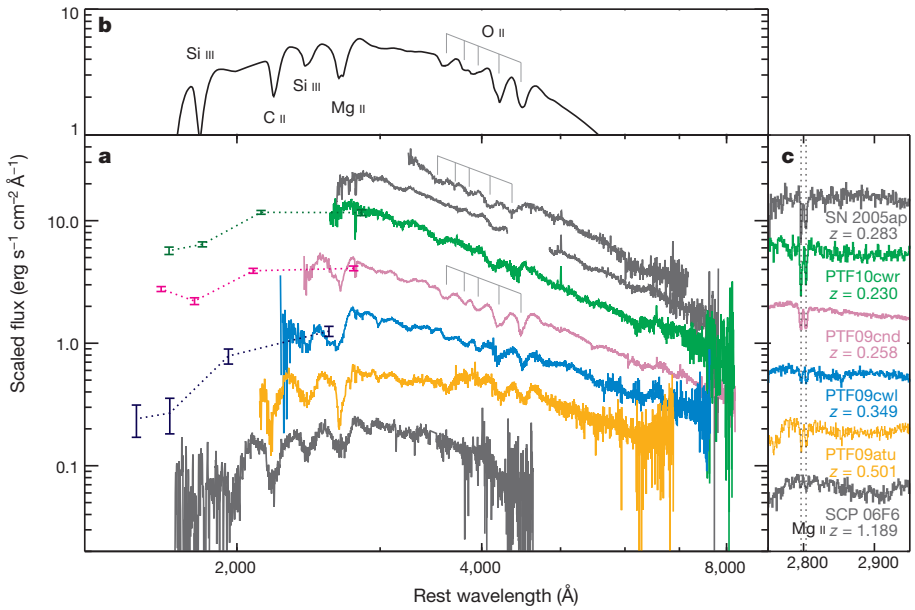


Figure 2 | Spectral energy distributions of the SN 2005ap-like sample. **a**, Top to bottom: SN 2005ap on 2005 March 7 and again on March 16 (grey), PTF10cwr on 2010 March 18 (green), PTF09cnd on 2009 August 25 (purple), PTF09cwl on 2009 August 25 (blue), PTF09atu on 2009 July 20 (orange) and an average of all three SCP 06F6 spectra presented in ref. 6 (grey). The spectra have been adjusted to remove the effect of redshift, binned and scaled arbitrarily for display purposes. Broadband flux densities from the Swift observations (scaled to join the spectra in the u band) are plotted for PTF09cnd, PTF09cwl and PTF10cwr with 1σ error bars. Five absorption bands are marked by the combs above SN 2005ap and PTF09cnd, the former being ~7,000 km s^{-1} faster. **b**, These features can be well fitted by O II using the highly parametric spectral synthesis code SYNOW[21] (Supplementary Information, section 2). SYNOW fits also suggest that C II and Mg II can account for the respective features at 2,200 Å and 2,700 Å. The fit to the line at 2,500 Å is improved with the addition of Si III. The model shown has a photospheric speed of 15,000 km s^{-1}. **c**, Close-up views of the narrow Mg II doublet, from which we derive the redshifts.

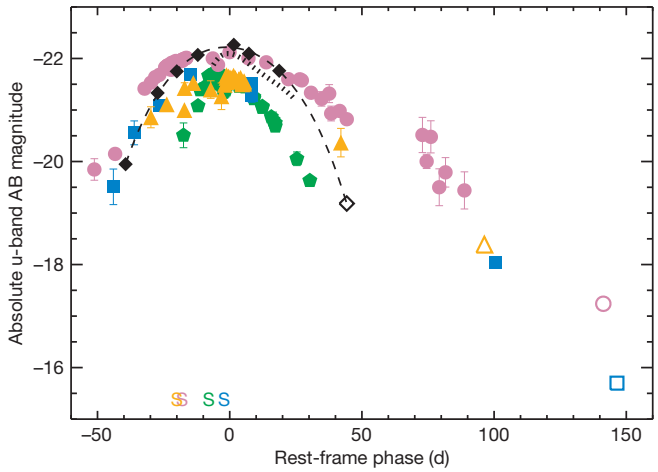


Figure 3 | Luminosity evolution of the SN 2005ap-like sample. Shown are the SCP 06F6 transient (diamonds), SN 2005ap (hatching), PTF09atu (orange triangles), PTF09cnd (purple dots), PTF09cwl (blue squares) and PTF10cwr (green pentagons). In each case, we transform the observed photometry to absolute u-band magnitudes by correcting both for distance and differences in the effective rest-frame band pass introduced by the redshifts. For SCP 06F6, the observed i band is similar to the rest frame u band, so the correction factor is nearly independent of the spectral properties[22]. For the PTF sample, however, the correction factor varies over time as the supernovae cool. We interpolated the observed spectra of PTF09cnd to phases appropriate for the B-, g-, V- and r-band observations of the PTF sample to calculate the correction factors. Errors bars representing 1σ (excluding the colour correction) are shown when larger than the plotting symbols. The colour corrections for PTF09atu and PTF09cwl near day 100 are very uncertain (~0.5 mag). We have not removed possible host light contaminating the late-time observations of PTF09atu, PTF09cnd and PTF09cwl (open symbols). Thus, these measurements represent upper limits on the supernova light. Host galaxy light may in fact dominate the final PTF09cwl observation. The final observation of SCP 06F6 (open diamond) is a 2.5σ detection made from the ground. Along the abscissa, with appropriately colour-coded 'S's we note the phases of spectra shown in Fig. 2.

Such a situation is naturally produced by extremely massive stars with initial masses in the range $90\ M_{\odot} \lesssim M_{\rm i} \lesssim 130\ M_{\odot}$, which are expected[16,17] to undergo violent pulsations—perhaps driven by the pair instability—that strip their outer layers and expel massive, hydrogen-poor shells. The star eventually dies by becoming a stripped-envelope, core-collapse supernova, which may interact with previously ejected carbon- or oxygen-rich shells to drive the observed luminosity[18]. Alternatively, the power source can be a prolonged energy injection by a central engine. For example, a spinning-down nascent magnetar[19,20] can account for the peak luminosities ($>10^{44}\,{\rm erg\,s^{-1}}$) and time to peak light (30–50 d) observed for these events, assuming a magnetic field $B \approx (1\text{–}3) \times 10^{14}\,{\rm G}$ and a natal spin period of 1–3 ms.

The high luminosities, exceptionally blue spectral energy distributions and volumetric event rates (Supplementary Information, section 7) of this new class of supernovae make them prime targets for high-redshift studies (Supplementary Fig. 3). Because they remain at around maximum luminosity for months to years in the observer frame, these events provide a steady light source to illuminate their environs and any intervening clouds of gas and dust. This creates new opportunities for high-resolution spectroscopy to probe distant star-forming regions in primitive galaxies without the need for rapid scheduling and with the benefit that the luminous supernova beacon eventually fades, allowing the study of the galaxy itself. Indeed, these events promise to be a rich source of results for future 30-m-class-telescope science.

Received 17 February; accepted 1 April 2011.
Published online 8 June 2011.

1. Woosley, S. E. & Weaver, T. A. The physics of supernova explosions. *Annu. Rev. Astron. Astrophys.* **24,** 205–253 (1986).
2. Arnett, W. D. Type I supernovae. I - Analytic solutions for the early part of the light curve. *Astrophys. J.* **253,** 785–797 (1982).
3. Grassberg, E. K., Imshennik, V. S. & Nadyozhin, D. K. On the theory of the light curves of supernovae. *Astrophys. Space Sci.* **10,** 28–51 (1971).
4. Chevalier, R. A. Self-similar solutions for the interaction of stellar ejecta with an external medium. *Astrophys. J.* **258,** 790–797 (1982).
5. Quimby, R. M. *et al.* SN 2005ap: a most brilliant explosion. *Astrophys. J.* **668,** L99–L102 (2007).
6. Barbary, K. *et al.* Discovery of an unusual optical transient with the Hubble Space Telescope. *Astrophys. J.* **690,** 1358–1362 (2009).
7. Law, N. M. *et al.* The Palomar Transient Factory: system overview, performance, and first results. *Publ. Astron. Soc. Pacif.* **121,** 1395–1408 (2009).
8. Rau, A. *et al.* Exploring the optical transient sky with the Palomar Transient Factory. *Publ. Astron. Soc. Pacif.* **121,** 1334–1351 (2009).
9. Drake, A. J. *et al.* Supernova 2009jh. *Centr. Bur. Electron. Telegr.* 1958 (2009).
10. Mahabal, A. A. *et al.* Supernova candidates and classifications from CRTS. *Astronomer's Telegr.* 2490 (2010).
11. Quimby, R. M. *et al.* Discovery of a luminous supernova, PTF10cwr. *Astronomer's Telegr.* 2492 (2010).
12. Pastorello, A. *et al.* Detection of PTF10cwr/CSS100313 on PS1 sky survey images and host galaxy identification. *Astronomer's Telegr.* **2504** (2010).
13. Pastorello, A. *et al.* Ultrabright optical transients are linked with type Ic supernovae. *Astrophys. J.* **724,** L16–L21 (2010).
14. Oke, J. B. & Gunn, J. E. Secondary standard stars for absolute spectrophotometry. *Astrophys. J.* **266,** 713–717 (1983).
15. Gal-Yam, A. *et al.* Supernova 2007bi as a pair-instability explosion. *Nature* **462,** 624–627 (2009).
16. Woosley, S. E., Blinnikov, S. & Heger, A. Pulsational pair instability as an explanation for the most luminous supernovae. *Nature* **450,** 390–392 (2007).
17. 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).
18. Chevalier, R. A. & Irwin, C. M. Shock breakout in dense mass loss: luminous supernovae. *Astrophys. J.* **729,** L6–L9 (2011).
19. Woosley, S. E. Bright supernovae from magnetar birth. *Astrophys. J.* **719,** L204–L207 (2010).
20. Kasen, D. & Bildsten, L. Supernova light curves powered by young magnetars. *Astrophys. J.* **717,** 245–249 (2010).
21. Jeffery, D. J. & Branch, D. in *Supernovae, Jerusalem Winter School for Theoretical Physics* (eds Wheeler, J. C., Piran, T. & Weinberg, S.) 149–247 (World Scientific, 1990).
22. Hogg, D. W., Baldry, I. K., Blanton, M. R. & Eisenstein, D. J. The K correction. Preprint at ⟨http://arxiv.org/abs/astro-ph/0210394⟩ (2002).

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

Acknowledgements Observations were obtained with the Samuel Oschin Telescope and the 60-inch telescope at Palomar Observatory as part of the PTF project, a scientific collaboration between the California Institute of Technology, Columbia University, Las Cumbres Observatory, the Lawrence Berkeley National Laboratory, the National Energy Research Scientific Computing Center, the University of Oxford and the Weizmann Institute of Science. Some of the data presented here were obtained at the W. M. Keck Observatory and the William Herschel Telescope. The National Energy Research Scientific Computing Center, which is supported by the Office of Science of the US Department of Energy, provided staff, computational resources and data storage for this project. Observations by the Weizmann PTF partnership and from the Wise Observatory are supported by grants from the Israel Science Foundation and the US–Israel Binational Science Foundation. We acknowledge support from the US Department of Energy Scientific Discovery through Advanced Computing programme, the Hale Fellowship from the Gordon and Betty Moore foundation, the Bengier Foundation, the Richard and Rhoda Goldman Fund, and the Royal Society.

Author Contributions R.M.Q. initiated, coordinated and managed the project, carried out photometric and spectroscopic observations and analysis, and wrote the manuscript. S.R.K. is the PTF principal investigator and contributed to manuscript preparation. M.M.K. obtained spectroscopy from the Keck I telescope and helped with the P60 observations. A.G.-Y. oversaw the Wise observations and contributed to analysis and manuscript writing. I.A. extracted the Wise photometry and helped obtain Keck I spectra. M.S. carried out and analysed spectroscopic observations from the William Herschel Telescope. P.N. designed and implemented the image-subtraction pipeline that detected the PTF events. R.T. analysed the combined spectra using his automated SYNOW code. D.A.H. helped to identify the PTF spectra as being like that of SN 2005ap. E.N. contributed to the physical interpretation and manuscript writing. L.B. advised during the preparation of the manuscript. C.T. helped vet potential candidates and first identified PTF09atu and PTF09cwl. N.M.L. is the PTF project scientist and oversaw the PTF system. R.D., G.R., D.H., R.S., E.O.O., J.Z., V.V., R.W., J.H., K.B. and D.M. helped to build and commission the PTF system. D.P., S.B.C. and D.L. helped to vet PTF candidates and obtain spectroscopic observations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.M.Q. (quimby@astro.caltech.edu).

LETTER

doi:10.1038/nature10119

Experimental non-classicality of an indivisible quantum system

Radek Lapkiewicz[1,2], Peizhe Li[1], Christoph Schaeff[1,2], Nathan K. Langford[1,2]†, Sven Ramelow[1,2], Marcin Wieśniak[1]† & Anton Zeilinger[1,2]

In contrast to classical physics, quantum theory demands that not all properties can be simultaneously well defined; the Heisenberg uncertainty principle is a manifestation of this fact[1]. Alternatives have been explored—notably theories relying on joint probability distributions or non-contextual hidden-variable models, in which the properties of a system are defined independently of their own measurement and any other measurements that are made. Various deep theoretical results[2–5] imply that such theories are in conflict with quantum mechanics. Simpler cases demonstrating this conflict have been found[6–10] and tested experimentally[11,12] with pairs of quantum bits (qubits). Recently, an inequality satisfied by non-contextual hidden-variable models and violated by quantum mechanics for all states of two qubits was introduced[13] and tested experimentally[14–16]. A single three-state system (a qutrit) is the simplest system in which such a contradiction is possible; moreover, the contradiction cannot result from entanglement between subsystems, because such a three-state system is indivisible. Here we report an experiment with single photonic qutrits[17,18] which provides evidence that no joint probability distribution describing the outcomes of all possible measurements—and, therefore, no non-contextual theory—can exist. Specifically, we observe a violation of the Bell-type inequality found by Klyachko, Can, Binicioğlu and Shumovsky[19]. Our results illustrate a deep incompatibility between quantum mechanics and classical physics that cannot in any way result from entanglement.

The Heisenberg uncertainty principle is perhaps one of the most curious and surprising features of quantum physics: it prohibits certain properties of physical systems (for example the position and momentum of a single particle) from being simultaneously well defined[1]. Such incompatibility of properties, however, contrasts strongly with what we experience in our everyday lives. If we look at a globe of the Earth, we can see only one hemisphere at any given time, but we suppose that the shapes of the continents on the far side remain the same irrespective of the observer's vantage point. Thus, by spinning the globe around to view different continents, we are able to construct a meaningful picture of the whole. It is reasonable to assume that observation reveals features of the continents that are present independent of which other continent we might be looking at. In an analogous way, classical physics allows us to assign properties to a system without actually measuring it. All these properties can be assumed to exist in a consistent way, whether or not they are measured.

The world view in which system properties are defined independently of both their own measurement and what other measurements are made is called non-contextual realism. From this viewpoint, mathematically speaking there must be a joint probability distribution for these properties, defining the outcome probabilities for an experiment in which they are observed (if a joint probability distribution exists, rolling appropriately weighted dice would reproduce all behaviour of such an experiment). The reverse is not necessarily true. Nature could in principle be such that although joint probability distributions exist they do not relate to properties of a system.

To derive the Bell-like inequality in ref. 19, consider five numbers, a_1, a_2, a_3, a_4 and a_5, each equal to $+1$ or -1. For any choice of them, the following algebraic inequality is true:

$$a_1a_2 + a_2a_3 + a_3a_4 + a_4a_5 + a_5a_1 \geq -3 \qquad (1)$$

Let these numbers now be the results of five corresponding two-outcome measurements, A_1, A_2, A_3, A_4 and A_5. Then, assuming that there exists a joint probability distribution for the 2^5 possible measurement outcome combinations, taking the average of inequality (1) gives (see Supplementary Information, section 1)

$$\langle A_1A_2\rangle + \langle A_2A_3\rangle + \langle A_3A_4\rangle + \langle A_4A_5\rangle + \langle A_5A_1\rangle \geq -3 \qquad (2)$$

Here, angle brackets denote averages of measurement outcomes and not quantum mechanical expectation values.

We would like to emphasize that, given only the result of inequality (2), an experimental violation of this inequality precludes the description of the measurement results using a joint probability distribution. The same form of argument can be used to show that such a violation also precludes the existence of any non-contextual realistic model for the results.

The above inequality can be tested experimentally if we assume that a long series of individual experimental runs would result in a fair sampling of a joint probability distribution (if one were to exist). Our experimental implementation consists of five main stages, depicted in Fig. 1b–f. We prepare single photons, each distributed among three modes (Fig. 1a) monitored by detectors. At each stage, the two outcomes of a given measurement are defined by determining whether the corresponding detector clicked. For example, at stage one (Fig. 1b), the outcomes of measurement A_1 are given by the response of the upper detector, and we assign numbers to the outcomes as above: that is, $A_1 = -1$ if the detector clicked and $A_1 = +1$ if it did not. Similarly, the outcomes of measurement A_2 are given by the response of the lower detector. By measuring A_1 and A_2 together for a number of photons, we obtain the average value $\langle A_1A_2\rangle$, the first term of inequality (2).

To move to the second stage, we perform a transformation, T_1, on the two upper modes (Fig. 1c). Because the mode monitored by the lower detector is not affected by T_1, this detector still measures the outcome of A_2. The upper detector, however, defines a different measurement—we call it A_3. Most significantly, for any specific run of the experiment it seems reasonable to assume that whether or not detector A_2 clicks must be independent of whether or not we apply the transformation to the other two modes. In the remaining three stages, we apply three more transformations, each time changing one measurement and leaving the other unaffected.

The last transformation, T_4, is chosen such that the new measurement should be equivalent to the original measurement of A_1. Unlike

[1]Vienna Center for Quantum Science and Technology, Faculty of Physics, University of Vienna, Boltzmanngasse 5, Vienna A-1090, Austria. [2]Institute for Quantum Optics and Quantum Information, Austrian Academy of Sciences, Boltzmanngasse 3, Vienna A-1090, Austria. †Present addresses: Clarendon Laboratory, Department of Physics, University of Oxford, Parks Road, Oxford OX1 3PU, UK (N.K.L.); Institute of Theoretical Physics and Astrophysics, University of Gdansk, PL-80-952 Gdansk, Poland (M.W.).

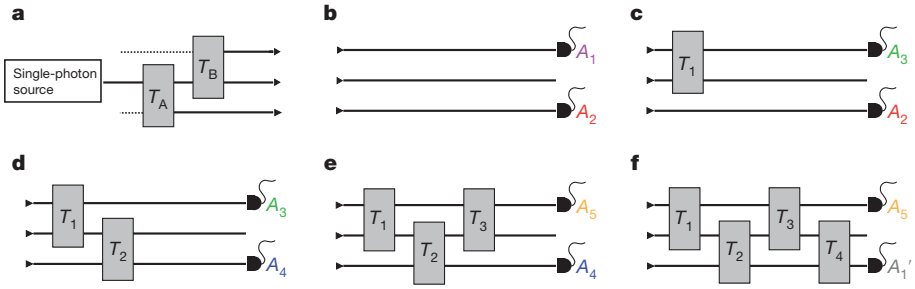


Figure 1 | Experimental preparation and five successive measurement stages. Straight, black lines represent the optical modes (beams) and grey boxes represent transformations on the optical modes. **a**, Single photons are distributed among three modes by transformations T_A and T_B. This preparation stage is followed by one of the five measurement stages. **b**–**f**, At each stage, the response of two detectors monitoring the optical modes defines one of the pairs of measurements in inequality (2). Outcomes of the measurements are defined by determining whether the corresponding detector clicks. If a detector clicks then a value of -1 is assigned to the corresponding measurement; otherwise, a value of $+1$ is assigned. A key aspect of our experimental implementation is that each transformation acts only on two modes, leaving the other mode completely unaffected. Thus, the part of the physical set-up corresponding to measurement A_2 is exactly the same in **b** and **c** (likewise, A_3 is the same in **c** and **d**, and so on). We note that this set-up can also be arranged such that the choice between A_1 and A_3 is made long after A_2 is measured. Thus, it seems reasonable to assume that measurement A_2 is independent of whether it is measured together with A_1 or A_3. The same reasoning can be applied to measurements A_3, A_4 and A_5.

for the other measurements; however, this new measurement apparatus (Fig. 1f) is not physically the same as the one measuring A_1 in Fig. 1b. We therefore call the sixth measurement A_1' and derive the following new inequality to replace inequality (2) (Supplementary Information, section 2):

$$\langle A_1A_2\rangle+\langle A_2A_3\rangle+\langle A_3A_4\rangle+\langle A_4A_5\rangle+\langle A_5A_1'\rangle\geq-3-\varepsilon \quad (3)$$

Here $\varepsilon=1-\langle A_1'A_1\rangle$ and we note that for the ideal case of $A_1'=A_1$, this extended inequality reduces to inequality (2).

Measurements A_1 and A_1' occur at different stages of the experiment, so the expectation value $\langle A_1'A_1\rangle$ cannot be calculated in the same way as the other terms. Instead, we note that in the ideal case, whenever detector A_1 fires, A_1' must also fire, and vice versa. Therefore, if the upper beam is blocked where A_1 would be measured, then A_1' should never click. Likewise, blocking the other two modes should not change the count rate at detector A_1'. Therefore, we can rewrite $\langle A_1'A_1\rangle$ as

$$\begin{aligned}\langle A_1'A_1\rangle=1-2(&\mathrm{P}(A_1'=-1\ \ |A_1=1)\mathrm{P}(A_1=1)\\&+\mathrm{P}(A_1'=1\ \ |A_1=-1)\mathrm{P}(A_1=-1))\end{aligned} \quad (4)$$

with the conditional probabilities experimentally accessible by blocking and monitoring the appropriate modes (Supplementary Fig. 1). The extra term, ε, in inequality (3) therefore completely accounts for any differences between A_1 and A_1'. It is an open question whether any experimental apparatus can be designed where A_1 and A_1' are physically the same.

In our measurements, we find that $\varepsilon=0.081(2)$, thus bounding the left-hand side of inequality (3) by $-3.081(2)$, and that the terms on the left-hand side are each less than -0.7 adding up to give $-3.893(6)$. This represents a violation of inequality (3) by more than 120 standard deviations, demonstrating that no joint probability distribution is capable of describing our results. The violation also excludes any non-contextual hidden-variable model. The result does, however, agree well with quantum mechanical predictions, as we will show now.

A single photon distributed among three modes can be described by the mathematical formalism used for spin-one particles. Using this formalism, the measurements performed in our experiment can be expressed by spin operators as $A_i=2\hat{S}_i^2-1$, where $\hat{S}_i$ is a spin projection onto the direction $\mathbf{l}_i$ in real three-dimensional space. Two measurements $\hat{S}_i^2$ and $\hat{S}_j^2$ are compatible if and only if the directions $\mathbf{l}_i$ and $\mathbf{l}_j$ are orthogonal. Thus, the five measurement directions have to be pairwise orthogonal to make the measurements themselves pairwise compatible (Fig. 2). In our experiment, we have three modes which by design represent orthogonal states. These can be seen as orthogonal directions in the spin case. An essential feature of a spin-one system is that out of three projections squared onto three orthogonal directions, exactly two are equal to 1 and the remaining one must be equal to 0. Correspondingly, a single photon distributed among three different modes will cause a click in exactly one detector only and none in the other two.

For the maximal violation of inequalities (2) and (3), the five directions form a regular pentagram and the input state has zero spin along the symmetry axis of the pentagram[19]. In the ideal case, that is, when $A_1=A_1'$ ($\varepsilon=0$), if the optimal state is taken quantum mechanics predicts the value of the left-hand side of inequality (3) to be $5-4\sqrt{5}\approx-3.944$. Smaller violations are also predicted for a range of non-ideal pentagrams and other input states. In our case, the departure from the maximum achievable violation can be attributed to residual errors in the settings of experimental parameters (Supplementary Information, section 5).

The five pairs of compatible measurements correspond to the five measurement devices described in Fig. 1b–f. Spin rotations, necessary to switch between various measurement bases, can be realized by combining the optical modes, for example on a tunable beam splitter.

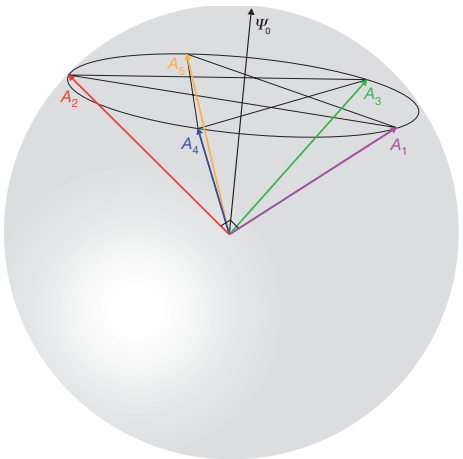


Figure 2 | Representation of the measurements and a state providing maximal violation of the inequality (2) by directions in three-dimensional space[19]. The measurement directions are labelled by the measurements A_i ($i=1, 2, \ldots, 5$). They are given as $2\hat{S}_i^2-1$, where $\hat{S}_i$ is a spin projection onto the direction $\mathbf{l}_i$. The five measurement directions are pairwise orthogonal, making the measurements A_i pairwise compatible. These five pairs correspond to the five measurement devices from Fig. 1b–f. For a maximal violation, the directions form a regular pentagram and the input state Ψ_0 has zero spin along the symmetry axis of the pentagram.

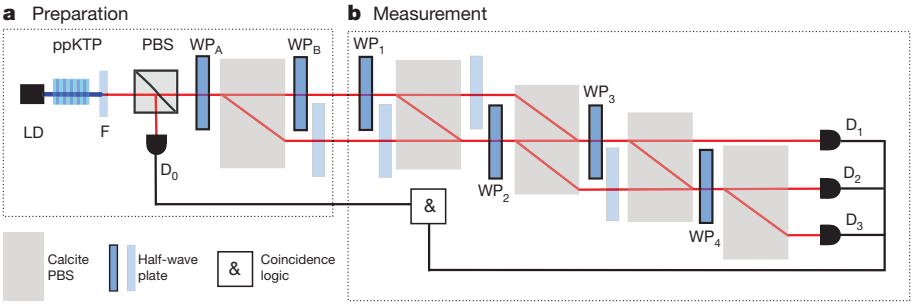


Figure 3 | Experimental set-up. a, Preparation of the required single-photon state. About 3 mW of power from a grating-stabilized laser diode (LD) at a wavelength of 405 nm is used to pump the nonlinear crystal (20 mm long, periodically poled potassium titanyl phosphate (ppKTP)), producing pairs of orthogonally polarized photons by means of spontaneous parametric down-conversion. The pump is filtered out with the help of a combination of dichroic mirrors and interference filters (labelled jointly as F). The photon pairs are split up at a polarizing beam splitter (PBS). Detection of the reflected photon heralds the transmitted one. Half-wave plates WP_A and WP_B transform the transmitted photon into the desired three-mode state. Calcite polarizing beam splitters separate and combine orthogonally polarized modes. **b**, In the measurement apparatus, half-wave plates WP_1 to WP_4 realize the transformations T_1 to T_4 on pairs of modes (wave-plate orientations are listed in Supplementary Table 1). Each transformation can be 'turned off' by setting the optical axis of the corresponding wave plate vertically (at 0°). The unlabelled (light blue) wave plates serve to balance the path lengths and to switch between horizontal and vertical polarization (the second unlabelled wave plate is set to 0°; the rest are set to 45°). Detecting heralded single photons in practice means registering coincidences between single-photon detectors: D_0 and each of D_1, D_2 and D_3. Registrations in two of the detectors D_1, D_2 and D_3 give the values A_i necessary to evaluate the inequality (Table 1). The third detector is used to identify the trials when the photon is lost. We note that the assignment of measurements to detectors in the experimental set-up differs in some cases from that described in the simplified conceptual scheme (Fig. 1). We use home-built avalanche photodiode single-photon detectors and coincidence logic. The effective coincidence window (including the jitter of the detector) is about 2.3 ns.

Each rotation of a measurement basis leaves one of the axes unaffected because only two modes are combined and one is untouched. Operationally, the measurements at each stage are compatible (co-measurable) because they are defined by independent detectors.

We realize the above scheme using heralded, single, 810-nm photons. In the type-II collinear spontaneous parametric down-conversion in a nonlinear crystal, pairs of orthogonally polarized photons are produced in a polarization product state. We separate them at a polarizing beam splitter, with detection of a vertically polarized photon heralding the horizontally polarized photon in the measurement set-up. In practice, therefore, we record pairwise coincidences between the heralding detector and any one of the three detectors in the measurement apparatus (heralded single clicks).

To realize a three-state system, our photons propagate in three modes. Two of the three modes are realized as two polarizations of a single spatial mode. Conveniently, two-mode transformations can then be implemented using half-wave plates acting on the two polarization modes propagating in the same spatial mode. Different spatial modes are combined using calcite crystals (acting as polarizing beam splitters). Thus, we are able to apply transformations to any pair of modes (Fig. 3).

The experiment consists in total of seven stages. The first five stages (corresponding to the five left-hand terms of inequality (3)) are the measurement configurations illustrated in Fig. 1b–f, whereas the final two, depicted in Supplementary Fig. 1, give us the value of ε. All of these measurements are realized with a single experimental apparatus tuned to one of seven configurations. Configurations one to five differ in the number of transformations that are 'active'. We activate and deactivate the transformations by changing the orientation of wave plates (see Supplementary Table 1 for the specific settings). For the two measurements in the final stage, where we measure conditional probabilities by blocking the appropriate modes (Supplementary Fig. 1), we insert a polarizer in two orthogonal orientations between wave plates WP_B and WP_1 (Fig. 3).

For each measurement, we record clicks for 1 s, registering about 3,500 heralded single photons. We repeat each stage 20 times, average the results and calculate the standard deviation of the mean to estimate the standard uncertainties which we then propagate to the final results (Table 1). Owing to photon loss, sometimes no photon is detected in the measurement, despite the observation of a heralding event. We therefore discard all events for which only the trigger detector (D_0) and none of the measurement detectors (D_1, D_2 and D_3) fire. We assume

Table 1 | Collected experimental results

(a)

D_1		D_2		D_3		Calculated contribution	
Condition	Value	Condition	Value	Condition	Value	Term	Value
$P(A_1 = 1, A_2 = -1)$	0.471(3)	$P(A_1 = -1, A_2 = 1)$	0.432(3)	$P(A_1 = 1, A_2 = 1)$	0.097(1)	$\langle A_1A_2\rangle$	−0.805(2)
$P(A_2 = -1, A_3 = 1)$	0.473(4)	$P(A_2 = 1, A_3 = 1)$	0.098(2)	$P(A_2 = 1, A_3 = -1)$	0.429(4)	$\langle A_2A_3\rangle$	−0.804(3)
$P(A_3 = 1, A_4 = 1)$	0.146(2)	$P(A_3 = 1, A_4 = -1)$	0.429(2)	$P(A_3 = -1, A_4 = 1)$	0.426(2)	$\langle A_3A_4\rangle$	−0.709(3)
$P(A_4 = 1, A_5 = -1)$	0.466(2)	$P(A_4 = -1, A_5 = 1)$	0.439(2)	$P(A_4 = 1, A_5 = 1)$	0.095(1)	$\langle A_4A_5\rangle$	−0.810(2)
$P(A_5 = -1, A_1' = 1)$	0.469(2)	$P(A_5 = 1, A_1' = -1)$	0.414(2)	$P(A_5 = 1, A_1' = 1)$	0.117(2)	$\langle A_5A_1'\rangle$	−0.766(3)
						Sum	−3.893(6)

(b)

D_1	D_3	$D_1 + D_3$		D_2		Calculated contribution	
Value	Value	Condition	Value	Condition	Value	Term	Value
0.788(2)	0.196(2)	$P(A_1' = 1 \mid A_1 = 1)$	0.983(1)	$P(A_1' = -1 \mid A_1 = 1)$	0.017(1)	$-3 - \varepsilon$	−3.081(2)
0.010(1)	0.062(2)	$P(A_1' = 1 \mid A_1 = -1)$	0.072(2)	$P(A_1' = -1 \mid A_1 = -1)$	0.928(2)		

Value indicates the measured probability corrected for relative efficiencies (Supplementary Information, section 3). Estimates of standard uncertainties (standard deviations of the means) are given in the brackets. Condition indicates assigned measurement values corresponding to heralded single-click events. Because of low detection efficiency, we need to use a third detector. It enables us to identify and discard trials in which a photon was lost (heralded no-click events). The rates of heralded double clicks (simultaneous responses of the heralding detector and two other detectors) are negligible—typically two orders of magnitude smaller than the standard deviation in the rate of heralded single clicks. **a,** Results for the Bell-like inequality in ref. 19. Rows 1–5 correspond to terms 1–5 of inequality (3), and are measured with the corresponding devices illustrated in Fig. 1b–f. The last column is given by $\langle A_iA_j\rangle = P(A_i = A_j = 1) - P(A_i = -1, A_j = 1) - P(A_i = 1, A_j = -1)$. **b**, Extended bound. Rows 1 and 2 (corresponding to Supplementary Fig. 1) display the contributions to the conditional probabilities that are necessary to evaluate the additional terms in extended inequality (3) using equation (4).

that the photons we do detect are a representative sample of all created photons (fair-sampling assumption).

A key aspect of Kochen–Specker experiments is that the co-measured observables must commute; if they do not, the 'compatibility loophole'[20] is potentially opened. The construction of the measurements in our experiment enforces their compatibility and thus makes the experiment immune to the compatibility loophole. Detector efficiencies and losses in the set-up prevent us from closing the detection loophole. Instead, we assume that the statistics of unregistered events would have been the same as the statistics of observed ones.

Our experimental results are in conflict with any description of nature that relies on a joint probability distribution of outcomes of a simple set of measurements. This also precludes any description in terms of non-contextual hidden-variable models. To our knowledge, this is the first observation of such a conflict for a single three-state system, which, apart from being the most basic one where such a contradiction is possible, cannot even in principle contain entanglement. For such a system, inequality (2) involves the smallest number of measurements possible.

Our result sheds new light on the conflict between quantum and classical physics. To finish, we want to point out that any model based on a joint probability distribution can in principle be non-deterministic. The experimental preclusion of such models highlights the fact that even for a single, indivisible quantum system, allowing randomness is not sufficient to allow its description with a conceptually classical model.

Received 9 March; accepted 15 April 2011.

1. Heisenberg, W. *The Physical Principles of Quantum Theory* 13–52 (Univ. Chicago Press, 1930).
2. Specker, E. Die Logik nicht gleichzeitig entscheidbarer Aussagen. *Dialectica* **14,** 239–246 (1960).
3. Bell, J. S. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.* **38,** 447–452 (1966).
4. Kochen, S. & Specker, E. P. The problem of hidden variables in quantum mechanics. *J. Math. Mech.* **17,** 59–87 (1967).
5. Gleason, A. M. Measures on the closed subspaces of a Hilbert space. *J. Math. Mech.* **6,** 885–893 (1957).
6. Mermin, N. D. Simple unified form for the major no-hidden-variables theorems. *Phys. Rev. Lett.* **65,** 3373–3376 (1990).
7. Peres, A. Two simple proofs of the Kochen-Specker theorem. *J. Phys. A* **24,** L175–L178 (1991).
8. Peres, A. *Quantum Theory: Concepts and Methods* Ch. 7 (Kluwer, 1993).
9. Clifton, R. Getting contextual and nonlocal elements-of-reality the easy way. *Am. J. Phys.* **61,** 443–447 (1993).
10. Cabello, A., Estebaranz, J. M. & García-Alcaine, G. Bell-Kochen-Specker theorem: a proof with 18 vectors. *Phys. Lett. A* **212,** 183–187 (1996).
11. Huang, Y.-F., Li, C.-F., Zhang, Y.-S., Pan, J.-W. & Guo, G.-C. Experimental test of the Kochen-Specker theorem with single photons. *Phys. Rev. Lett.* **90,** 250401 (2003).
12. Bartosik, H. *et al.* Experimental test of quantum contextuality in neutron interferometry. *Phys. Rev. Lett.* **103,** 040403 (2009).
13. Cabello, A. Experimentally testable state-independent quantum contextuality. *Phys. Rev. Lett.* **101,** 210401 (2008).
14. Kirchmair, G. *et al.* State-independent experimental test of quantum contextuality. *Nature* **460,** 494–497 (2009).
15. Amselem, E., Rådmark, M., Bourennane, M. & Cabello, A. State-independent quantum contextuality with single photons. *Phys. Rev. Lett.* **103,** 160405 (2009).
16. Moussa, O., Ryan, C. A., Cory, D. G. & Laflamme, R. Testing contextuality on quantum ensembles with one clean qubit. *Phys. Rev. Lett.* **104,** 160501 (2010).
17. Lapkiewicz, R, *et al. Most Basic Experimental Falsification of Non-Contextuality* (Poster, 13th Workshop on Quantum Information Processing, 2010).
18. Lapkiewicz, R, *et al. Experimental Non-Classicality of an Indivisible System*. Abstr. Q29.00008 (41st Annu. Meeting Div. Atom. Mol. Opt. Phys., American Physical Society, 2010).
19. Klyachko, A. A., Can, M. A., Binicioğlu, S. & Shumovsky, A. S. Simple test for hidden variables in spin-1 systems. *Phys. Rev. Lett.* **101,** 20403 (2008).
20. Gühne, O. *et al.* Compatibility and noncontextuality for sequential measurements. *Phys. Rev. A* **81,** 22121 (2010).

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

Acknowledgements This work was supported by the ERC (Advanced Grant QIT4QAD), the Austrian Science Fund (Grant F4007), the EC (Marie Curie Research Training Network EMALI), the Vienna Doctoral Program on Complex Quantum Systems and the John Templeton Foundation. We acknowledge A. A. Klyachko for discussion of the proposal made in ref. 19; M. Hentschel, M. Kacprowicz and G. J. Pryde for discussions of technical issues; A. Cabello, S. Osnaghi, H. M. Wiseman and M. Żukowski, with whom we discussed the conceptual issues; and M. Nespoli for help during the early stages of the experiment.

Author Contributions All authors contributed to the design of the experiment. R.L., P.L. and C.S. performed the experiment and all authors wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.Z. (anton.zeilinger@univie.ac.at).

LETTER

doi:10.1038/nature10164

Flying in a flock comes at a cost in pigeons

James R. Usherwood[1], Marinos Stavrou[1], John C. Lowe[1], Kyle Roskilly[1] & Alan M. Wilson[1]

Flying birds often form flocks, with social[1], navigational[2] and anti-predator[3] implications. Further, flying in a flock can result in aerodynamic benefits, thus reducing power requirements[4], as demonstrated by a reduction in heart rate and wingbeat frequency in pelicans flying in a V-formation[5]. But how general is an aerodynamic power reduction due to group-flight? V-formation flocks are limited to moderately steady flight in relatively large birds, and may represent a special case. What are the aerodynamic consequences of flying in the more usual 'cluster'[6,7] flock? Here we use data from innovative back-mounted Global Positioning System (GPS) and 6-degrees-of-freedom inertial sensors to show that pigeons (1) maintain powered, banked turns like aircraft, imposing dorsal accelerations of up to 2*g*, effectively doubling body weight and quadrupling induced power requirements; (2) increase flap frequency with increases in all conventional aerodynamic power requirements; and (3) increase flap frequency when flying near, particularly behind, other birds. Therefore, unlike V-formation pelicans, pigeons do not gain an aerodynamic advantage from flying in a flock. Indeed, the increased flap frequency, whether due to direct aerodynamic interactions or requirements for increased stability or control, suggests a considerable energetic cost to flight in a tight cluster flock.

We recorded 18 pigeons during seven bouts of voluntary straight and circling flight around their home loft over a period of more than 9 pigeon-hours of flight, 400 pigeon-km, and over 243,000 flaps. Speed and position relative to the ground were determined using back-mounted 10 Hz raw Doppler and pseudorange GPS data. GPS data were post-processed relative to local base station data using Waypoint GrafNav 8.10 (Novatel). Relative airspeeds were calculated by taking account of air velocity measured using a three-dimensional ultrasonic anemometer sited on an adjacent rooftop. Within-wingbeat motions were recorded with 300 Hz inertial measurement units consisting of three axes of accelerometer and three axes of gyroscope. Flap frequencies and body amplitudes were determined from the dorsal acceleration signal. The quantity and quality of these measurements allow, for the first time, the relationship of flap frequency and body motions with airspeed, induced, climbing and accelerating power, and proximity to other pigeons, to be separated and quantified using multiple polynomial regression.

The pigeons flew in a flock of a range of densities and positions, at a range of speeds, ascended and descended while circling at a range of radii (Supplementary Information shows a movie of one of seven sequences). By treating GPS velocity measurements and accelerometer and gyroscope recordings independently, we determined that pigeons bank optimally according to aerodynamic theory. Turns require centripetal acceleration $a_{centrip}$ (Fig. 1a). If the pigeons bank appropriately, this combines with the acceleration due to gravity g to require an increase in dorsal acceleration a_{dorsal}; this is indeed what is observed. Further, to maintain the level of banking for a given turn rate (derived only from GPS measurements) requires a net yaw and pitch every flap (Supplementary Information), and these are close to those calculated by integrating gyroscope-derived pitch and yaw rates (Fig. 1b, c). Therefore, during fast, sweeping turns (radius approximately 15–20 m), pigeons bank appropriately to load both wings evenly, and increase the average loading on their wings by up to double (>1,500 flaps), presenting a relatively unconfounded self-manipulation of induced power via changes in apparent weight. Such flapping turns do not seem to be near a power limit: ascending flight is achieved at over 4 m s^{-1} (>8,000 flaps). If this climbing power was available to pay the additional induced power costs of turning during constant-speed, level flight, a turn radius of 6.6 m would be achievable at 15 m s^{-1}, concomitantly imposing a 3.6-fold increase in effective body weight. It is likely that powered circling is, at least in part, force-limited, similar to bend-running humans[8,9] because, unlike wheeled vehicles, propeller-driven aircraft and greyhounds[10], the main motors powering bird flight cannot be dissociated from weight-support.

This self-imposed variability in effective weight support, along with the range of airspeeds (flight speed with respect to the air) V_{air}, ascent speeds V_z and rostral accelerations $a_{rostral}$, imposes largely independent demands of four aerodynamic power components: (1) parasite/profile power (due to the drag on the body and wings, dependent on airspeed); (2) induced power (due to supporting weight with finite wings); (3) climbing power (due to changes in gravitational potential energy); and (4) accelerating power (due to changes in kinetic energy). Two kinematic parameters are related to variation in speed and power requirements. Flap frequency is expected to have a strong bearing on power[11,12], as it relates to both the aerodynamic forces on the wings—and hence the work per cycle—and the cycle rate. Amplitude of the dorsal body displacement, derived from double integration of the dorsal accelerometer measurements over each flap period, gives an indirect measure of the amplitude of the wing motions. However, the exact relationship between body motions and wing motions is unclear[13,14], depending on both inertial and aerodynamic reactions, and so will be additionally related to speed.

To test whether increases in either kinematic parameter consistently indicate an increase in power and provide insight into the mechanisms of power modulation, while making few a priori assumptions about how they might vary with speed and power requirement, we fit third-order polynomials to each factor, and find partial coefficients that give the best fit for the whole model. In addition, we include two 'flock factors' as a metric for the proximity of other birds. This is the proportion of forward or rearward hemisphere view that would be covered by all other pigeons at the instant of each flap (assuming each other pigeon to occlude a circle with diameter of 0.5 m) (Fig. 3c).

Each component of the regression model has a strong relationship with flap frequency and dorsal amplitude (Figs 2 and 3), with increases in each associated with an increase in flap frequency. The influence of airspeed, if the effects of induced power (proportional to the inverse of airspeed) are treated independently, reflects the demands of parasite and profile power. Combining the airspeed polynomial results with those of induced power allows the frequency–speed relationship to be calculated for steady, level flight at a large distance from other pigeons. This (blue line, Fig. 2a) shows the familiar U-shaped curve of frequency with airspeed, with higher frequencies at lower speeds, demanded by high induced power requirements, and higher frequencies at higher speeds, related to higher parasite and profile drag requirements. It must be emphasized, however, that although wingbeat frequency is probably closely correlated with power at around the

[1]Structure and Motion Laboratory, The Royal Veterinary College, University of London, North Mymms, Hatfield AL9 7TA, UK.

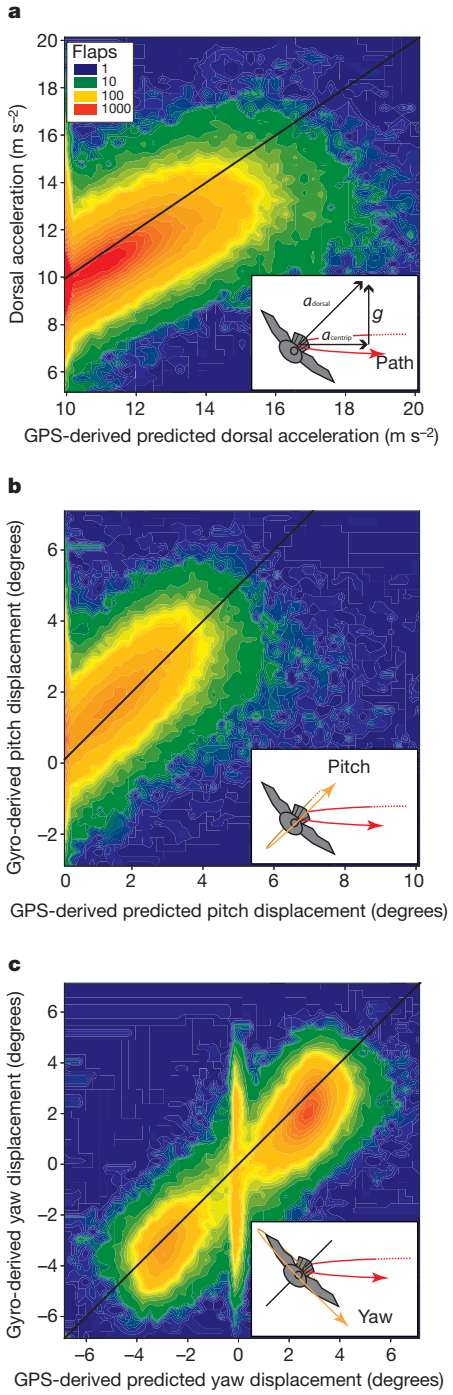


Figure 1 | Flap-number histogram contour plots. **a**, Flap-averaged dorsal acceleration; **b**, net pitch and **c**, yaw angular displacements (averaged over five flaps). Turning with optimal banking requires an increase in dorsal acceleration, and pitch and yaw displacement for every flap (insets). Black lines show the predicted (from GPS alone) against observed (inertial measurement units alone) relationships for birds, assuming they bank optimally during turns.

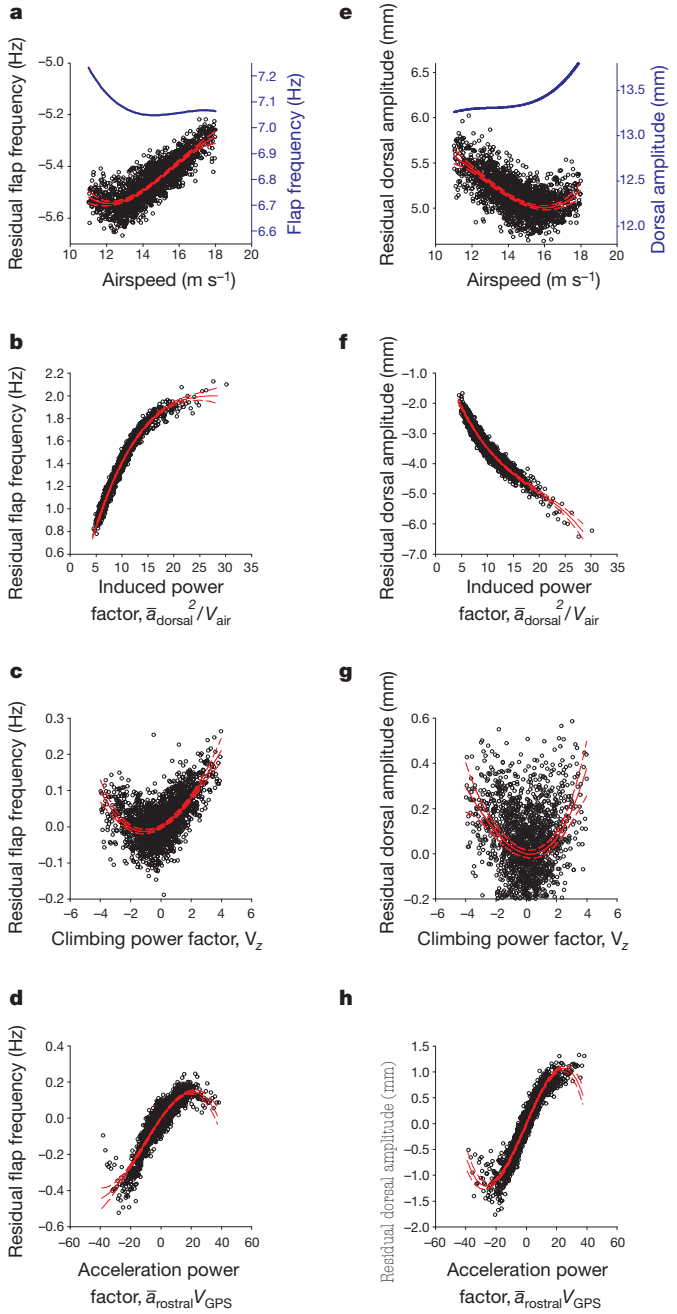


Figure 2 | Influence of speed, and power and flock factors on flap frequency and dorsal amplitude. **a**–**h**, The influence attributable to airspeed (**a**, **e**), induced power (**b**, **f**), climbing power (**c**, **g**) and accelerating power (**d**, **h**) on flap frequency (**a**–**d**) and dorsal amplitude (**e**–**h**) over each flap (18 pigeons, 171,209 flaps). Red curves show third-order polynomial fits; points show the values once the influence of all other factors have been removed. Each point represents the average of 100 flaps, binned along the *x*-axis. Dashed red lines show ±99.99% confidence intervals. Blue curves (**a**, **e**) show the relationship predicted if the effect of both airspeed and induced power (which includes airspeed as a term) are combined; in effect, the relationship that would be observed for steady, straight, level flight.

minimum power speed[12], it cannot be taken as a direct proxy for aerodynamic power over a range of speeds; the best measurements to date show that minimum power speeds do not occur at the flight speeds of minimum flap frequency[15,16]. But it is reasonable to conclude from our measurements that, once the effects of speed are removed—that is, for a given speed—an increased power requirement is strongly associated with increases in flap frequency.

The relationship of speed and power with body motion is less straightforward: whereas increases in mechanical power requirements due to ascending or accelerating flight are, perhaps unsurprisingly, related to increases in dorsal amplitude, the opposite is true for speed and induced power (Fig. 2e–h). Pigeons take higher frequency but shallower wingbeats as they fly around tighter circles. In physiological terms, this is counterintuitive: one would expect extra power to be produced with, in addition to increases in frequency and muscle stress (directly related to the increased wing forces due to increased effective weight), an increase in muscle strain from an increased flap amplitude (as is observed for climbing and accelerating powers). So is there an

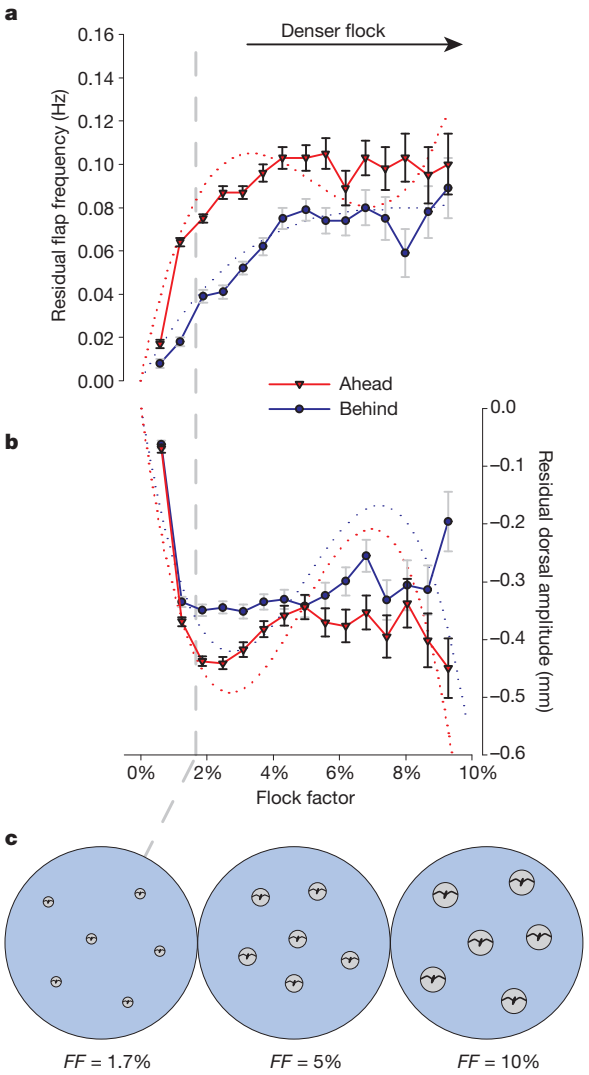


Figure 3 | Relationship between flock factor and flap frequency or dorsal displacement. **a**–**c**, The relationship between flock factor—the proportion of hemisphere-view covered by other pigeons for every flap, illustrated graphically with six neighbouring birds in **c**—and flap frequency (**a**) or dorsal displacement amplitude (**b**). The vertical grey dashed line indicates the mean flock factor ($FF = 1.7\%$), the underlying dotted lines show the third-order polynomial fits used in the statistical separation of factors, and the error bars show s.e.m. Flight in a cluster flock, particularly when flying behind other birds, is associated with an increase in flap frequency and decrease in dorsal amplitude.

aerodynamic account for this method of modulation in response to induced power demands? Indeed: such flight would ameliorate the increasingly significant induced power costs by reducing the spatial and temporal fluctuations in the wake, pushing the correction factor traditionally applied to induced power calculations[11,17–19] towards unity. In effect, the momentum jet in the wake would be directed ventrally more evenly, and the mass flux increased. If this is the case, some counter-cost must be invoked to account for why this flight style is not always adopted. Given all other aerodynamic powers are either equivalent or also improved with lower amplitude flapping, the inertial power cost of flapping non-massless wings must be considered. Inertial power is proportional to the square of flap amplitude, but the cube of frequency; while a high wingbeat frequency would result in a high aerodynamic efficiency, particularly important under high loads, it would also require a higher inertial power. This observation therefore provides some support for the notion[20,21] that flapping flight involves a trade-off between aerodynamic efficiency (requiring low wingbeat amplitudes at high frequencies) and inertial power costs (requiring low wingbeat frequencies), even during relatively fast flight.

Flight within a flock, if the effects of airspeed and conventional aerodynamic power relationships are accounted for, results in an increase in flap frequency and decrease in flap amplitude (Fig. 3). Given that (1) V-formation-flying birds benefiting from aerodynamic interaction fly with reduced wingbeat frequency, and (2) an increase in each form of aerodynamic power is associated with an increase in frequency in pigeons, the increase in flap frequency with proximity to other birds indicates an increased power requirement due to flying within a cluster flock. The ultimate cause for this relationship is currently unclear. Whereas aerodynamic interactions are certainly feasible—modal nearest neighbour distance was 1.49 m (close to the mean of 1.54 m reported from photogrammetric studies of a turning pigeon flock[22]), well under a flapping wavelength—the magnitude of the change is not well predicted by simple downwash calculations. The downward velocity of the wake close behind another pigeon would, from momentum-flux considerations, be approximately $0.4\ m\ s^{-1}$; if this was evenly distributed, even following tightly in a flock with a flock factor of 10%, this would require a relative climbing speed of only $0.04\ m\ s^{-1}$. This ascent velocity would result in an increase in flap frequency of only 0.001 Hz instead of the observed 0.1 Hz. So either the assumption of uniform downwash is incorrect, the pigeons are disturbed by some other aerodynamic interaction, or they increase their flap frequency for reasons that are not directly aerodynamic. This last option seems most parsimonious. The benefits of high frequency, low amplitude flight may well be increased opportunity for control and collision avoidance: higher flap frequencies can allow both a greater manoeuvrability and passive stability[23]; lower amplitudes (Fig. 3c) would allow a higher flock density without risk of collision or requirement of accurate wingpath coherence between birds (of which we find no evidence). But whatever the ultimate cause driving the higher flap frequency in denser pigeon flocks, this observation indicates that, unlike the V-formation, flight in a cluster flock comes at an energetic cost.

METHODS SUMMARY

Measurements. We recorded a flock of 18 trained racing pigeons (*Columba livia*, mean mass 0.461 kg, s.d. 0.045 kg; mean span 0.691 m, s.d. 0.039 m) during voluntary flights with synchronized back-mounted 10 Hz raw Doppler and pseudo-range GPS and 300-Hz 12-bit precision tri-axial accelerometers ($\pm 12g$) and tri-axial gyroscopes ($\pm 2{,}000$ degrees per second). Sensor, logger and mounting mass (35 g) constituted less than 8% body mass. Although this is larger than usually accepted on wild-bird studies, and would also have imposed a slight increase in body drag, the fact that extended, voluntary flights were performed, including circling at $2g$, demonstrates that this load was not prohibitive. Experiment protocols were approved by the RVC local Ethics and Welfare committee. A GPS base station (Novatel FlexPak G2L and GPS-702-GG antenna) and ultrasonic anemometer (1 Hz, WindMaster; Gill Instruments) on a nearby rooftop provided synchronised measurements. For further details on post-processing and turning calculations, see Supplementary Information.

Power models. We found best fit (least squares) partial regression coefficients K for third order polynomials of airspeed, induced power factor, climbing power factor, acceleration power factor (Fig. 2) and flock factors FF_{ahead} and FF_{behind} using the following expressions:

	Frequency (Hz) or dorsal amplitude (m) =
Airspeed component	$K_1 V_{air} + K_2 {V_{air}}^2 + K_3 {V_{air}}^3$
Induced power component	$+ K_4 \frac{\overline{a_{dorsal}}^2}{V_{air}} + K_5 \left(\frac{\overline{a_{dorsal}}^2}{V_{air}}\right)^2 + K_6 \left(\frac{\overline{a_{dorsal}}^2}{V_{air}}\right)^3$
Climbing power component	$+ K_7 V_z + K_8 {V_z}^2 + K_9 {V_z}^3$
Acceleration power component	$+ K_{10} \overline{a_{rostral}} V_{GPS} + K_{11} (\overline{a_{rostral}} V_{GPS})^2 + K_{12} (\overline{a_{rostral}} V_{GPS})^3$
Forward flock factor component	$+ K_{13} FF_{ahead} + K_{14} {FF_{ahead}}^2 + K_{15} {FF_{ahead}}^3$
Rear flock factor component	$+ K_{16} FF_{behind} + K_{17} {FF_{behind}}^2 + K_{18} {FF_{behind}}^3$
Constant	+ Constant

Each power component takes account of the kinematic contribution towards power; for instance, induced power is proportional to the square of body weight (hence dorsal acceleration to take account of turns) and inversely proportional to airspeed[11]. Each individual pigeon was included as a categorical variable, thus taking some account of the morphological variability between pigeons. The data set was first bounded to remove extreme values (frequencies 5–10 Hz; air speed 11–18 m s^{-1}; dorsal acceleration 8.81–20.81 m s^{-2}; $V_z = -4$ to 4 m s^{-1}; flock factors < 0.1; acceleration power within ± 39.2 W kg^{-1}, equivalent to the V_z bounding), leaving 171,209 flaps. The overall model fitted the data with $R^2 = 0.48$ (frequency) or $R^2 = 0.37$ (dorsal amplitude) (statistical tables in Supplementary Information).

Received 23 March; accepted 28 April 2011.

1. Nagy, M., Ákos, Z., Biro, D. & Vicsek, T. Hierarchical group dynamics in pigeon flocks. *Nature* **464,** 890–893 (2010).
2. Dell'Ariccia, G., Dell'Omo, G., Wolfer, D. P. & Lipp, H.-P. Flock flying improves pigeons' homing: GPS track analysis of individual flyers versus small groups. *Anim. Behav.* **76,** 1165–1172 (2008).
3. Tinbergen, N. *The Study of Instinct* (Clarendon Press, 1951).
4. Lissaman, P. B. S. & Schollenberger, C. A. Formation flight of birds. *Science* **168,** 1003–1005 (1970).
5. Weimerskirch, H., Martin, J., Clerquin, Y., Alexandre, P. & Jiraskova, S. Energy saving in flight formation. *Nature* **413,** 697–698 (2001).
6. Higdon, J. J. L. & Corrsin, S. Induced drag of a bird flock. *Am. Nat.* **112,** 727–744 (1978).
7. Heppner, F. H. Avian flight formations. *Bird-Banding* **45,** 160–169 (1974).
8. Greene, P. R. & McMahon, T. A. Running in circles. *Physiologist* **22,** S35–S36 (1979).
9. Usherwood, J. R. & Wilson, A. M. Accounting for elite indoor 200 m sprint results. *Biol. Lett.* **2,** 47–50 (2006).
10. Usherwood, J. R. & Wilson, A. M. No force limit on greyhound sprint speed. *Nature* **438,** 753–754 (2005).
11. Pennycuick, C. J. *Bird Flight Performance: A Practical Calculation Manual* (Oxford Univ. Press, 1989).
12. Pennycuick, C. J., Klaasen, M., Kvist, A. & Lindström, Å. Wingbeat frequency and the body drag anomaly: wind-tunnel observations on a thrush nightingale (*Luscinia luscinia*) and a teal (*Anas crecca*). *J. Exp. Biol.* **199,** 2757–2765 (1996).
13. Bilo, D., Lauck, A. & Nachtigall, W. in *Biona-Report 3* (ed. Nachtigall, W.) 87–108 (Gustav Fischer, 1984).
14. Hedrick, T. L., Usherwood, J. R. & Biewener, A. A. Wing inertia and whole-body acceleration: an analysis of instantaneous aerodynamic force production in cockatiels (*Nymphicus hollandicus*) flying across a range of speeds. *J. Exp. Biol.* **207,** 1689–1702 (2004).
15. Tobalske, B. W., Hedrick, T. L., Dial, K. P. & Biewener, A. A. Comparative power curves in bird flight. *Nature* **421,** 363–366 (2003).
16. Schmidt-Wellenburg, C. A., Biebach, H., Daan, S. & Visser, G. H. Energy expenditure and wing beat frequency in relationship to body mass in free flying barn swallows (*Hirundo rustica*). *J. Comp. Physiol. B* **177,** 327–337 (2007).
17. Rayner, J. M. V. A vortex theory of animal flight. Part 1. The vortex wake of a hovering animal. *J. Fluid Mech.* **91,** 697–730 (1979).
18. Ellington, C. P. The aerodynamics of hovering insect flight. V. A vortex theory. *Phil. Trans. R. Soc. Lond. B* **305,** 115–144 (1984).
19. Spedding, G. R. & McArthur, J. Span efficiencies of wings at low Reynolds numbers. *J. Aircr.* **47,** 120–128 (2010).
20. Lilienthal, O. *Birdflight as the Basis of Aviation* (Markowski International, 2001) [transl.].
21. Usherwood, J. R. Inertia may limit efficiency of slow flapping flight, but mayflies show a strategy for reducing the power requirements of loiter. *Bioinspir. Biomim.* **4,** 015003 (2009).
22. Pomeroy, H. & Heppner, F. Structure of turning in airborne rock dove (*Columba livia*) flocks. *Auk* **109,** 256–267 (1992).
23. Hedrick, T. L., Cheng, B. & Deng, X. Wingbeat time and the scaling of passive rotational damping in flapping flight. *Science* **324,** 252–255 (2009).

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

Acknowledgements We would like to thank T. Hubel, H. Chapman, V. Unt and T. Demes for practical assistance, and The Wellcome Trust (J.R.U.), The Royal Society (A.M.W.), BBSRC (K.R.) and EPSRC for funding.

Author Contributions J.R.U. and A.M.W. conceived and designed the project. J.R.U. analysed the data, and wrote the paper with input from all other authors. M.S. trained the pigeons and helped perform the experiments. J.C.L., K.R. and A.M.W. developed and built the equipment.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.R.U. (jusherwood@rvc.ac.uk).

LETTER

doi:10.1038/nature10190

City living and urban upbringing affect neural social stress processing in humans

Florian Lederbogen[1]*, Peter Kirsch[1]*, Leila Haddad[1]*, Fabian Streit[1], Heike Tost[1], Philipp Schuch[1], Stefan Wüst[1], Jens C. Pruessner[2], Marcella Rietschel[1], Michael Deuschle[1] & Andreas Meyer-Lindenberg[1]

More than half of the world's population now lives in cities, making the creation of a healthy urban environment a major policy priority[1]. Cities have both health risks and benefits[1], but mental health is negatively affected: mood and anxiety disorders are more prevalent in city dwellers[2] and the incidence of schizophrenia is strongly increased in people born and raised in cities[3–6]. Although these findings have been widely attributed to the urban social environment[2,3,7,8], the neural processes that could mediate such associations are unknown. Here we show, using functional magnetic resonance imaging in three independent experiments, that urban upbringing and city living have dissociable impacts on social evaluative stress processing in humans. Current city living was associated with increased amygdala activity, whereas urban upbringing affected the perigenual anterior cingulate cortex, a key region for regulation of amygdala activity, negative affect[9] and stress[10]. These findings were regionally and behaviourally specific, as no other brain structures were affected and no urbanicity effect was seen during control experiments invoking cognitive processing without stress. Our results identify distinct neural mechanisms for an established environmental risk factor, link the urban environment for the first time to social stress processing, suggest that brain regions differ in vulnerability to this risk factor across the lifespan, and indicate that experimental interrogation of epidemiological associations is a promising strategy in social neuroscience.

Urbanization, a process that started in North America and Western Europe but is now mainly occurring in developing nations, is a major socio-ecological change confronting mankind. By 2050, 69% of humans will live in urban areas[1]. Although city dwellers, on average, are wealthier and receive improved sanitation, nutrition, contraception and health care[1], urban living is also associated with increased risk for chronic disorders, a more demanding and stressful social environment and greater social disparities. The biological components of this complex landscape of risk and protective factors remain largely uncharacterized.

Some of the best-established effects of urbanization concern mental health. Meta-analyses show that current city dwellers have a substantially increased risk for anxiety disorders (by 21%) and mood disorders (by 39%)[2]. For the major brain disorder, schizophrenia, incidence is about doubled in subjects born and brought up in cities[3], with evidence of a dose–response relationship[5] that probably reflects causation[3]. Genetically vulnerable individuals are more at risk[6], in agreement with the assumption that schizophrenia represents a neurodevelopmental disorder[11]. Importantly, urbanicity effects on schizophrenia later in life are minor[2,5], providing an epidemiological dissociation between current and early life urbanicity effects, which are associated with mood and anxiety disorders and schizophrenia, respectively.

Because longitudinal studies indicate that urbanicity effects on mental illness are causal and not mediated by other epidemiological variables[3,7], attempts to explain these associations must consider the specifics of the urban situation affecting the brain[12]. Increased social evaluative threat[13], including social defeat and chronic social stress, might constitute such a factor[8]. Consequently, many authors have proposed that social stress processing in the urban environment underlies the greater risk for mental illness[2,3,7,8], and contributes to the manifestation of these disorders in adults. To test experimentally the hypothesis that urban living and upbringing modulate neural processing of acute social evaluative stress, we studied the neural responses of healthy German volunteers undergoing such stress during functional magnetic resonance imaging (fMRI). We confirmed our findings in a second study using a different social stress paradigm and then tested for cognitive specificity by ascertaining the effect of urbanicity on brain activation during cognitive processing without stress. Importantly, our subjects did not have a mental disorder nor were they at high risk for one; the link to these illnesses from the environmental risk factor that we studied is established by the epidemiological evidence discussed earlier.

In our first (discovery) study, we used the Montreal Imaging Stress Task (MIST)[14], a social stress paradigm where participants solve arithmetic tasks under time pressure. Difficulty was varied adaptively to keep success rates—visually presented on a 'performance scale'—at between 25–40%. Study investigators provided further negative feedback after each test segment through headphones. Subjective stress levels were measured before and after the session using a visual analogue scale, and effects of the MIST on salivary cortisol, heart rate and blood pressure were recorded repeatedly. Urbanicity was quantified as follows[4]: city with more than 100,000 inhabitants (3); town with more than 10,000 inhabitants (2); and rural area (1). For urban upbringing, these numbers were multiplied by the number of years living in the area up to age fifteen and added. Thirty-two participants with rural as well as urban upbringing and habitation entered the final analysis (Supplementary Table 1a). City dwellers did not differ in subjective health, depressed mood, social support, or personality dimensions. Baseline circadian cortisol measures were normal[15]. The MIST increased cardiovascular and hormonal measures (Supplementary Fig. 1a and Supplementary Table 2), indicating that stress was successfully induced. Stress-related brain activations (compared to a control condition without social evaluative threat) were most prominent in the right temporoparietal junction ($t = 9.53$, $P = 0.001$, all significance values are family-wise error (FWE) corrected for multiple comparisons), anterior cingulate cortex (ACC) and posterior cingulate cortex (anterior: $t = 7.91$, $P < 0.001$; posterior: $t = 8.09$, $P < 0.001$), insular cortex (right: $t = 8.18$, $P < 0.001$; left: $t = 6.69$, $P = 0.003$) and hypothalamus (right: $t = 7.12$, $P < 0.001$; left: $t = 6.86$, $P = 0.002$, see Supplementary Fig. 1b and Supplementary Table 3 for complete list). Differential brain activation correlated significantly with the test-induced rise in cortisol for hippocampus (right: $r = -0.59$, $P = 0.041$; left: $r = -0.59$, $P = 0.040$) and amygdala (right: $r = -0.61$, $P = 0.016$; left: $r = -0.55$, $P = 0.048$), confirming previous reports[16].

Although current and early life (birth to age 15) urbanicity shared moderate variance ($r = 0.37$, $P < 0.05$), their neural effects were fully

[1]Central Institute of Mental Health, University of Heidelberg/Medical Faculty Mannheim, 68159 Mannheim, Germany. [2]Douglas Mental Health University Institute, McGill University, Montreal, Quebec H4H 1R3, Canada.

*These authors contributed equally to this work.

distinct. Current urban living was associated with amygdala activity ($t = 3.6$; Fig. 1a), which increased stepwise from subjects living in the country to those living in small towns, and was highest in city dwellers ($r = 0.55$, $P < 0.001$; Fig. 1b). In contrast, urban upbringing was associated with differential activity in the perigenual ACC (pACC, $t = 3.8$; Fig. 2a), increasing linearly with highest activation in participants entirely brought up in cities ($r = 0.56$, $P < 0.001$; Fig. 2b). Results were not explained by demographic or clinical variables (see Supplementary Methods). These effects were regionally specific as no other brain regions showed any urbanicity association at exploratory thresholds ($P < 0.001$, uncorrected).

To test whether our findings were related to specific aspects of the sample or task, we studied a second cohort of 23 participants (characteristics in Supplementary Table 1b) with a modified stress paradigm. Subjects performed two cognitive tasks (arithmetic and mental rotation) while being continuously visually exposed to disapproving investigator feedback through video (see Supplementary Methods). This experiment fully replicated the findings from the previous sample: city living was again specifically associated with activity in the amygdala and was highest in city dwellers ($t = 3.3$, $P < 0.05$, FWE corrected; Fig. 1c, d), whereas a cluster within the pACC ($t = 4.0$, $P < 0.05$, FWE corrected; Fig. 2c) showed a significant linear correlation with urban upbringing ($r = 0.64$, $P < 0.001$; Fig. 2d). Post hoc, to study the effect of current urbanicity in a larger and better distributed sample, we additionally tested 24 predominantly town and rural dwellers. The analysis of the combined sample again confirmed the effect (Supplementary Fig. 2). As before, no other brain regions showed urbanicity associations even at exploratory thresholds. These findings indicated that specifics of the procedure did not confound the urbanicity effects.

Because acute social stress interacts with cognitive processing[13], the question arose as to whether the observed effect of urbanicity in these

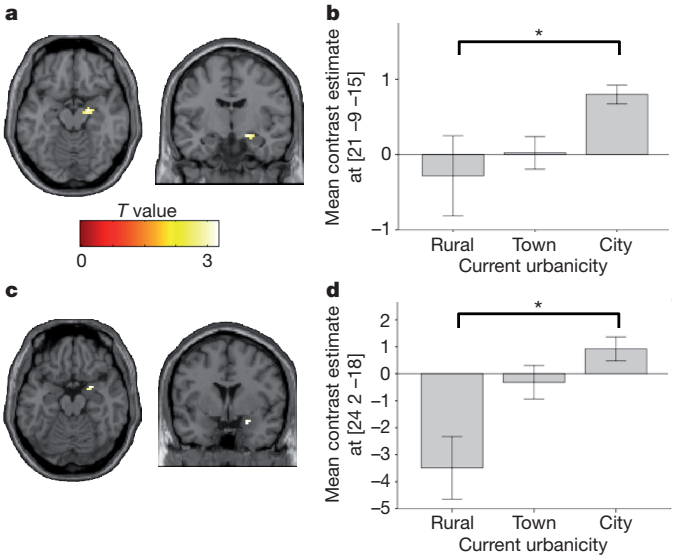


Figure 1 | Relationship between current urbanicity and amygdala activation. **a**, Discovery study ($N = 32$): T map of significant correlations between stress-related activations (in the experimental versus control contrast) and current urbanicity scores shown at a threshold of $P < 0.005$, uncorrected. **b**, Discovery study: contrast estimates at the most significantly correlated voxel in the amygdala (located at $x = 21$, $y = -9$, $z = -15$) for the experimental compared to control contrast for the three current urbanicity groups (*$P < 0.05$; error bars indicate s.e.m.). **c**, Replication study ($N = 23$): T map of significant correlations between activations in the experimental compared to control contrast and current urbanicity scores (shown at $P < 0.05$, FWE corrected for the right amygdala as region of interest (ROI)). **d**, Replication study: contrast estimates at the most significantly correlated voxel in the amygdala (located at $x = 24$, $y = 2$, $z = -18$) for the experimental compared to control contrast for the three current urbanicity groups (*$P < 0.05$, error bars indicate s.e.m.).

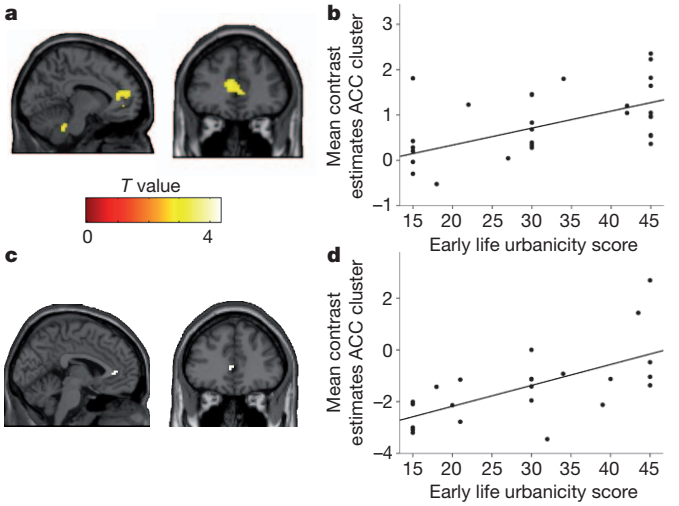


Figure 2 | Relationship between early life urbanicity scores and pACC activation. **a**, Discovery study ($N = 32$): T map of significant correlations between stress-related activations (in the experimental versus control contrast) correlating with urbanicity scores shown at a threshold of $P < 0.005$, uncorrected. **b**, Discovery study: scatterplot of urbanicity scores and mean contrast estimates of the significantly (at $P < 0.005$) correlating voxels within the ACC in the experimental compared to control contrast. Results indicate a linear relationship between these two variables ($r = 0.56$, $P = 0.001$). **c**, Replication study ($N = 23$): T map of significant correlations between activations (in the experimental compared to control contrast) and urbanicity scores (shown at $P < 0.05$, FWE corrected for the rostral ACC as ROI). **d**, Replication study: scatterplot between contrast estimates for the stress compared to control contrast and the urbanicity score shown for the mean of all significantly ($P < 0.005$) correlated voxels ($r = 0.64$, $P < 0.001$).

experiments was related to social evaluative stress per se or to the cognitive tasks used. To investigate this issue, we studied a sample of 37 healthy adults from an ongoing study[17] (Supplementary Table 1b) during a working memory and an emotional face matching task, both without stress. Even at a threshold of $P < 0.01$, uncorrected, there were no voxels within the ACC or amygdala whose activations were correlated to current or early life urbanicity, indicating that our findings were indeed related to social stress with a degree of specificity. In a secondary analysis, we increased the sample size for this experiment to 80 (Supplementary Table 1c), with power to detect an effect of similar size to the discovery study of 0.99. Again, no significant associations with urbanicity were observed.

In this initial examination of the effects of urbanicity, no neural circuit was hypothesized a priori. Nevertheless, the regions identified, controlling for false positives, can plausibly be related to previous epidemiological observations. In the amygdala, activity during social stress was specifically related to city living. The amygdala, which among other functions signals negative affect and environmental threat[18], has been strongly implicated in anxiety disorders, depression, and other behaviours that are increased in cities, such as violence[19]. Conversely, urban upbringing showed a distinct, but equally regionally specific effect on the pACC, a major part of the limbic stress regulation system[18] that exhibits high neuronal glucocorticoid receptor expression[20], modulates hypothalamic–pituitary–adrenal axis activation during stress[10], and is implicated in processing chronic social stressors such as social defeat. In schizophrenia, reduced cingulate grey matter volume[21] has been reported in patients, emerging during adolescence[22]. Connectivity abnormalities of the pACC with the amygdala during processing of affectively negative stimuli were seen in schizophrenic patients, but not in genetically at-risk individuals[23], suggesting a link to environmental factors. Therefore, the epidemiological distinction between current and early life urbanicity maps onto distinct neural regions that are associated with the disease phenotypes implicated by the environmental risk data. A direct link between psychopathology

and these circuits during social evaluative stress must be established in future work, including the study of subjects with mental illness.

In principle, any of the multiple factors related to urban living[1], such as pollution, toxins, crowding, noise, or demographic factors not captured in our analysis, could be responsible for the observed associations. However, in light of the epidemiological evidence that urbanicity is causal for mental disorders[2,3,7], it is interesting to consider the parsimonious proposal that social stress contributes causally to the impact of urbanicity on the neural circuits identified here. Importantly, although urbanicity effects on the pACC and amygdala were dissociable, these two structures are functionally closely linked: the pACC is a key regulatory region for amygdala[18] activity in the context of negative affect, which is critical for gene–environment interactions[9] and extinction. Further, synaptic and neuronal remodelling of the pACC and amygdala have been described in animals exposed to social stressors[24], and amygdala and cingulate volume relate to social network size in humans[25]. Therefore, we investigated functional connectivity between the pACC and amygdala during the MIST stress paradigm using previously published methods[17]. Urban upbringing was associated with reduced connectivity (Spearman's Rho $= -0.39$, $P = 0.013$; see Supplementary Fig. 3), whereas current urbanicity had no effect, supporting an effect of early urban exposure on this regulatory circuit. Because early life neurochemical alterations in the serotonin system linked to social support have enduring effects on the cingulate in animals[26] and humans[27], these differential effects of early life and current urbanicity may reflect a developmental vulnerability of the cingulate. In line with this, the cortisol stress response has been found to be exaggerated in human adults who were exposed to maternal stress *in utero*[28].

Beyond mental illness, our data are of general interest in showing a link between cities and social stress sensitivity. This indicates that an experimental approach to dissecting epidemiological associations is feasible and that it could be used to characterize further the underlying psychosocial components; for example, the effects of finer-grained quantifiers of individuals' social networks or individual social experience in urban contexts. One such potential component is unstable hierarchical position—a social stressor related to general health that might be relevant in the context of increased socioeconomic disparities in cities—which also affects medial prefrontal cortex and amygdala function[29]. Further, 'prosocial' neuropeptides, molecular mediators of social interactions, modulate the pACC-amygdala circuit[30], indicating that social risk and protective factors might converge on this system.

This first series of studies of the neural effects of urbanicity has several limitations. First, our cross-sectional study does not prove that the observed association is causal. Second, our subjects grew up in relative safety and prosperity in Germany, a developed country, whereas greater urban–rural discrepancies are found elsewhere; however, this would have probably attenuated our findings. Third, the pronounced differences in brain processing did not correlate with cortisol levels, possibly reflecting the greater sensitivity of neural measures compared to downstream peripheral markers; nevertheless, the cortisol stress response should be studied in a larger sample. Fourth, the lack of random population sampling and the fact that our confirmation study largely used college students potentially limits the generalizability of our findings, making replication of these results in different and larger samples important.

Our data reveal neural effects of urban upbringing and habitation on social stress processing in humans. These findings contribute to our understanding of urban environmental risk for mental disorders and health in general. Further, they point to a new empirical approach for integrating social sciences, neurosciences and public policy to respond to the major health challenge of urbanization.

METHODS SUMMARY

The study was approved by the Ethics Committee of Heidelberg University. Three groups of healthy participants were studied after written informed consent. A thorough clinical and psychiatric examination was performed to exclude relevant illness. As in previous work, urbanicity was scored as follows[4]: city with more than 100,000 inhabitants (3); town with more than 10,000 inhabitants (2); and rural area (1). For urban upbringing until age 15, these numbers were multiplied by the years spent in each area and added.

Blood-oxygen-level-dependent (BOLD) fMRI was performed on a 3.0 Tesla Siemens Trio scanner using an echo-planar-imaging (EPI) sequence and analysed using SPM5 (MIST study and non-stress control study) and SPM8 (stress replication study) (http://www.fil.ion.ucl.ac.uk/spm). In the discovery study, participants performed the MIST, in the replication study a variant social stress task (see Supplementary Methods), in the control study a working memory task (the n-back task) and an emotional face matching task.

All imaging results were corrected for multiple comparisons at a significance level of $P < 0.05$ via FWE. For main task effects, correction was performed over the whole brain. For hypothesis-driven analyses, bilateral a priori anatomical regions of interest (ROI) were taken from the Harvard Oxford Atlas (http://www.cma.mgh.harvard.edu). For correlations with cortisol, the amygdala, hypothalamus and ACC were specified based on previous results[16]. For urbanicity analyses, on the basis of the correlations observed with early life and current urbanicity in the discovery study (MIST), we defined a priori anatomical ACC and amygdala ROI for the replication and control studies.

Received 3 December 2010; accepted 12 May 2011.

1. Dye, C. Health and urban living. *Science* **319,** 766–769 (2008).
2. Peen, J., Schoevers, R. A., Beekman, A. T. & Dekker, J. The current status of urban-rural differences in psychiatric disorders. *Acta Psychiatr. Scand.* **121,** 84–93 (2010).
3. Krabbendam, L. & van Os, J. Schizophrenia and urbanicity: a major environmental influence–conditional on genetic risk. *Schizophr. Bull.* **31,** 795–799 (2005).
4. Mortensen, P. B. *et al.* Effects of family history and place and season of birth on the risk of schizophrenia. *N. Engl. J. Med.* **340,** 603–608 (1999).
5. Pedersen, C. B. & Mortensen, P. B. Evidence of a dose-response relationship between urbanicity during upbringing and schizophrenia risk. *Arch. Gen. Psychiatry* **58,** 1039–1046 (2001).
6. van Os, J., Pedersen, C. B. & Mortensen, P. B. Confirmation of synergy between urbanicity and familial liability in the causation of psychosis. *Am. J. Psychiatry* **161,** 2312–2314 (2004).
7. van Os, J., Kenis, G. & Rutten, B. P. The environment and schizophrenia. *Nature* **468,** 203–212 (2010).
8. Selten, J. P. & Cantor-Graae, E. Social defeat: risk factor for schizophrenia? *Br. J. Psychiatry* **187,** 101–102 (2005).
9. Pezawas, L. *et al.* 5-HTTLPR polymorphism impacts human cingulate–amygdala interactions: a genetic susceptibility mechanism for depression. *Nature Neurosci.* **8,** 828–834 (2005).
10. Diorio, D., Viau, V. & Meaney, M. J. The role of the medial prefrontal cortex (cingulate gyrus) in the regulation of hypothalamic-pituitary-adrenal responses to stress. *J. Neurosci.* **13,** 3839–3847 (1993).
11. Weinberger, D. R. Implications of normal brain development for the pathogenesis of schizophrenia. *Arch. Gen. Psychiatry* **44,** 660–669 (1987).
12. Meyer-Lindenberg, A. From maps to mechanisms through neuroimaging of schizophrenia. *Nature* **468,** 194–202 (2010).
13. Dickerson, S. S. & Kemeny, M. E. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychol. Bull.* **130,** 355–391 (2004).
14. Dedovic, K. *et al.* The Montreal Imaging Stress Task: using functional imaging to investigate the effects of perceiving and processing psychosocial stress in the human brain. *J. Psychiatry Neurosci.* **30,** 319–325 (2005).
15. Lederbogen, F. *et al.* Salivary cortisol in a middle-aged community sample: results from 990 men and women of the KORA-F3 Augsburg study. *Eur. J. Endocrinol.* **163,** 443–451 (2010).
16. Pruessner, J. C. *et al.* Deactivation of the limbic system during acute psychosocial stress: evidence from positron emission tomography and functional magnetic resonance imaging studies. *Biol. Psychiatry* **63,** 234–240 (2008).
17. Esslinger, C. *et al.* Neural mechanisms of a genome-wide supported psychosis variant. *Science* **324,** 605 (2009).
18. LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23,** 155–184 (2000).
19. Meyer-Lindenberg, A. *et al.* Neural mechanisms of genetic risk for impulsivity and violence in humans. *Proc. Natl Acad. Sci. USA* **103,** 6269–6274 (2006).
20. Herman, J. P., Ostrander, M. M., Mueller, N. K. & Figueiredo, H. Limbic system mechanisms of stress regulation: hypothalamo-pituitary-adrenocortical axis. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **29,** 1201–1213 (2005).
21. Wright, I. C. *et al.* Meta-analysis of regional brain volumes in schizophrenia. *Am. J. Psychiatry* **157,** 16–25 (2000).
22. Vidal, C. N. *et al.* Dynamically spreading frontal and cingulate deficits mapped in adolescents with schizophrenia. *Arch. Gen. Psychiatry* **63,** 25–34 (2006).
23. Rasetti, R. *et al.* Evidence that altered amygdala activity in schizophrenia is related to clinical state and not genetic risk. *Am. J. Psychiatry* **166,** 216–225 (2009).
24. Poeggel, G. *et al.* Juvenile emotional experience alters synaptic composition in the rodent cortex, hippocampus, and lateral amygdala. *Proc. Natl Acad. Sci. USA* **100,** 16137–16142 (2003).

25. Bickart, K. C., Wright, C. I., Dautoff, R. J., Dickerson, B. C. & Barrett, L. F. Amygdala volume and social network size in humans. *Nature Neurosci.* **14,** 163–164 (2011).
26. Spinelli, S. *et al.* Early-life stress induces long-term morphologic changes in primate brain. *Arch. Gen. Psychiatry* **66,** 658–665 (2009).
27. Cohen, R. A. *et al.* Early life stress and morphometry of the adult anterior cingulate cortex and caudate nuclei. *Biol. Psychiatry* **59,** 975–982 (2006).
28. Entringer, S., Kumsta, R., Hellhammer, D. H., Wadhwa, P. D. & Wust, S. Prenatal exposure to maternal psychosocial stress and HPA axis regulation in young adults. *Horm. Behav.* **55,** 292–298 (2009).
29. Zink, C. F. *et al.* Know your place: neural processing of social hierarchy in humans. *Neuron* **58,** 273–283 (2008).
30. Zink, C. F., Stein, J. L., Kempf, L., Hakimi, S. & Meyer-Lindenberg, A. Vasopressin modulates medial prefrontal cortex–amygdala circuitry during emotion processing in humans. *J. Neurosci.* **30,** 7017–7022 (2010).

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

Acknowledgements We thank D. Gass and C. Niemeyer for technical assistance. We also thank C. Sauer, O. Grimm, M. Plichta and A. Schäfer for support on data analyses. The research leading to these results has received funding from the European Community's Seventh Framework Programme under grant agreement No. HEALTH-F2-2010-241909 (Project EU-GEI), German Research Foundation (Deutsche Forschungsgemeinschaft SFB 636-B7) and Federal Ministry of Education and Research (MooDS) to A.M.L. EU-GEI is an acronym for the project ''European network of National Schizophrenia Networks Studying Gene–Environment Interactions''. F.S. is a member of the International Research Training Group ''Psychoneuroendocrinology of Stress'', Institute of Psychobiology, University of Trier, Germany, granted by the German Research Foundation.

Author Contributions F.L., P.K and L.H. designed and performed experiments, analysed data and wrote the paper; F.S., P.S. and S.W. designed and performed experiments, analysed data and reviewed the manuscript; H.T. analysed data and reviewed the manuscript; M.D. and M.R. designed experiments and reviewed the manuscript; J.C.P. developed the MIST paradigm and reviewed the manuscript. A.M.-L. obtained funding, designed the study and experiments and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.M.-L. (a.meyer-lindenberg@zi-mannheim.de).

LETTER

doi:10.1038/nature09992

Non-adaptive origins of interactome complexity

Ariel Fernández[1,2] & Michael Lynch[3]

The boundaries between prokaryotes, unicellular eukaryotes and multicellular eukaryotes are accompanied by orders-of-magnitude reductions in effective population size, with concurrent amplifications of the effects of random genetic drift and mutation[1]. The resultant decline in the efficiency of selection seems to be sufficient to influence a wide range of attributes at the genomic level in a non-adaptive manner[2]. A key remaining question concerns the extent to which variation in the power of random genetic drift is capable of influencing phylogenetic diversity at the subcellular and cellular levels[2–4]. Should this be the case, population size would have to be considered as a potential determinant of the mechanistic pathways underlying long-term phenotypic evolution. Here we demonstrate a phylogenetically broad inverse relation between the power of drift and the structural integrity of protein subunits. This leads to the hypothesis that the accumulation of mildly deleterious mutations in populations of small size induces secondary selection for protein–protein interactions that stabilize key gene functions. By this means, the complex protein architectures and interactions essential to the genesis of phenotypic diversity may initially emerge by non-adaptive mechanisms.

Here we examine whether established gene orthologies reveal a role for drift in phylogenetic patterns of protein structural evolution. Although evolutionary change at the structural level is unlikely to destabilize greatly the native fold of an essential protein, as the complete loss of function would generally be unbearable, the drift hypothesis predicts a negative relation between population size (N) and the accumulation of mildly deleterious amino-acid substitutions. The following examination of the structures of orthologous proteins from vastly different lineages suggests that the enhanced power of drift in eukaryotes (multicellular species in particular) results in a qualitative reduction in the stability of protein–water interfaces (PWIs) through the partial exposure of paired backbone polar groups (amides and carbonyls) that are otherwise protected in prokaryotes. In effect, the reduced efficiency of selection in small-N species encourages the accumulation of mild structural deficiencies in the form of solvent-accessible backbone hydrogen bonds (SABHBs), which lead to protein structures that are more 'open' and vulnerable to fold-disruptive hydration (Fig. 1a) and create protein–water interfacial tension (PWIT; Supplementary Fig. 1)[5] by hindering the hydrogen-bonding capabilities of nearby water molecules.

We argue that the emergence of unfavourable PWIs promotes the secondary recruitment of novel protein–protein associations that restore structural stability by reducing PWI. Under this hypothesis, complex organisms may frequently develop protein–protein interactions not as immediate vehicles for novel adaptive functions, but as compensatory mechanisms for retaining key gene functions. Once in place, such physical contact between interacting proteins may provide a selective environment for the further emergence of entirely novel protein–protein interactions underlying cellular and organismal complexities. Our suggestion that the hallmark of eukaryotic evolution, the origin of interactome complexity, may have arisen in part as a passive consequence of the enhanced power of drift reduces the need to invoke direct long-term selective advantages of phenotypic complexity[6].

To gain insight into the evolution of interactome complexity, we derived quantitative measures of the PWIT as indicators of potential molecular interactivity. To estimate the PWIT of a protein, we computationally equilibrated the protein structure in surrounding water, using the function $g(\mathbf{r})$ to represent the time-averaged coordination (number of hydrogen bonds) associated with a water molecule at position $\mathbf{r}$ (Fig. 1a), and integrating over the entire protein surface all water molecules within a 10 Å radius (the thickness of four layers of water molecules). Compared with bulk water (where $g = 4$), interfacial water molecules may have reduced hydrogen-bonding opportunities ($g < 4$) and often counterbalance these losses by interacting with polar groups on the protein surface. Thus the PWIT parameter integrates information on unfavourable local decreases in g and favourable polarization contributions from the protein to yield the free-energy cost, ΔG_{if}, of spanning the protein–water interface (Methods). A high PWIT signals a high propensity for protein–protein associations, which reduce the PWI area.

To validate the use of PWIT as a measure of interactivity, we examined an exhaustive catalogue of contact topologies for protein complexes with one to six subunits, with each topology being evaluated with one or more non-homologous complexes using structures in the Protein Data Bank (PDB) (Supplementary Table 1). For each complex, we computed the total protein–protein interface area after identifying the residues engaged in intermolecular contacts[7]. For each protein subunit, the protein–protein interface is contained within the PWI region that generates tension in the free subunit, and there is a tight correlation between the surface areas for both regions, implying that regions on the protein surface generating PWIT (i.e. those with $g < 4$ for nearby water) actually promote associations (Supplementary Figs 1 and 2a). Next, we verified that protein surface regions generating PWIT coincide with the affinity-contributing regions at protein–protein interfaces. To this end, we tested the value of PWIT as a promoter of protein associations by focusing on the interface for the 1:1 human growth hormone (hGH)-receptor complex[8] (Supplementary Fig. 2b) for which the consequences of amino-acid substitutions have been extensively evaluated. Our analysis reveals a strong correlation between the change in PWIT induced by site-specific mutagenesis of interfacial residues and the association free-energy difference created by the alteration of the hormone–receptor interface (Supplementary Fig. 2c).

Comparison of orthologous proteins engaging in different levels of homo-oligomerization in different species[9] further supports the view that PWIT serves as a measure of the propensity for protein–protein association. The ratio of protein–protein interface areas (lower to higher degrees of complexation; Supplementary Table 2) exhibits a strong positive correlation with the ratio of PWITs for the respective free subunits (Fig. 1b). As complexes with higher degrees of oligomerization arise from lower-order complexes, this implies that the degree of cooperativity among subunits correlates with the PWIT of the basic subunit.

Hydrophobic regions on protein surfaces obviously contribute to PWIT, but analysis of proteins exhibiting association propensity (Supplementary Table 2) shows that the regions generating 73 ± 5%

[1]Department of Computer Science, The University of Chicago, Chicago, Illinois 60637, USA. [2]Department of Bioengineering, Rice University, Houston, Texas 77005, USA. [3]Department of Biology, Indiana University, Bloomington, Indiana 47405, USA.

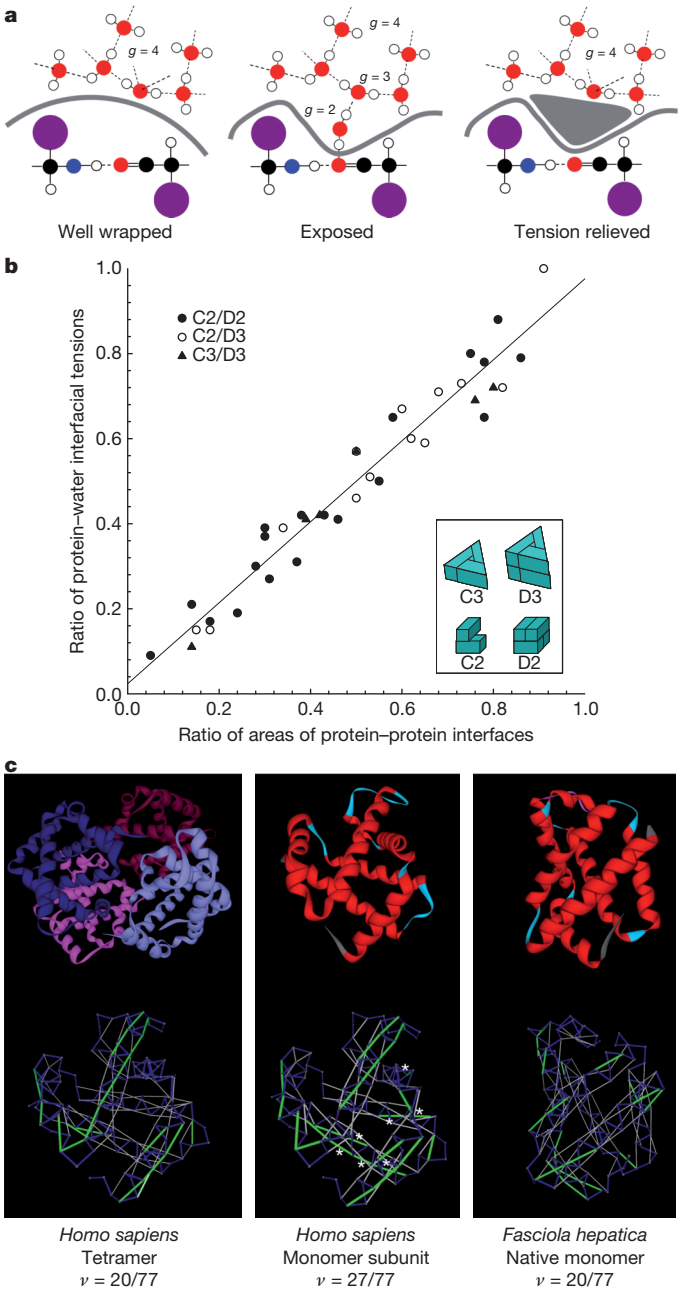


Figure 1 | Structural deficiencies in soluble proteins promote protein associations. **a**, Hydration of exposed polar backbone induces interfacial tension by causing water molecules near the defect to relinquish part of their coordination ($g < 4$) relative to the level in surrounding bulk solvent ($g = 4$). White represents hydrogen atoms; red, oxygen; blue, nitrogen; black, carbon; the larger purple circles denote side chains for amino acids. Hydrogen bonds are denoted by dashed lines. Thick grey lines outline the external surface of the overall protein molecule, and the underlying structure represents two amino acids made adjacent by the protein architecture and bound by a hydrogen bond between the backbone amide (blue:white) of one amino acid and carbonyl (red:black) of the other. Water molecules are shown as angular red and white segments, with the coordination number g denoting the number of hydrogen bonds associated with a water molecule ($g = 4$ for bulk water; $g < 4$ for confined interfacial water). In the centre, the structure of the protein causes local exposure and unfavourable hydration of the polar backbone, whereas the absence of such local interactions between water molecules and the well-wrapped proteins on the left and right reduces interfacial tension (interfacial water is bulk-like, retaining the maximum coordination $g = 4$). **b**, Comparison of orthologous proteins with different levels of homo-oligomerization reveals that the PWIT is an indicator of the propensity for cooperative improvement/refinement of protein function through complexation. The ratio of protein–protein interfaces (small to large) was determined for pairs of orthologous proteins with different levels of oligomerization in different species (Supplementary Table 2) and plotted against the ratio of PWITs for the respective free subunits. The tight correlation ($r^2 = 0.94$) reveals that interspecific differences in PWIT accompany differences in levels of oligomerization, thus providing a measure of potential allosteric or cooperative improvement of basic protein function. Complexes with cyclic rotational symmetry (C2, C3, …) can further oligomerize into complexes with dihedral (D2, D3, …) symmetry, as shown in the idealized diagrams in the lower right. For example, C2 complexes can dimerize into D2 complexes, trimerize into D3 complexes, etc., whereas a D3 complex can also be obtained by dimerization of a C3 complex. For the protein–protein interface and PWIT ratios examined, the interface for the subunit in the complex with lower-order symmetry is compared with that in the complex with higher-order symmetry, yielding analyses based on protein pairs contrasted within three groupings: C2 versus D2, C2 versus D3, and C3 versus D3. **c**, The SABHB patterns from two haemoglobins with different oligomerization levels in their native states are compared. In the bottom panels, the protein backbone is represented by virtual bonds in blue joining α-carbons, with well-protected BHBs shown as light grey and SABHBs as green lines joining the α-carbons of the paired residues. The ribbon representations of the human complex and dissociated subunit (chain A in PDB.2DN2, left and centre, respectively) are included as aids to the eye, representing the structuring of the backbone in each subunit. The free subunit isolated from the tetramer in *H. sapiens* (PDB.2DN2, chain A, centre) has seven excess SABHBs (denoted by stars) when compared with the subunit within the tetrameric complex, where they are well-protected intermolecularly, alleviating interfacial tension. As a consequence of this better wrapping, the overall extent of structural deficiency (ν value) for the subunit within the human complex is identical to that of the natively monomeric haemoglobin from the trematode *F. hepatica* (PDB.2VYW). This raises the possibility that the accumulation of structural deficiencies in the mammalian haemoglobin subunit promoted the emergence of an oligomeric association as a means of reducing excess interfacial tension. The structural displays were obtained by uploading the PDB text files into the program YAPview, a displayer of local backbone desolvation of soluble proteins that can be downloaded from the link 'Dehydron Calculator' at http://www.owlnet.rice.edu/~arifer/.

of the PWIT (Supplementary Information and Supplementary Fig. 1) arise from SABHBs. The resultant hydration of backbone polar groups (amides and carbonyls) causes a loss of coordination for local water molecules, which increases surface tension and creates an unstable PWI, as the cavities cannot accommodate a bulk-like water molecule[10]. As an example of how such a structural deficiency can be alleviated through a protein association, an isolated α-subunit of the human haemoglobin tetramer has seven SABHBs that become protected within the tetrameric complex, such that the ratio (ν) of SABHBs to total BHBs in the complex-associated subunit is the same as that in the natively monomeric unit for haemoglobin from the trematode *Fasciola hepatica* (Fig. 1c).

To evaluate whether the accumulation of structural deficiencies of proteins is generally encouraged by random genetic drift, and in turn enhances the propensity for establishing protein complexes, we examined a set of 106 orthologous water-soluble proteins (sequence identity greater than 30%)[11,12] with PDB-reported structures for at least two species. We considered 36 species with vastly different population sizes[1,2], each containing proteins in at least 90 of the 106 orthologous groups (Supplementary Tables 3–5). Template-based three-dimensional structures for orthologues lacking PDB-reported structures were constructed by homology threading[13,14], and evaluated, ranked and selected according to the energetic proximity between template and model[15]. The accuracy of this homology-based prediction of PWIT was determined with a test set of proteins with PDB-reported structures from two species, subjecting one member of each orthologous pair to homology threading through the other. Comparison of the indirect and direct estimates of PWIT demonstrates that when sequence identities are greater than 35%, the predicted PWIT diverges less than 10% from the more direct estimate for the same protein (Supplementary Fig. 3).

For each protein structure, $g(\mathbf{r})$ was obtained as described in Methods, and the relative propensities for protein association across orthologues were then determined by assessing differences in the free-energy cost

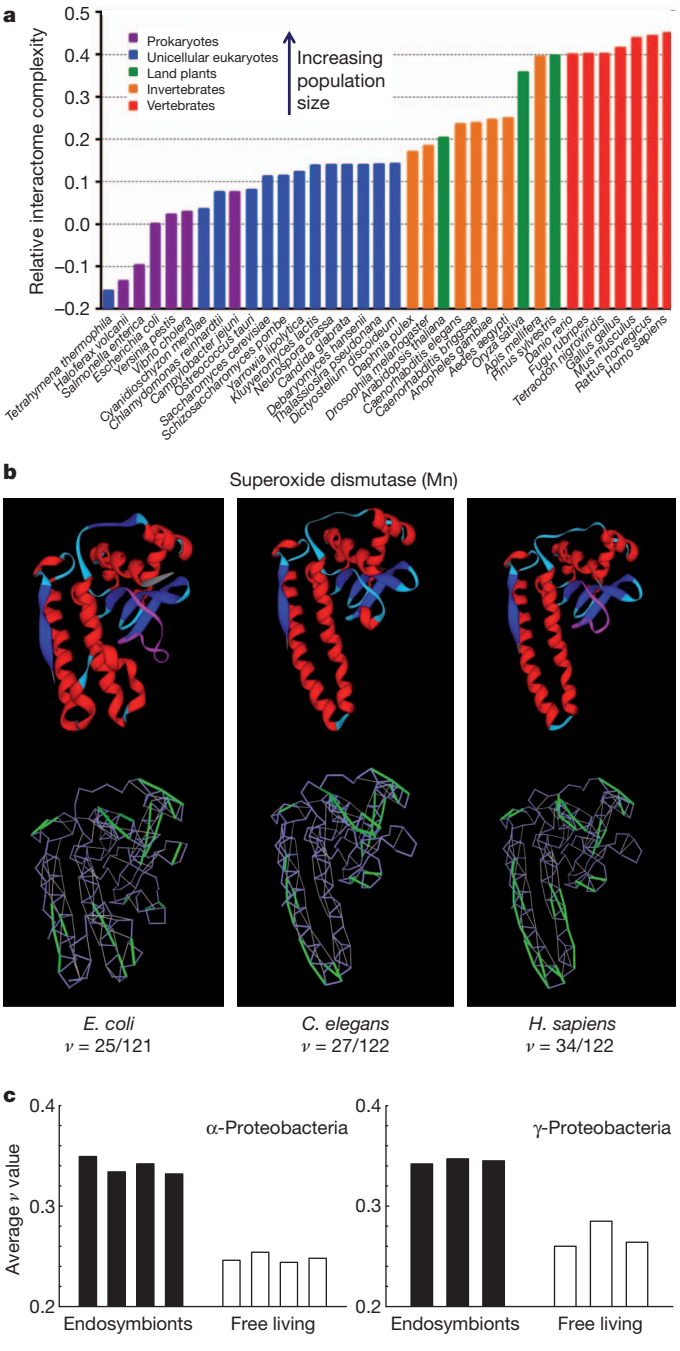


Figure 2 | Structural degradation enhances PWIT and promotes protein interactivity in species with small population sizes. **a**, Potential for interactome complexity of 36 species with diverse population sizes (Supplementary Table 2), relative to *E. coli*. To highlight the relative power of random genetic drift, bars are colour-coded to reflect groupings of species in broad population-size categories. **b**, Overall structural deficiency of orthologues of the enzyme superoxide dismutase (Mn), revealing a progressive accumulation of SABHBs in the orthologues of the bacterium *E. coli*, the nematode *Caenorhabditis elegans* and *H. sapiens*. The upper ribbon representations illustrate the structural conservation across orthologues (respective PDB accession numbers 3ot7, 3dc6, 2adq). The conventional colour coding is red, blue, magenta and light blue for helix, β-strand, loop and turn, respectively. **c**, Average structural deficiency (ν value) of protein orthologues for intracellular and free-living bacterial species. Species identities, progressing from left to right are as follows: α-Proteobacteria—*Rickettsia typhi*, *Orientia tsutsugamushi*, *Anaplasma centrale* str. Israel, *Wolbachia* sp. wRi, *Rhodospirillum centenum* SW, *Magnetospirillum magneticum*, *Silicibacter TM1040*, *Erythrobacter litoralis*; γ-Proteobacteria—*Buchnera aphidicola*, *Wigglesworthia brevipalpis*, Candidatus *Blochmannia pennsylvanicus*, *Marinomonas* MWYL1, *E. coli*, *Pseudomonas aeruginosa*. Only proteins with orthologues across the full set of species within each group were considered for analysis (Supplementary Tables 6 and 7).

ΔG_{if} among species. We estimated the relative complexation propensity $M_{j,n}$ of a protein in orthologue group j (1, …, 106) from species n (1, …, 36) by adopting *Escherichia coli* as a reference species ($n = 1$):

$$M_{j,n} = [(\Delta G_{if})_{j,n} - (\Delta G_{if})_{j,1}]/(\Delta G_{if})_{j,1} \quad (1)$$

With this index, $M_{j,1} = 0$ for all proteins in *E. coli*, and taxa with less well-wrapped proteins (and hence greater propensity for complexation), have positive values.

The mean value of species-specific estimates of $M_{j,n}$ over all proteins evaluated is negatively correlated with the approximate effective population sizes of species (Fig. 2a), given that the average ranking of the latter is prokaryotes > unicellular eukaryotes > invertebrates > vertebrates and land plants[1,2]. A specific example of a trend towards increasing structural openness with reduced population sizes is illustrated in Fig. 2b, where the SABHB patterns and ν values for orthologues of the enzyme superoxide dismutase are compared across three species.

The results from Fig. 2a and an additional analysis (Supplementary Fig. 4) support the hypothesis that large organisms with small population sizes experience a significant enough increase in the power of random genetic drift to magnify the accumulation of mild structural deficiencies in the form of SABHBs, resulting on average in proteins with a more solvent-exposed or 'open' structure. By contrast, mutations to SABHBs are more frequently excluded by selection in species with larger population sizes (for example, prokaryotes). Thus, because SABHBs are the main determinants of interfacial tension (Supplementary Fig. 1), the proteins of large organisms have a greater inherent tendency to form novel protein–protein associations (Fig. 1a). This suggests that increases in protein-network complexity in multicellular species may in part owe their origins to modifications to the intracellular selective environment induced by non-adaptive structural degradation of individual proteins.

One concern with the preceding interpretation is the order of events: does an initial degradation of architectural integrity of individual proteins in response to random genetic drift induce secondary selection for the recruitment of interacting partners, or does the emergence of cellular complexity (and increased protein interactivity) precede secondary changes in protein sequence to accommodate such interactions? One way to evaluate this matter is to compare proteins from related species that have experienced relatively recent divergences in effective population sizes but no major modifications in intracellular complexity or emergence of multicellularity.

To achieve this task, we compared orthologous genes from endosymbiotic/intracellular bacteria and their free-living relatives, as the former are thought to have experienced substantial reductions in effective population sizes[16]. Previous suggestions that intracellular bacteria experience elevated levels of random genetic drift have been based on ratios of substitution rates at silent and replacement sites, which can be biased indicators of the efficiency of selection if there is selection on silent sites. Although the lack of protein structural information for endosymbiotic species requires a sequence-based identification of SABHBs derived from reliable scores of native disorder propensity (Methods), the resultant analyses are broadly consistent with the hypothesis that an increase in the power of drift in microbes encourages the accumulation of structural defects in protein architecture (Fig. 2c). Free-living species, with larger effective population sizes, have consistently smaller ν values for orthologous genes in both α- and γ-Proteobacteria. (Application of the same sort of analysis of disorder propensity across a set of 105 species and 541 proteins corroborates this result (Supplementary Figs 5–7).)

Taken together, our analyses support the hypothesis that the range of population sizes experienced by natural populations is sufficient to

induce significantly different patterns of evolution at the level of protein architecture. The resultant changes in the intracellular environment in small-N species provides an opportunity for the recruitment of stabilizing protein–protein interactions, yielding a plausible mechanism for the emergence of molecular complexities before their exploitation in phenotypic divergence[9,17]. This hypothesis does not deny a potentially significant role for natural selection in using such novelties subsequent to their establishment, nor does it deny the fact that intramolecular compensatory mutations can alleviate some structural defects associated with SABHBs. However, our results do raise questions about the necessity of invoking an intrinsic advantage to organismal complexity, and provide a strong rationale for expanding comparative studies in molecular evolution beyond linear sequence analysis to evaluations of molecular structure.

METHODS SUMMARY

We determined the propensity of proteins to be engaged in associations that reduce the PWI by computing the PWIT. This thermodynamic parameter gives ΔG_{if}, the free-energy cost of spanning the PWI. The PWIT is computed as

$$\Delta G_{if} = \tfrac{1}{2}\int\{a|\nabla g|^2 - |\mathbf{P}[g(\mathbf{r})]|^2\}\mathrm{d}\mathbf{r}, \quad (2)$$

where the term $\frac{1}{2}a|\nabla g|^2$, with a = 9.02 mJ m^{-1} at T = 298 K (Methods), accounts for tension-generating reductions in water coordination, and the polarization $\mathbf{P}[g(\mathbf{r})]$ accounts for dipole–electrostatic field interactions (Methods). For a given protein structure or template-based structural model, the field $g = g(\mathbf{r})$ used in the numerical integration of equation (2) was determined by equilibrating the water-embedded structure within an isothermal–isobaric (NPT) ensemble (with fixed parameters N = number of particles, P = pressure and T = temperature; Methods)[10,18,19]. From structural coordinates, we determined the structural deficiencies (SABHBs)[20] that generate 73±5% of the PWIT (Supplementary Information). We examined 106 groups of orthologous proteins identified using OrthoMCL[11,12] for which there are PDB representatives from at least two species (usually *E. coli* and *Homo sapiens*, Supplementary Tables 3–5). We considered 36 representative species, each containing proteins in at least 90 of the 106 orthologue groups. Template-based three-dimensional structures for orthologues lacking a PDB-reported structure[14] were constructed using MODELLER[13], with side chains directly positioned with SCWRL[21]. The template and resulting model were evaluated, ranked and finally selected using ProSA[15]. The accuracy of homology models is shown in Supplementary Fig. 3. In cases where orthologous structural templates were unavailable, like the comparison of endosymbionts with free-living species, a sequence-based inference of SABHBs was performed based on an established anti-correlation between backbone protection and disorder propensity (Supplementary Fig. 5)[22]. The cross validation of homology- and disorder-based estimations of v values is given in Supplementary Fig. 6.

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

Received 24 August 2010; accepted 9 March 2011.
Published online 18 May 2011.

1. Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302,** 1401–1404 (2003).
2. Lynch, M. *The Origins of Genome Architecture* (Sinauer, 2007).
3. Stoltzfus, A. On the possibility of constructive neutral evolution. *J. Mol. Evol.* **49,** 169–181 (1999).
4. Gray, M. W., Lukes, J., Archibald, J. M., Keeling, P. J. & Doolittle, W. F. Cell biology. Irremediable complexity? *Science* **330,** 920–921 (2001).
5. Rowlinson, J. S. & Widom, B. *Molecular Theory of Capillarity* (Oxford Univ. Press, 1982).
6. Lynch, M. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl Acad. Sci. USA* **104** (Suppl.), 8597–8604 (2007).
7. Levy, E. D., Pereira-Leal, J. B., Chothia, C. & Teichmann, S. A. 3D Complex: a structural classification of protein complexes. *PLOS Comput. Biol.* **2,** e155 (2006).
8. Clackson, T., Ultsch, M. H., Wells, J. A. & de Vos, A. M. Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. *J. Mol. Biol.* **277,** 1111–1128 (1998).
9. Levy, E. D., Boeri Erba, E., Robinson, C. V. & Teichmann, S. A. Assembly reflects evolution of protein complexes. *Nature* **453,** 1262–1265 (2008).
10. Fenimore, P. W., Frauenfelder, H., McCammon, B. H. & Young, R. D. Bulk solvent and hydration-shell fluctuations, similar to α- and β-fluctuations in glasses, control protein motions and functions. *Proc. Natl Acad. Sci. USA* **101,** 14408–14413 (2004).
11. Ostlund, G. *et al.* InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res.* **38,** D196–D203 (2010).
12. Gabaldon, T. *et al.* Joining forces in the quest for orthologs. *Genome Biol.* **10,** 403 (2009).
13. Sali, A. & Blundell, T. L. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* **234,** 779–815 (1993).
14. Zhou, H. & Skolnick, J. Improving threading algorithms for remote homology modeling by combining fragment and template comparisons. *Proteins* **78,** 2041–2048 (2010).
15. Wiederstein, M. & Sippl, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **35,** W407–W410 (2007).
16. Moran, N. A. Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. *Proc. Natl Acad. Sci. USA* **93,** 2873–2878 (1996).
17. Kuriyan, J. & Eisenberg, D. The origin of protein interactions and allostery in colocalization. *Nature* **450,** 983–990 (2007).
18. Rizzo, R. C. & Jorgensen, W. L. OPLS All-atom model for amines: resolution of the amine hydration problem. *J. Am. Chem. Soc.* **121,** 4827–4836 (1999).
19. Jorgensen, W. L., Chandrasekhar, J., Madura, J., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79,** 926–935 (1983).
20. Fernández, A. & Berry, R. S. Golden rule for buttressing vulnerable soluble proteins. *J. Proteome Res.* **9,** 2643–2648 (2010).
21. Canutescu, A. A., Shelenkov, A. & Dunbrack, R. L. A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci.* **12,** 2001–2014 (2003).
22. Pietrosemoli, N., Crespo, A. & Fernández, A. Dehydration propensity of order-disorder intermediate regions in soluble proteins. *J. Proteome Res.* **6,** 3519–3526 (2007).

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

Acknowledgements A.F. was supported by National Institutes of Health grant R01GM072614, and by the Institute of Biophysical Dynamics and the Department of Computer Science at The University of Chicago. M.L. was supported by National Institutes of Health grant R01GM036827 and National Science Foundation grant EF-0827411.

Author Contributions A.F. and M.L. conceived the project and wrote the paper. A.F. collected the orthologue groups across 36 species with sufficient structural representation, performed the structural analysis and determined the interaction propensities across orthologues.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.L. (milynch@indiana.edu) or A.F. (ariel@uchicago.edu).

METHODS

Computation of PWIT. The parameter a in equation (2) is obtained from the interfacial tension of a large non-polar sphere with radius θ in the limit $\theta/1\,\mathrm{nm} \to \infty$. Thus we get $a = 9.02\,\mathrm{mJ\,m^{-1}} = \lim_{\theta/1\,\mathrm{nm}\to\infty}[\gamma(4\pi\theta^2)/\int \tfrac{1}{2}|\nabla g|^2 \mathrm{d}\mathbf{r}]$, where $\gamma = 72\,\mathrm{mJ\,m^{-2}}$ is the bulk surface tension of water at 298 K, and $\int|\nabla g|^2 \mathrm{d}\mathbf{r} = O(\theta^2)$ since $\nabla g \neq 0$ only in the vicinity of the interface. To determine the g-dependence of polarization $\mathbf{P} = \mathbf{P}(\mathbf{r})$, we adopt the Fourier-conjugate frequency space ($\boldsymbol{\omega}$ space) and represent the dipole correlation kernel $K_\mathrm{p}(\boldsymbol{\omega})$ and the electrostatic field $\mathbf{E} = \mathbf{E}(\mathbf{r})$ in this space. In contrast with other treatments[23], we note that $\mathbf{P}$ and $\mathbf{E}$ are indeed proportional but the proportionality constant is $\boldsymbol{\omega}$ dependent[24]. Thus, in $\boldsymbol{\omega}$ space, we get

$$F(\mathbf{P})(\boldsymbol{\omega}) = K_\mathrm{p}(\boldsymbol{\omega})F(\mathbf{E})(\boldsymbol{\omega}), \tag{3}$$

where F denotes three-dimensional Fourier transform $F(\mathbf{f})(\boldsymbol{\omega}) = (2\pi)^{-3/2}\int \mathrm{e}^{i\boldsymbol{\omega}.\mathbf{r}}\mathbf{f}(\mathbf{r})\mathrm{d}\mathbf{r}$, and the kernel $K_\mathrm{p}(\boldsymbol{\omega})$ is the Lorentzian $K_\mathrm{p}(\boldsymbol{\omega}) = (\varepsilon_\mathrm{b} - \varepsilon_\mathrm{o})/(1 + (\tau(\mathbf{r})c)^2|\boldsymbol{\omega}|^2)$, with $\tau(\mathbf{r})c$ = position-dependent dielectric relaxation scale ≈ 3 cm for $\tau = \tau_\mathrm{b} \approx 100$ ps (c = speed of light), ε_b = bulk permittivity and ε_o = vacuum permittivity. Because $\mathbf{P}(\mathbf{r})$ satisfies the Debye relation $\nabla.(\varepsilon_\mathrm{o}\mathbf{E} + \mathbf{P})(\mathbf{r}) = \rho(\mathbf{r})$, where $\rho(\mathbf{r})$ = charge density, equation (4) yields the following equation in $\mathbf{r}$ space[25]:

$$\nabla.[\int F^{-1}(\mathrm{K})(\mathbf{r} - \mathbf{r}')\mathbf{E}(\mathbf{r}')\mathrm{d}\mathbf{r}')] = \rho(\mathbf{r}), \tag{4}$$

with $K(\boldsymbol{\omega}) = \varepsilon_\mathrm{o} + K_\mathrm{p}(\boldsymbol{\omega})$. The convolution $\int F^{-1}(\mathrm{K})(\mathbf{r} - \mathbf{r}')\mathbf{E}(\mathbf{r}')\mathrm{d}\mathbf{r}'$ captures the correlation of the dipoles with the electrostatic field. Note that equation (4) is not the Poisson–Boltzmann equation, which requires a proportionality between the fields $\mathbf{E}$ and $\mathbf{P}$ under the *ad hoc* assumption $K(\boldsymbol{\omega}) \equiv$ constant.

Upon water confinement, the dielectric relaxation undergoes a frequency red-shift arising from the reduction in hydrogen-bond partnerships that translates to a reduction in dipole orientation possibilities. Thus, at position $\mathbf{r}$, the relaxation time is $\tau = \tau_\mathrm{b}\exp(B(g(\mathbf{r}))/k_BT)$, where the kinetic barrier $B(g(\mathbf{r})) = -k_BT\ln(g(\mathbf{r})/4)$ yields $\tau(\mathbf{r}) = \tau_\mathrm{b}(g(\mathbf{r})/4)^{-1}$. Thus, for charge distribution,

$$\rho(\mathbf{r}) = \Sigma_{m\,\in\,L} 4\pi q_m \delta(\mathbf{r} - \mathbf{r}_m), \tag{5}$$

with L = set of charges on the protein surface labelled by index m, the g-dependent polarization is obtained from equation (4) (Supplementary Information):

$$\mathbf{P}(\mathbf{r}) = \int F^{-1}(K_\mathrm{p})(\mathbf{r} - \mathbf{r}')\mathbf{E}(\mathbf{r}')\mathrm{d}\mathbf{r}'$$

$$= (2\pi)^{-3}\boldsymbol{\Sigma}_{\,m\,\in\,L}\int \mathrm{d}\mathbf{r}' F^{-1}(K_\mathrm{p})(\mathbf{r} - \mathbf{r}')\nabla_{\mathbf{r}'}\int \mathrm{d}\boldsymbol{\omega}\mathrm{e}^{-i\boldsymbol{\omega}.(\mathbf{r}' - \mathbf{r}_m)} 4\pi q_m/[|\omega|^2 K(\boldsymbol{\omega})]. \tag{6}$$

Spatially dependent coordination $g = g(\mathbf{r})$**.** The time-averaged scalar field $g = g(\mathbf{r})$ was obtained from classical trajectories generated by molecular dynamics. The computations started with the PDB structure of a free (uncomplexed) protein molecule embedded in a pre-equilibrated cell of explicitly represented water molecules and counterions[18,19]. The molecular-dynamics trajectories were generated by adopting an integration time step of 2 fs in an NPT ensemble with box size $10^3\,\mathrm{nm}^3$ and periodic boundary conditions[26]. The box size was calibrated so that the solvation shell extended at least 10 Å from the protein surface at all times. The long-range electrostatics were treated using the particle mesh Ewald summation method[27]. A Nosé–Hoover thermostat[28] was used to maintain the temperature at 300 K, and a Tip3P water model with the optimized potential for liquid simulations (OPLS) force field was adopted[18,19]. A barostat scheme was maintained through a dedicated routine with the pressure held constant at 1 atm. using a weak-coupling algorithm[29]. After equilibration for 300 ns, g values averaged over a time span of 100 ns were determined for each point in space.

PWIT as promoter of protein–protein associations. The PWIT computed using equations (2) and (6) is generated by interfacial hotspots of red-shifted dielectric relaxation ($g(\mathbf{r}) < 4$, $\tau(\mathbf{r}) > \tau_\mathrm{b}$). The most common spots involve hindered polar hydration generated by SABHBs (Fig. 1a). Taken collectively, the SABHBs contribute $73 \pm 5\%$ to the interfacial tension (Supplementary Information). The results are validated by showing that the inferred patches of interfacial tension promote protein associations, a conclusion supported by the tight correlation ($r^2 = 0.83$) between the total area of surface patches begetting PWIT (increasing the value of the integral in equation (2)) in free complex subunits, and the total protein–protein interfacial area of protein complexes (Supplementary Fig. 2a). The relevance of PWIT as a molecular determinant of protein–protein interactions is further validated by showing that inferred tension patches actually coincide with hotspots at complex interfaces experimentally identified by mutational scanning (Supplementary Fig. 2b, c).

Identification of SABHBs in soluble proteins. The extent of protection of a backbone hydrogen bond, ζ, was computed directly from PDB structural coordinates by determining the number of side-chain non-polar groups contained within a desolvation domain around the bond[20,22]. This domain was defined as two intersecting spheres of fixed radius (approximate thickness of three water layers) centred at the α-carbons of the residues paired by the hydrogen bond. In structures of soluble proteins, backbone hydrogen bonds are protected on average by $\zeta = 26.6 \pm 7.5$ non-polar groups for a desolvation sphere of radius 6 Å. SABHBs lie in the tails of the distribution: that is, their microenvironment contains 19 or fewer non-polar groups ($\zeta \leq 19$), so their ζ value is below the mean minus one standard deviation.

Sequence-based identification of SABHBs. SABHBs represent structural vulnerabilities that have been characterized as belonging to a twilight zone between order and native disorder. This characterization is justified by a strong correlation between intramolecular hydrogen-bond protection, ζ, and propensity for structural disorder (f_d) (Supplementary Fig. 5). The correlation reveals that the inability to exclude water intramolecularly from pre-formed hydrogen bonds is causative of the loss of structural integrity. The disorder propensity is accurately quantified by a sequence-based score generated by the program PONDR-VLXT[30], a predictor of native disorder that takes into account residue attributes such as hydrophilicity, aromaticity and their distribution within the window interrogated. The disorder score ($0 \leq f_\mathrm{d} \leq 1$) is assigned to each residue within a sliding window, representing the predicted propensity of the residue to be in a disordered region ($f_\mathrm{d} = 1$, certainty of disorder; $f_\mathrm{d} = 0$, certainty of order). Only 6% of 1,100 non-homologous PDB proteins gave false-positive predictions of disorder in sequence windows of 40 amino acids[22,30]. The strong correlation (Supplementary Fig. 5) between the disorder score of a residue and extent of protection of the hydrogen bond engaging the residue (if any) provides a sequence-based method of inference of SABHBs and supports the picture that such bonds belong to an order–disorder twilight zone[22]. Thus SABHBs can be safely inferred in regions where the disorder score lies in the range $0.35 \leq f_\mathrm{d} < 0.95$, which corresponds to a marginal BHB protection with $7 \leq \zeta \leq 19$ (Supplementary Fig. 5).

Evaluation of homology models. The homology models based on template PDB structures from orthologous proteins were evaluated, ranked and ultimately selected using ProSA[15], based on the minimization of $(Z_\mathrm{mod} - Z_\mathrm{temp})/Z_\mathrm{temp}$, where Z_mod and Z_temp are the Z scores of model and template. The Z score of a structure or template-based model is the energetic gap between the structure and an average over an ensemble of random conformations for the protein chain[15].

23. Schutz, C. N. & Warshel, A. What are the dielectric constants of proteins and how to validate electrostatic models? *Proteins Struct. Funct. Genet.* **44,** 400–417 (2001).
24. Scott, R., Boland, M., Rogale, K. & Fernández, A. Continuum equations for dielectric response to macromolecular assemblies at the nanoscale. *J. Phys. A* **37,** 9791–9803 (2004).
25. Fernández, A., Sosnick, T. R. & Colubri, A. Dynamics of hydrogen-bond desolvation in folding proteins. *J. Mol. Biol.* **321,** 659–675 (2002).
26. Lindahl, E., Hess, B. & Van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7,** 302–317 (2001).
27. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an *N* log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* **98,** 10089–10092 (1993).
28. Hoover, W. G. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A* **31,** 1695–1697 (1985).
29. Berendsen, H. J., Postma, J. P., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81,** 3684–3690 (1984).
30. Li, X., Romero, P., Rani, M., Dunker, A. K. & Obradovic, Z. Predicting protein disorder for N-, C-, and internal regions. *Genome Informat.* **10,** 30–40 (1999).

LETTER

doi:10.1038/nature10111

A nuclear-receptor-dependent phosphatidylcholine pathway with antidiabetic effects

Jae Man Lee[1], Yoon Kwang Lee[2,3], Jennifer L. Mamrosh[2], Scott A. Busby[4], Patrick R. Griffin[4], Manish C. Pathak[5], Eric A. Ortlund[5] & David D. Moore[1,2]

Nuclear hormone receptors regulate diverse metabolic pathways and the orphan nuclear receptor LRH-1 (also known as NR5A2) regulates bile acid biosynthesis[1,2]. Structural studies have identified phospholipids as potential LRH-1 ligands[3–5], but their functional relevance is unclear. Here we show that an unusual phosphatidylcholine species with two saturated 12 carbon fatty acid acyl side chains (dilauroyl phosphatidylcholine (DLPC)) is an LRH-1 agonist ligand *in vitro*. DLPC treatment induces bile acid biosynthetic enzymes in mouse liver, increases bile acid levels, and lowers hepatic triglycerides and serum glucose. DLPC treatment also decreases hepatic steatosis and improves glucose homeostasis in two mouse models of insulin resistance. Both the antidiabetic and lipotropic effects are lost in liver-specific *Lrh-1* knockouts. These findings identify an LRH-1 dependent phosphatidylcholine signalling pathway that regulates bile acid metabolism and glucose homeostasis.

Increased fat accumulation in the liver—steatosis—is tightly correlated with insulin resistance and type 2 diabetes[6]. Modestly raised bile acid levels decrease steatosis[7]. Loss of the nuclear receptor LRH-1 decreases bile acid levels[1,2], indicating that an LRH-1 agonist could increase them and improve fatty liver. In screens of a number of different phosphatidylcholine (PC) and other phospholipid species for effects on human LRH-1 transactivation, dilauroyl PC (DLPC; C12:0/C12:0) and diundecanoyl PC (DUPC; C11:0/C11:0) showed strong stimulation (Fig. 1a). Comparable responses were not observed with closely related PCs differing in acyl chain length by only a single methylene group, or with any other C12:0/C12:0 phospholipid species (Supplementary Fig. 1a–c).

DLPC and DUPC, but not the bile acid chenodeoxycholic acid (CDCA) or the more conventional phospholipid dipalmitoyl PC (DPPC; C16:0/C16:0), also activated the synthetic LRH-1 reporter in several other cell lines, including CV-1 and HEK293T cells (data not shown), and specifically increased basal LRH-1 transactivation of the native mouse SHP promoter[8] by approximately twofold in HeLa cells (Supplementary Fig. 2a). DLPC and DUPC also induced a similar response with the OCT4 promoter, which was dependent on both LRH-1 cotransfection and an intact LRH-1 response element[9] (Supplementary Fig. 2a). DLPC and DUPC responsiveness was not altered in mutant LRH-1 derivatives previously shown to inactivate responses to LRH-1 phosphorylation[10] or sumoylation[11], but was strongly decreased by mutations shown to block phospholipid binding[4] (Supplementary Fig. 2d).

Mouse and human LRH-1 showed essentially equivalent responses to DLPC and DUPC, and both DLPC and DUPC also activate the close LRH-1 relative SF-1 (also known as NR5A1; Supplementary Fig. 2). The LRH-1 responses were dose dependent (Supplementary Fig. 2c). Neither DUPC nor DLPC showed significant activation of any of a number of additional nuclear receptors outside of the NR5A subgroup (Supplementary Fig. 2b). In particular, DLPC and DUPC failed to activate PPARα, which was recently reported to be specifically bound and activated by 1-palmitoyl-2-oleoyl (C16:0/C18:1) PC[12], and C16:0/C18:1 PC failed to affect LRH-1 transactivation (Supplementary Fig. 1a). DLPC rapidly induced expression of the LRH-1 target CYP8B1 in the C3A derivative of HepG2 cells (Supplementary Fig. 3a). This response as well as CDCA repression of CYP8B1 expression and transactivation of a synthetic LRH-1 reporter plasmid was specifically compromised in cells transfected with LRH-1 short interfering RNA (siRNA; Supplementary Fig. 3b, c).

We used the mammalian two-hybrid assay and a simple GST pull-down approach to initially test the predicted function of DLPC and DUPC as LRH-1 agonist ligands. In the mammalian two-hybrid analysis, interaction of a VP16–human LRH-1 ligand-binding-domain fusion with a second fusion of the Gal4 DNA-binding domain to the nuclear receptor interaction domain of the coactivator SRC-3 (also known as NCOA3) was unaffected by vehicle, CDCA or DPPC, but was stimulated by either DUPC or DLPC (Supplementary Fig. 4a). *In vitro*, SRC-3 protein did not bind to GST alone but showed a significant basal interaction with a GST–LRH-1-ligand-binding-domain fusion protein, as expected[4]. DLPC and DUPC further increased binding of the coactivator by approximately 3 fold, but vehicle, CDCA, or any of a number of other PC species, including DPPC, had little or no effect (Supplementary Fig. 4b). DLPC also unexpectedly but specifically decreased binding of an SRC-2 peptide to the LRH-1 ligand-binding domain with a half-maximum inhibitory concentration (IC_{50}) of approximately 500 nM, but DPPC had no effect (Supplementary Fig. 4c), and DLPC did not affect rosiglitazone binding to PPARγ (Supplementary Fig. 4d).

As a stringent test of specific binding, the purified bacterially expressed human LRH-1 ligand-binding domain was incubated with DLPC or DPPC at molar ratios of 1:1 or 1:5 (protein:PC), or with buffer alone, and the protein was then repurified to eliminate unbound lipids. Specifically bound lipids were extracted and compared to DLPC or DPPC by electrospray ionization mass spectrometry. Phosphatidylethanolamine (PE) and phosphatidylglycerol (PG) species with 16–22 carbon acyl chain lengths occupy the ligand-binding pocket in the buffer-treated control, with the most abundant peak corresponding to 16:1/18:1 PG (Fig. 1b). DLPC completely replaced these *Escherichia coli* phospholipids, even at an added lipid to protein molar ratio of only 1:1, but DPPC showed no detectable displacement, even at a ratio of 1:5 (Fig. 1b). On the basis of these functional and *in vitro* biochemical results, as well as the extensive structural studies demonstrating phospholipid binding to NR5A receptors[3–5,13,14], we conclude that DLPC and DUPC act *in vitro* as LRH-1 agonists. The functional results indicate that they may also act directly as agonists *in vivo*, although it remains unclear how they might transit the cell membrane and cytosol and enter the nucleus.

PCs are normal dietary nutrients that are efficiently absorbed in the small intestine, and we used the simple route of oral gavage to deliver cholic acid (CA), DLPC, DUPC and DPPC to C5BL/6 mice. These treatments had no apparent toxic effects and did not alter normalized

[1]Program in Developmental Biology, Baylor College of Medicine, Houston, Texas 77030, USA. [2]Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. [3]Department of Integrative Medical Sciences, Northeastern Ohio Universities Colleges of Medicine and Pharmacy, Rootstown, Ohio 44272, USA. [4]The Scripps Research Molecular Screening Center, The Scripps Research Institute, Scripps Florida, Jupiter, Florida 33458, USA. [5]Department of Biochemistry, Emory University School of Medicine, Atlanta, Georgia 30322, USA.

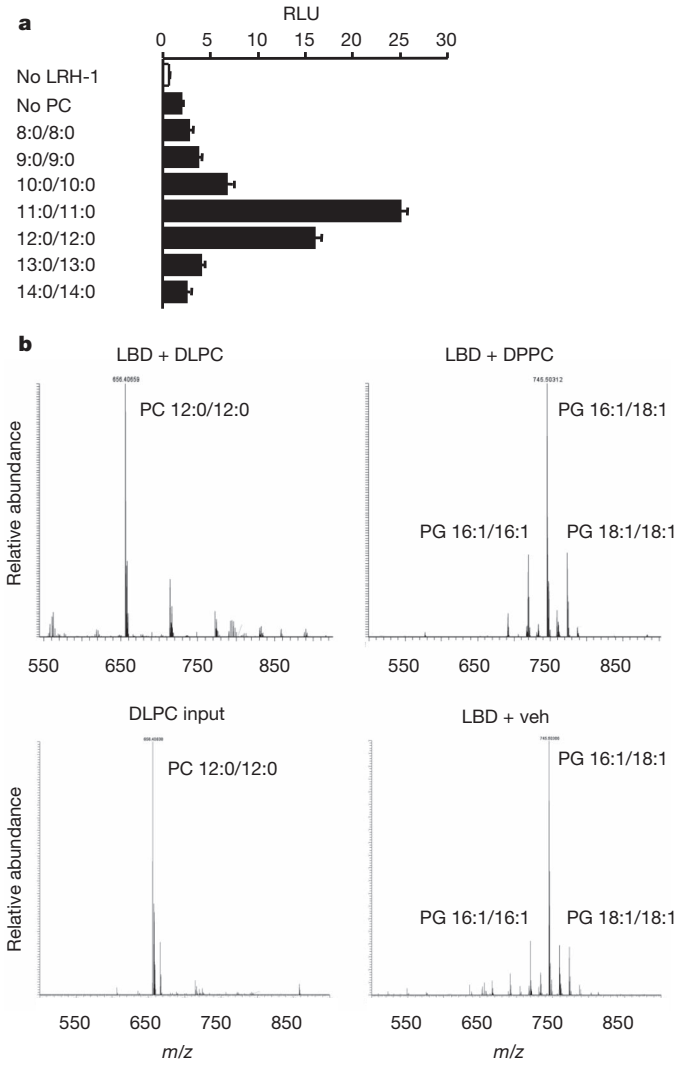


Figure 1 | DLPC activates and binds human LRH-1. a, HeLa cells were transfected with a human LRH-1 expression vector and a luciferase reporter and treated with 100 μM of indicated PCs. Error bars represent mean ± s.e.m. RLU, relative luciferase units. **b**, The human LRH-1 ligand-binding domain (LBD) was expressed and purified as described previously[5] and was incubated at molar ratios of 1:1 and 1:5 (human LRH-1 LBD:PC) with DLPC, DPPC or vehicle (veh) for two hours at 37 °C, and then repurified by size exclusion chromatography to remove unbound phospholipids. Bound lipids were analysed using electrospray mass injection mass spectrometry in the negative-ion mode. Results with DLPC (1:1), DPPC (1:5) and vehicle are shown, along with analysis of re-extracted DLPC; DLPC (1:5) and DPPC (1:1) incubations were very similar to those shown. The re-extracted DPPC peak is at 768.5, and is not detectable in any of the DPPC incubations.

liver weight (Supplementary Fig. 5a) or increase serum indicators of liver damage (Supplementary Fig. 5b). CA reduced expression of CYP7A1 and CYP8B1 and induced SHP, as expected, and DPPC was without significant effect (Fig. 2a). Both DLPC and DUPC significantly induced expression of CYP7A1, CYP8B1 and SR-B1, and repressed SHP (Fig. 2a). The substantial induction of CYP8B1, particularly by DLPC, is in accord with the opposite response in liver-specific *Lrh-1* knockouts[1,2], which otherwise show relatively limited alterations in gene expression or liver physiology. The decreased SHP expression is consistent with the induction of the bile acid biosynthetic enzymes, but was not expected based on the acute response of the isolated SHP promoter in HeLa cells (Supplementary Fig. 2a). Because SHP represses its own expression in the liver[8], it is possible that an initial inductive response is followed by an autoregulatory decrease. The induction of CYP7A1 and CYP8B1 is lost in liver-specific *Lrh-1* knockouts generated by infecting LRH-1 floxed (f/f) mice[2] with adenoviral Cre (Ad-Cre) expression vectors (Supplementary Fig. 6a, b).

These gene expression changes were associated with a modest but significant increase in the total bile acid pool and serum bile acid levels in DLPC- and DUPC-treated mice (Fig. 2b), consistent with the opposite effect in liver-specific *Lrh-1* knockouts[1,2]. These DLPC and DUPC effects were lost in Ad-Cre-mediated liver-specific knockouts (Supplementary Fig. 6c). Both CA- and DUPC/DLPC-treated mice showed significantly decreased serum non-esterized fatty acids (NEFAs) (Fig. 2b) and hepatic triglycerides (Fig. 2c), which were associated with decreased serum glucose in DUPC/DLPC-treated, but not CA-treated mice (Fig. 2b). Serum and hepatic cholesterol levels were unaffected (Fig. 2b, c). As anticipated based on the effective clearance of phospholipid-containing gut-derived chylomicrons by the liver, neither DLPC nor DUPC altered expression of SF-1 target genes in the adrenal gland (Supplementary Fig. 5c).

Prompted by the lipid and glucose effects in the normal mice, we focused on DLPC, a natural product, and treated insulin-resistant leptin-receptor-deficient *db/db* mice for 2 weeks by oral gavage, followed by a glucose tolerance test (GTT). An insulin tolerance test (ITT) was carried out after 1 week of additional treatment. Glucose homeostasis was improved in DLPC-treated mice, as shown by the GTT and ITT (Supplementary Fig. 7a), as well as lower fasting serum insulin levels (Supplementary Fig. 7e). DLPC treatment did not affect body weight or a number of other parameters, but decreased expression of the lipogenic transcription factor SREBP-1c (also known as SREBF1) and its downstream targets, and significantly lowered hepatic triglyceride levels in the *db/db* mice (Supplementary Figs 7b–f and 8).

To critically test the role of hepatic LRH-1 in these antidiabetic effects, wild-type *Lrh-1*$^{f/f}$ and liver-specific *Lrh-1*$^{-/-}$ mice generated using an albumin-Cre transgene (Supplementary Fig. 9a) were fed a high-fat diet to induce obesity and insulin resistance (diet-induced obesity (DIO)) for 15 weeks. Continuing on the diet, they were treated daily by oral gavage with vehicle or DLPC for 3 weeks, and glucose homeostasis was assessed by GTT and ITT. Loss of LRH-1 did not affect glucose homeostasis in the *Lrh-1*$^{-/-}$ DIO mice relative to the *Lrh-1*$^{f/f}$ DIO mice (Fig. 3a). As in the *db/db* mice, DLPC treatment substantially improved glucose homeostasis in the *Lrh-1*$^{f/f}$ DIO mice as indicated by GTT and ITT, and these responses were absent in the *Lrh-1*$^{-/-}$ DIO mice (Fig. 3a). The DLPC treated *Lrh-1*$^{f/f}$ DIO mice also had decreased fasting serum glucose and insulin levels, resulting in an 80% decrease in the homeostatic model assessment of insulin resistance (HOMA-IR) (Fig. 3b). Increased insulin sensitivity was confirmed using the hyperinsulinaemic–euglycaemic clamp, which showed both increased glucose disposal and markedly decreased hepatic glucose production in the DLPC-treated mice (Fig. 3c). Increased overall insulin sensitivity was also confirmed by increased insulin-dependent phosphorylation of the insulin receptor, IRS2 and AKT in the DLPC-treated *Lrh-1*$^{f/f}$ livers, but not *Lrh-1*$^{-/-}$ livers (Fig. 3d and Supplementary Fig. 9c).

Total body weight and food intake (Supplementary Fig. 9b), as well as weights of liver, reproductive fat pads, or brown fat did not differ between the *Lrh-1*$^{f/f}$ and *Lrh-1*$^{-/-}$ DIO mice. However, the livers of DLPC-treated *Lrh-1*$^{f/f}$ DIO mice were less pale and fatty, and decreased lipid deposition was confirmed both histologically and by direct measurement of hepatic triglyceride levels (Fig. 4a, b). NEFA levels were also decreased by DLPC in *Lrh-1*$^{f/f}$, but not *Lrh-1*$^{-/-}$ DIO livers and serum (Fig. 4b). Hepatic and serum bile acid levels were significantly increased by DLPC in the *Lrh-1*$^{f/f}$, but not the *Lrh-1*$^{-/-}$ DIO mice (Fig. 4b). DLPC significantly induced both CYP7A1 and CYP8B1 expression in the *Lrh-1*$^{f/f}$ mice, and this specific response was absent in the *Lrh-1*$^{-/-}$ mice (Supplementary Fig. 10a). The expression of additional bile-acid-related genes, including the biosynthetic CYP7B1 and CYP27A1 and the hepatic bile acid transporters BSEP (also known as ABCB11) and NTCP (also known as SLC10A1) was not significantly affected by DLPC treatment in *Lrh-1*$^{f/f}$ DIO mice (Supplementary Fig. 10a).

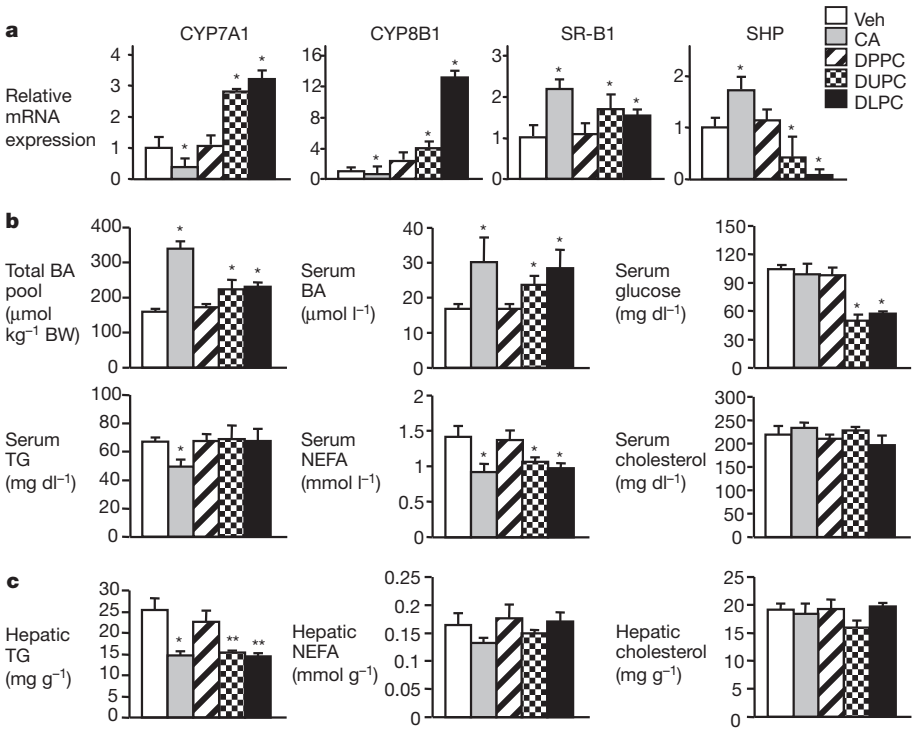


Figure 2 | DLPC and DUPC modulate expression of LRH-1 target genes in liver. a, Eight-week-old male C57BL/6 mice were challenged orally with vehicle (Veh), CA, DPPC, DUPC and DLPC for 3 days. Total liver RNA was isolated and prepared for the complementary DNA. Hepatic gene expression was determined using quantitative polymerase chain reaction (PCR). mRNA levels are relative to 36B4. **b**, Total bile acid (BA) pool and serum BA, glucose, triglyceride (TG), NEFA and cholesterol were measured in the same mice. **c**, Hepatic TG, NEFA and cholesterol were measured in the same mice. Error bars represent mean ± s.e.m. $*P < 0.05$, $**P < 0.01$ versus vehicle; $n = 5$ mice per group.

In accord with *db/db* results (Supplementary Fig. 8), there was little or no effect on hepatic expression of a number of glucose homeostasis and fatty acid oxidation genes (Supplementary Fig. 10d, e). However, DLPC markedly decreased expression of genes associated with *de novo* lipogenesis (Fig. 4c), including the lipogenic transcription factor SREBP-1c and its key downstream targets ACC-2, SCD-1 and FASN in *Lrh-1*$^{f/f}$ DIO mice (Fig. 4c). The beneficial effects of DLPC on glucose homeostasis and fatty liver in *Lrh-1*$^{f/f}$ mice fed a high-fat diet and infected with a control Ad–GFP vector were also lost in mice in which the *Lrh-1*$^{f/f}$ allele was deleted by Ad-Cre expression (Supplementary Fig. 11). Overall, we conclude that LRH-1 is required for the antidiabetic effects of DLPC. However, it remains to be determined whether its effects are a consequence of being a direct ligand for LRH-1, and it remains possible that DLPC activates an alternative signalling cascade or induces biosynthesis of an endogenous LRH-1 ligand.

Here we have identified DLPC as a specific agonist ligand for LRH-1 *in vitro*. Further studies will be needed to address the intriguing questions of whether phospholipid transfer proteins[15] facilitate its transport to the large and dynamic intranuclear pool of phosphatidylcholine[16], and whether DLPC is an endogenous LRH-1 agonist. The ligand responsiveness of LRH-1 is consistent with the identification of synthetic agonists that activate both LRH-1 and SF-1 (ref. 17). When expressed in an adrenal cell line, SF-1 is bound by a relatively low molecular weight form of phosphatidic acid with two saturated 14 carbon acyl chains, which acts as an agonist for SF-1 but not LRH-1 (ref. 18). Earlier results identified the sphingolipids sphingosine and lyso-sphingomyelin as potential endogenous antagonists of SF-1 transactivation[19]. DLPC does not activate other nuclear receptors, including PPARα or PPARγ, which have previously been reported to be activated by more conventional longer chain phospholipid species[12,20,21]. In the opposite direction, LRH-1 is not activated by conventional PC species, including the C16:0/C18:1 PC reported to specifically bind and activate PPARα in the liver[12]. Phospholipids are emerging as a structurally diverse class of highly specific nuclear receptor ligands.

The beneficial effect of DLPC on steatosis is associated with significantly decreased expression of the transcription factor SREBP-1c and its downstream lipogenic targets. At least two complementary mechanisms could contribute to this decrease. As SREBP-1c autoregulates its own expression[22], the reported functional antagonism of SREBP-1c transactivation by LRH-1 (ref. 23) could directly inhibit SREBP-1c promoter activity. As SREBP-1c expression is induced by insulin[24], the DLPC-dependent decrease in serum insulin should also decrease SREBP-1c messenger RNA. The combination of these two mechanisms could set up a positive regulatory loop in which the initial LRH-1-dependent repression of SREBP-1c expression would decrease steatosis and increase insulin sensitivity, resulting in a decrease in serum insulin. This decrease would then reinforce the decline in SREBP-1c expression and activity, further ameliorating fatty liver and thereby continuing a beneficial cycle (Supplementary Fig. 12). This essentially reverses the lipogenic vicious cycle to insulin resistance proposed previously by McGarry[25], and supported by more recent results with SREBP-1c[26].

These beneficial effects are probably complemented by an increase in fatty acid β-oxidation due to the decrease in acetyl-CoA carboxylase-2 (ACC-2) and its product, malonyl-CoA, which allosterically inhibits CPT-1a enzymatic activity and mitochondrial fatty acid uptake[27]. Decreasing ACC-2 activity in response to either specific antisense oligonucleotides[28] or activation of the nuclear receptor CAR[29] increases β-oxidation and has beneficial effects on both steatosis and insulin resistance. Because SCD-1 ablation also protects against hepatic steatosis by decreasing lipogenesis and increasing β-oxidation[30], reduced SCD-1 expression may also increase β-oxidation in response to LRH-1 activation.

We conclude that the identification of DLPC as a useful tool for analysis of LRH-1 function has uncovered an unexpected, LRH-1-dependent PC signalling pathway that can improve fatty acid and

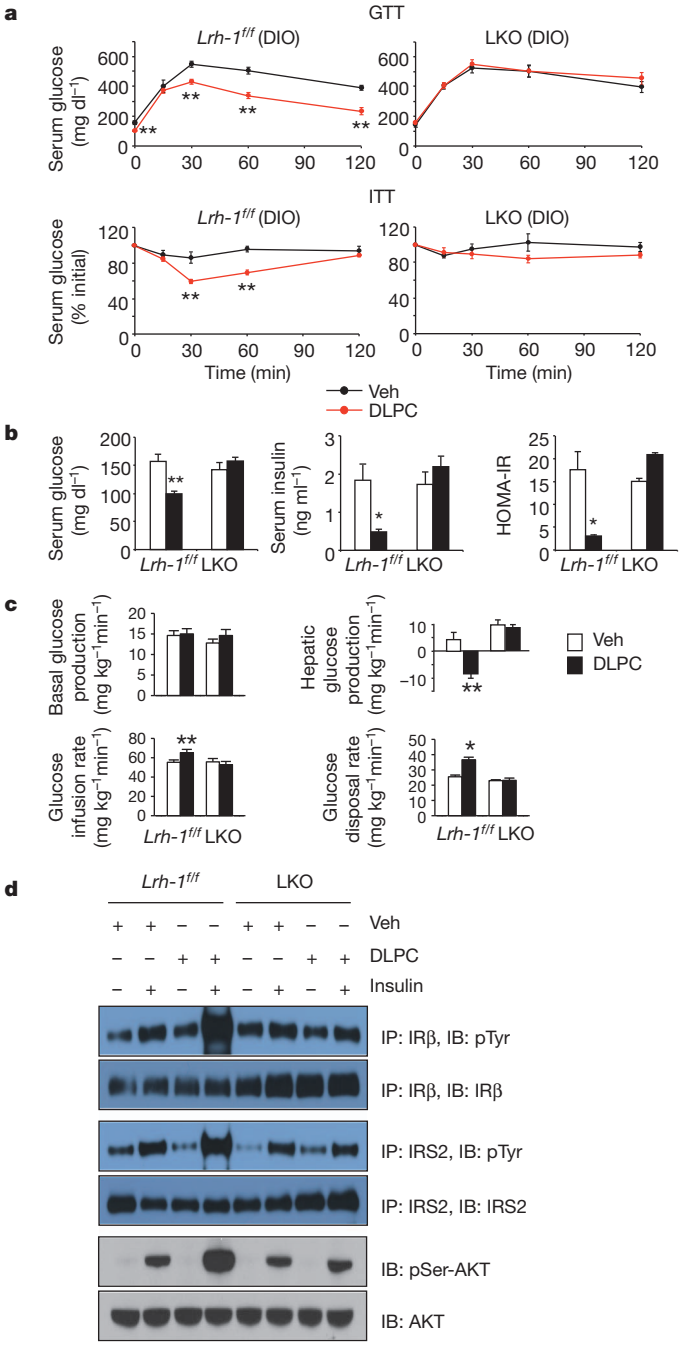


Figure 3 | DLPC improves glucose homeostasis in mouse models of insulin resistance. **a**, Glucose and insulin tolerance were assessed in *Lrh-1*$^{f/f}$ and *Lrh-1*$^{-/-}$ knockout (LKO) DIO mice 2–3 weeks after vehicle or DLPC treatment. **b**, Fasting serum glucose and insulin levels were measured in the same mice shown in **a**. HOMA-IR was calculated from fasting serum glucose and insulin levels. **c**, The high dose (10 mU kg^{-1} min^{-1}) hyperinsulinaemic–euglycaemic clamp (insulin dose of 10 mU kg^{-1} min^{-1}) was used to assess glucose homeostasis in *Lrh-1*$^{f/f}$ DIO mice after 3 weeks of vehicle or DLPC treatment. **d**, Hepatic insulin signalling was examined in *Lrh-1*$^{f/f}$ and LKO DIO mice 2 weeks after vehicle or DLPC treatment. Liver tissue homogenates from 3 mice per group were pooled and immunoprecipitation (IP) and immunoblotting (IB) were as indicated. Results are representative of three independent experiments. Error bars represent mean ± s.e.m. $*P < 0.05$, $**P < 0.01$ versus *Lrh-1*$^{f/f}$ DIO mice treated with vehicle; $n = 4$ mice per group.

glucose homeostasis. These studies indicate that DLPC is a promising therapeutic agent for the treatment of metabolic disorders, and we have initiated a human clinical trial to explore potential beneficial effects in prediabetic patients.

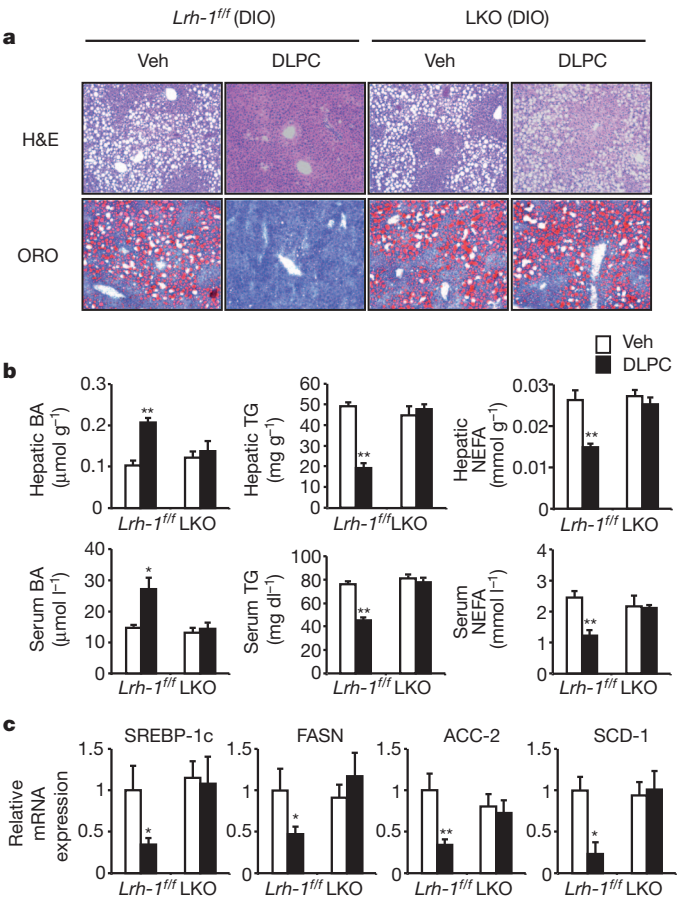


Figure 4 | DLPC reduces liver fat accumulation by suppressing lipogenesis. **a**, Liver sections from *Lrh-1*$^{f/f}$ and LKO DIO mice treated for 3 weeks with vehicle or DLPC were stained with haematoxylin and eosin (H&E) for general morphology or Oil Red O (ORO) for lipid accumulation. Original magnification, ×10. **b**, Hepatic and serum BA, TG and NEFA levels were measured in the same mice described in **a**. **c**, Lipogenic gene expression in the liver was determined using qPCR. mRNA levels are relative to 36B4. Error bars represent mean ± s.e.m. $*P < 0.05$, $**P < 0.01$ versus *Lrh-1*$^{f/f}$ DIO mice treated with vehicle; $n = 4$ mice per group.

METHODS SUMMARY

For transient transfection assays with HeLa, Cos-1 or C3A/HepG2 cells, candidate phospholipids dissolved in ethanol were added to cells for 24 h in medium containing 10% charcoal-treated FBS. Luciferase expression was assayed and normalized using β-galactosidase expression for transfection efficiency. Transfections were done in triplicate. For binding studies, the human LRH-1 ligand-binding domain, residues 291–541, was expressed as a maltose-binding protein fusion protein, cleaved and purified. It was incubated overnight with or without DLPC and specifically bound lipids were extracted with chloroform/methanol and analysed using electrospray mass injection mass spectrometry in the negative-ion mode to detect and identify phospholipids. LanthaScreen binding studies (Invitrogen) used full-length human LRH-1 and PPARγ. For short-term animal studies, C57BL/6 mice were orally gavaged with CA, DPPC, DUPC, or DLPC delivered in a standard vehicle every 12 h for a total of five treatments. Mice were killed 4 h after the final treatment on the morning of day 3. *Lrh-1* liver-specific knockout mice were generated as previously described[2]. For diabetes experiments, *db/db* mice were treated with vehicle or DLPC for 3 weeks. The GTT was performed 2 weeks after treatment. After an additional 1 week treatment, the ITT was performed. Eight-to-ten-week-old male control *Lrh-1*$^{f/f}$ or *Lrh-1*$^{-/-}$ mice were placed on a high-fat diet (45% kcal fat) for 15 weeks. The diet was maintained and mice were treated with vehicle or DLPC by oral gavage. GTT was performed in 18 h fasted mice after 2-week treatments. One week later, ITT was performed in *ad libitum* fed mice. Hyperinsulinaemic clamp (insulin dose of 10 mU kg^{-1} min^{-1}) was performed in the Diabetes and Endocrinology Research Center at the Baylor College of Medicine. All animal experiments were performed according to

procedures approved by the Baylor College of Medicine's Institutional Animal Care and Use Committee.

Statistics. Numbers of mice for each group used in experiments are indicated in the figure legends. Statistical analyses were performed with the two-tailed Student's *t*-test, and error bars represent means ± s.e.m. P value < 0.05 was considered statistically significant.

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

Received 19 March 2010; accepted 14 March 2011.

Published online 25 May; corrected online 29 May 2011 (see full-text HTML version for details).

1. Mataki, C. *et al.* Compromised intestinal lipid absorption in mice with a liver-specific deficiency of the liver receptor homolog 1. *Mol. Cell. Biol.* **27,** 8330–8339 (2007).
2. Lee, Y. K. *et al.* Liver receptor homolog-1 regulates bile acid homeostasis but is not essential for feedback regulation of bile acid synthesis. *Mol. Endocrinol.* **22,** 1345–1356 (2008).
3. Krylova, I. N. *et al.* Structural analyses reveal phosphatidyl inositols as ligands for the NR5 orphan receptors SF-1 and LRH-1. *Cell* **120,** 343–355 (2005).
4. Ortlund, E. A. *et al.* Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. *Nature Struct. Mol. Biol.* **12,** 357–363 (2005).
5. Wang, W. *et al.* The crystal structures of human steroidogenic factor-1 and liver receptor homologue-1. *Proc. Natl Acad. Sci. USA* **102,** 7505–7510 (2005).
6. Cusi, K. Nonalcoholic fatty liver disease in type 2 diabetes mellitus. *Curr. Opin. Endocrinol. Diabetes Obes.* **16,** 141–149 (2009).
7. Watanabe, M. *et al.* Bile acids lower triglyceride levels via a pathway involving FXR, SHP, and SREBP-1c. *J. Clin. Invest.* **113,** 1408–1418 (2004).
8. Lee, Y. K., Parker, K. L., Choi, H. S. & Moore, D. D. Activation of the promoter of the orphan receptor SHP by orphan receptors that bind DNA as monomers. *J. Biol. Chem.* **274,** 20869–20873 (1999).
9. Gu, P. *et al.* Orphan nuclear receptor LRH-1 is required to maintain Oct4 expression at the epiblast stage of embryonic development. *Mol. Cell. Biol.* **25,** 3492–3505 (2005).
10. Lee, Y. K., Choi, Y. H., Chua, S., Park, Y. J. & Moore, D. D. Phosphorylation of the hinge domain of the nuclear hormone receptor LRH-1 stimulates transactivation. *J. Biol. Chem.* **281,** 7850–7855 (2006).
11. Chalkiadaki, A. & Talianidis, I. SUMO-dependent compartmentalization in promyelocytic leukemia protein nuclear bodies prevents the access of LRH-1 to chromatin. *Mol. Cell. Biol.* **25,** 5095–5105 (2005).
12. Chakravarthy, M. F. *et al.* Identification of a physiologically relevant endogenous ligand for PPARα in liver. *Cell* **138,** 476–488 (2009).
13. Li, Y. *et al.* Crystallographic identification and functional characterization of phospholipids as ligands for the orphan nuclear receptor steroidogenic factor-1. *Mol. Cell* **17,** 491–502 (2005).
14. Sablin, E. P. *et al.* Structure of SF-1 bound by different phospholipids: evidence for regulatory ligands. *Mol. Endocrinol.* **23,** 25–34 (2009).
15. Kang, H. W., Wei, J. & Cohen, D. E. PC-TP/StARD2: Of membranes and metabolism. *Trends Endocrinol. Metab.* **21,** 449–456 (2010).
16. Hunt, A. N. Dynamic lipidomics of the nucleus. *J. Cell. Biochem.* **97,** 244–251 (2006).
17. Whitby, R. J. *et al.* Identification of small molecule agonists of the orphan nuclear receptors liver receptor homolog-1 and steroidogenic factor-1. *J. Med. Chem.* **49,** 6652–6655 (2006).
18. Li, D. *et al.* Cyclic AMP-stimulated interaction between steroidogenic factor-1 and diacylglycerol kinase θ facilitates induction of CYP17. *Mol. Cell Biol.* **27,** 6669–6685 (2007).
19. Urs, A. N., Dammer, E. & Sewer, M. B. Sphingosine regulates the transcription of CYP17 by binding to steroidogenic factor-1. *Endocrinology* **147,** 5249–5258 (2006).
20. Lee, H. *et al.* Role for peroxisome proliferator-activated receptor α in oxidized phospholipid-induced synthesis of monocyte chemotactic protein-1 and interleukin-8 by endothelial cells. *Circ. Res.* **87,** 516–521 (2000).
21. McIntyre, T. M. *et al.* Identification of an intracellular receptor for lysophosphatidic acid (LPA): LPA is a transcellular PPARγ agonist. *Proc. Natl Acad. Sci. USA* **100,** 131–136 (2003).
22. Amemiya-Kudo, M. *et al.* Promoter analysis of the mouse sterol regulatory element-binding protein-1c gene. *J. Biol. Chem.* **275,** 31078–31085 (2000).
23. Kanayama, T. *et al.* Interaction between sterol regulatory element-binding proteins and liver receptor homolog-1 reciprocally suppresses their transcriptional activities. *J. Biol. Chem.* **282,** 10290–10298 (2007).
24. Shimomura, I. *et al.* Insulin selectively increases SREBP-1c mRNA in the livers of rats with streptozotocin-induced diabetes. *Proc. Natl Acad. Sci. USA* **96,** 13656–13661 (1999).
25. McGarry, J. D. What if Minkowski had been ageusic? An alternative angle on diabetes. *Science* **258,** 766–770 (1992).
26. Li, S., Brown, M. S. & Goldstein, J. L. Bifurcation of insulin signaling pathway in rat liver: mTORC1 required for stimulation of lipogenesis, but not inhibition of gluconeogenesis. *Proc. Natl Acad. Sci. USA* **107,** 3441–3446 (2010).
27. Kim, K. H. Regulation of mammalian acetyl-coenzyme A carboxylase. *Annu. Rev. Nutr.* **17,** 77–99 (1997).
28. Savage, D. B. *et al.* Reversal of diet-induced hepatic steatosis and hepatic insulin resistance by antisense oligonucleotide inhibitors of acetyl-CoA carboxylases 1 and 2. *J. Clin. Invest.* **116,** 817–824 (2006).
29. Dong, B. *et al.* Activation of nuclear receptor CAR ameliorates diabetes and fatty liver disease. *Proc. Natl Acad. Sci. USA* **106,** 18831–18836 (2009).
30. Dobrzyn, P. *et al.* Stearoyl-CoA desaturase 1 deficiency increases fatty acid oxidation by activating AMP-activated protein kinase in liver. *Proc. Natl Acad. Sci. USA* **101,** 6409–6414 (2004).

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

Acknowledgements We thank S. A. Kliewer and D. J. Mangelsdorf (UT Southwestern Medical Center) for the gift of *Lrh-1*$^{f/f}$ mice, A. J. Cooney for the OCT4 promoter constructs, C. Mills and D. Kuruvilla for experimental assistance, the Baylor College of Medicine Diabetes Endocrine Research Center (supported by NIH DK-079638 and USDA ARS 6250-52000-055) and the services of the Mouse Metabolism Core for hyperinsulinaemic clamp studies, and the current and previous members of the D.D.M. laboratory for discussions and technical support. Supported by NIH R01 DK068804, the Alkek Foundation and the Robert R. P. Doherty Jr—Welch Chair in Science to D.D.M., and NIH R01 CA134873 to P.R.G.

Author Contributions J.M.L. designed and executed the experiments, interpreted data and co-wrote the manuscript. Y.K.L. and J.L.M. helped with experiments. S.A.B. and P.R.G. performed the fluorescence binding experiments, and M.C.P. and E.A.O. performed the mass spectrometry experiment. D.D.M. supervised the design and interpretation of the experiments and co-wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.D.M. (moore@bcm.edu).

METHODS

Materials. Phospholipids were purchased from Avanti Polar Lipids; fatty acids, CA and CDCA from Sigma-Aldrich; cell culture media and supplements from Invitrogen; insulin from Eli Lilly and Co.; human LRH-1 antibody from R&D systems, antibody against IRβ, IRS2, pSer-AKT and AKT was from Cell Signaling Technology, anti-phosphtyrosine antibody from Millipore; Ad5-CMV–GFP or Cre virus was prepared by the Vector Development Laboratory at the Baylor College of Medicine.

Cell culture and transient transactivation assays. HeLa, Cos-1 and C3A/HepG2 cells were cultured in DMEM containing 10% FBS and 1% penicillin/streptomycin antibiotics. 70–80% confluent cells were replated into a 24-well plate with a 1:5 ratio 24 h before transfection. Cell culture media were changed 1 h before transfection (calcium phosphate method). At 16 h after transfection, candidate phospholipids dissolved in ethanol were added to cells for 24 h in medium containing 10% charcoal treated FBS. Luciferase expression was assayed and normalized using β-galactosidase expression for transfection efficiency. Transfections were done in triplicate. Plasmids used were pcDNA3 for empty vector (100 ng per well), LRH-1/SF-1 luciferase reporter (200 ng per well), actin-β-galactosidase for internal control (150 ng per well), human LRH-1 expression plasmid (100 ng per well), mutant human LRH-1 expression plasmids (100 ng per well, phosphorylation mutant: S238, 243A; sumoylation mutant: K270R; ligand-binding mutant: F342W, I426W), Oct4-PP luciferase reporter (200 ng per well), and Oct4-PP$_{mut}$ luciferase (200 ng per well). Expression plasmids for receptors (100 ng per well) and their cognate luciferase reporters (200 ng per well) used were: human T_3Rβ, TK-28T-Luc; human RXRα, TK-CRBPII-Luc; human RARβ, TK-DR5-Luc; mouse PPARα, mouse PPARδ, mouse PPARγ, TK-PPRE × 3-Luc; mouse FXRα, human FXRα, MMTV-TK-ECRE × 5-Luc; human LXRα, TK-LXRE × 3-Luc; human ERα, TK-ERE-Luc; mouse CAR, human CAR, human PXR, TK-DR4-Luc; mouse SF-1, mouse LRH-1, human LRH-1, LRH-1/SF-1 Luc. For mammalian two-hybrid assays, replated HeLa cells in a 24-well plate were transfected with VP-16 (50 ng per well), VP16-human LRH-1 ligand-binding domain (50 ng per well), Gal4-SRC-3 RID (100 ng per well), G5-TK-Luc (200 ng per well), and actin-β-galactosidase plasmids (150 ng per well). For siRNA experiments, C3A/HepG2 cells were maintained with MEM containing 10% FBS, 1 mM sodium pyruvate, 0.1 mM nonessential amino acids, and 1.5 g l^{-1} sodium bicarbonate. C3A/HepG2 cells were replated into either a 24-well plate for luciferase assay or a 6-well plate for target gene expression and then transfected with human LRH-1 (Dharmacon, ON-TARGETplus SMARTpool, L-003430-00; J-0003430-06: AUAUGAAUAGCCCA UUAUGUU, J-0003430-07: UAGCUGUCCAAAUUUCUCUUU, J-0003430-08: AGGAUUAAGAGCUCACUCCUU, J-0003430-09: UCACCUGAGACAUGGC UUCUU) or control siRNA pool (Dharmacon, si*CONTROL* non-targeting siRNA pool, D-001206-13-05; UAGCGACUAAACACAUCAA, UAAGGCUAU GAAGAGAUAC, AUGUAUUGGCCUGUAUUAG, AUGAACGUGAAUUGC UCAA) using FuGENE 6 (Roche); For the luciferase assay, C3A/HepG2 cells were transfected with LRH-1/SF-1 luciferase and actin-β-galactosidase along with the siRNA pool. Twenty-four hours later, ligands were added, and cells were harvested 24 h later. To check for knockdown, cell extracts were analysed by immunoblot using an antibody against human LRH-1.

***In vitro* binding assays.** GST pull-down was used to examine the interaction between human LRH-1 ligand binding domain and SRC-3 *in vitro* as described[4]. GST alone and GST human LRH-1 ligand-binding domain (amino acid residues, 185–541) were expressed in DH5α strain *E. coli* with 0.5 mM IPTG for 4 h and then purified with glutathione-sepharose beads (GE Healthcare). GST proteins were incubated overnight at 4 °C in 50 mM Tris-HCL (pH 7.6), 0.2% Tween-20, 100 mg ml^{-1} BSA and 300 mM NaCl with various phospholipids. Full-length [^{35}S]methionine-labelled SRC-3 proteins (2 μl) were added to each reaction and incubated for 2 h at 4 °C. Unbound and nonspecific proteins were removed by washing five times with the same buffer. Specifically bound proteins were eluted by treatment with SDS sample buffer, subjected to SDS–PAGE, and visualized by autoradiography. The amount of specifically bound SRC-3 proteins was determined by densitometry (Personal Densitometer SI; Molecular Dynamics).

Lanthascreen assays were as described by the manufacturer (Invitrogen) using full-length human LRH-1 and a fluorescein-tagged SRC-2 coactivator peptide, and PPARγ and Fluoromone[31].

For mass spectrometry, the human LRH-1 ligand-binding domain (LBD), residues 291–541, was expressed as a maltose-binding protein fusion protein, cleaved and purified as described previously[4]. The pure protein was stored in a final buffer containing 150 mM NaCl, 20 mM HEPES and 5% glycerol. For binding studies, DLPC or DPPC dissolved in ethanol was evaporated in a clean glass cuvette at 50 °C under a stream nitrogen gas. Two millilitres of buffer containing 150 mM NaCl, 20 mM HEPES (pH = 7.5) and 5% glycerol was added to the cuvette containing dried DLPC or DPPC and was sonicated until the solution was optically clear. Human LRH-1 LBD was then added to the DLPC or DPPC vesicles at a ratio of 1:1 or 1:5 (human LRH-1 LBD:PC). The mixture was incubated for one hour at 37 °C followed by 24 h at 11 °C. The human-LRH-1–lipid complex was then purified by size exclusion chromatography to remove unbound phospholipids. Protein purity was assed by SDS–PAGE. Bound lipids were analysed using electrospray mass injection mass spectrometry (ESI-MS) in the negative-ion mode to detect and identify phospholipids. Approximately 6 mg of human LRH-1 LBD or the human LRH-1-LBD–lipid complexes were extracted with a 2:1 chloroform/methanol solution, diluted in 200 ml chloromethylene and analysed by negative-ion ESI-MS on a Thermo LTQ FTMS using direct injection analysis with electrospray ionization. The high-resolution analyses were performed in the FTMS at a resolution of 100,000 at 400 *m/z*. The MS/MS experiments were done in the ion trap portion of the instrument with a mass selection of 3 AMU and a normalized collision energy of 30 V. The major phospholipid species were identified by accurate mass measurements and MS/MS via collisionally induced dissociation (CID), which yields product ions characteristic of the head groups and attached fatty acids.

Animal studies. C57BL/6 mice and *db/db* mice were purchased from Harlan laboratories. Eight-week-old male C57BL/6 mice were orally gavaged with CA, DPPC, DUPC, or DLPC at a dose of 100 mg kg^{-1} body weight, delivered in a standard vehicle for delivery of hydrophobic compounds (4:1 of PEG-400 and Tween-80) every 12 h for a total of five treatments. Mice were killed 4 h after the final treatment on the morning of day 3. Harvested tissues were immediately frozen in liquid nitrogen for molecular studies. Twelve-week-old male *db/db* mice were used for diabetes studies. *db/db* mice were given compounds (vehicle or DLPC, $n = 5$ mice per group) at the dose of 100 mg kg^{-1} day^{-1}. After 2 weeks of treatments GTT (1.5 g kg^{-1} intraperitoneal injection) was performed in 18 h fasted mice. Treatments were continued for an additional week, and ITT (2 U kg^{-1} intraperitoneal injection) was performed in *ad libitum* fed mice. Serum insulin levels were determined using Rat/Mouse Insulin ELISA Kit from Linco Research. *Lrh-1$^{f/f}$* mice[2] were maintained on mixed C57BL6/129 backgrounds and were given by the Kliewer/Mangelsdorf laboratory. In brief, 4-month-old male *Lrh-1$^{f/f}$* littermates were tail vein injected with either Ad5-CMV–GFP (3×10^9 p.f.u.) or Ad5-CMV-Cre (3×10^9 p.f.u.). For acute experiments, these mice were orally gavaged daily with each compound starting 2 weeks after adenovirus injection as described in C57BL/6 mice above. Liver-specific *Lrh-1* ablation (LKO) was also achieved by crossing *Lrh-1$^{f/f}$* mice with albumin-Cre transgenic mice obtained from the O'Malley laboratory at the Baylor College of Medicine. To confirm tissue-specific deletion of exon 5 of *Lrh-1*, genomic DNA was extracted from tail, liver and intestine, and PCR analysis was performed as shown previously[2]. For diabetes experiments, 8–10-week-old male control *Lrh-1$^{f/f}$* or LKO mice were placed on a high-fat diet (Research diets; 45% kcal fat) for 15 weeks. The diet was maintained and mice were treated with vehicle or DLPC (dose of 100 mg kg^{-1} day^{-1}) by oral gavage. GTT (2 g kg^{-1} intraperitoneal injection) was performed in 18 h fasted mice after 2-week treatments. 1 week later, ITT (1 U kg^{-1}, intraperitoneal injection) was performed in *ad libitum* fed mice. Glucose levels were analysed using a glucometer (LifeScan). Insulin resistance (HOMA-IR) was calculated as following: fasting glucose (mg dl^{-1}) × fasting insulin (μU ml^{-1}) / 405. Hyperinsulinaemic clamp (insulin dose of 10 mU kg^{-1} min^{-1}) was performed and calculated as described in our previous publication[32]. Ad–GFP- or Ad-Cre-infected *Lrh-1$^{f/f}$* mice fed the high-fat diet for 15 weeks were used for the diabetes study as shown above. Mice were housed in a temperature-controlled room in pathogen-free facilities on a 12 h light/dark cycle (07:00 on, 19:00 off) and had free access to water and standard chow diet. All animals received humane care according to the criteria outlined in the "Guide for the Care and Use of Laboratory Animals" prepared by the National Academy of Sciences and published by the National Institutes of Health.

RNA isolation and mRNA quantification. Total RNA was isolated from C3A/HepG2 cells or snap-frozen liver tissues using Trizol Reagent (Invitrogen) and prepared for the cDNA with QuantiTect reverse transcriptase (Qiagen). Hepatic gene expression ($n = 4$–5) was determined by qPCR using FastStart SYBR Green master (ROX) (Roche). mRNA levels were normalized by the 36B4 gene. Primer information can be provided upon request.

Serum and tissue lipid analysis. Blood was collected from the orbital plexus and transferred into gel/clot activator tubes (Terumo). Samples were centrifuged at 6,000*g* for 5 min to separate serum. To extract bile acids from liver or intestine, each tissue was weighed and homogenized in 75% ethanol. The homogenate was incubated at 50 °C for 2 h to extract bile acids and centrifuged at 6,000*g* for 10 min at 4 °C. The bile acid content of the supernatant was determined and normalized with tissue weight used. To extract other lipids, snap-frozen liver fragments were weighed and homogenized in nine volumes of PBS. Two-hundred microlitres of the homogenate was transferred into 1,200 μl of chloroform:methanol (2:1; v/v) mixture and mixed vigorously for 30 s. One-hundred microlitres of PBS was then added, and the resulting suspension was mixed vigorously for 15 s then centrifuged at 4,200*g* for 10 min at 4 °C. Two-hundred microlitres of the chloroform:methanol layer (bottom phase) was transferred into a tube and evaporated for dryness. The

dried lipid residue was resuspended in 100 µl of 1% Triton X100 in absolute ethanol for 4 h with constant rotation. Bile acids levels were measured using the bile acid L3K assay kit (Diagnostic Chemicals). Cholesterol and triglyceride levels were determined by assay kits from Thermo DMA. Free fatty acids were assayed using a kit obtained from WAKO Chemicals.

***In vivo* insulin stimulation and analysis of insulin signalling.** Mice were fasted overnight and injected intraperitoneally with insulin (1 U kg^{-1}) or PBS. Five minutes after injection, tissues were removed, frozen in liquid nitrogen, and stored at −80 °C until use. For protein extraction, tissues were homogenized in a cold lysis buffer (50 mM Tris-HCl, pH 7.4; 1% NP-40; 0.5% sodium deoxycholate; 150 mM NaCl; 1 mM EDTA) containing protease and phosphatase inhibitor cocktails (Roche). After homogenization, the tissue lysates were allowed to solubilize for 1 h at 4 °C with rotation, and then were centrifuged at 19,700*g* for 30 min at 4 °C. The supernatants were used for immunoprecipitations and immunoblot analyses of insulin signalling proteins.

Histology. Liver was removed and pieces were fixed in 10% (v/v) neutralized formalin solution (J. T. Baker), embedded in paraffin, sectioned at 5 µm, and stained with haematoxylin and eosin. For Oil Red O staining, frozen liver tissues embedded in O.C.T. compound (Tissue-Tek) were used. Histological analysis performed in the Comparative Pathology Laboratory at Baylor College of Medicine.

Statistics. Numbers of mice for each group used in experiments are indicated in the figure legends. Statistical analyses were performed with the two-tailed Student's *t*-test, and error bars represent means ± s.e.m. P value < 0.05 was considered statistically significant.

31. Vidović, D., Busby, S. A., Griffin, P. R. & Schurer, S. C. A combined ligand- and structure-based virtual screening protocol identifies submicromolar PPARγ partial agonists. *ChemMedChem* **6,** 94–103 (2011).
32. Ma, K., Saha, P. K., Chan, L. & Moore, D. D. Farnesoid X receptor is essential for normal glucose homeostasis. *J. Clin. Invest.* **116,** 1102–1109 (2006).

LETTER

doi:10.1038/nature10085

Regulation of angiogenesis by a non-canonical Wnt–Flt1 pathway in myeloid cells

James A. Stefater III[1,2], Ian Lewkowich[3], Sujata Rao[1,2], Giovanni Mariggi[4], April C. Carpenter[1,2], Adam R. Burr[5], Jieqing Fan[1,2], Rieko Ajima[6], Jeffery D. Molkentin[5,7], Bart O. Williams[8], Marsha Wills-Karp[3], Jeffrey W. Pollard[9], Terry Yamaguchi[6], Napoleone Ferrara[10], Holger Gerhardt[4,11] & Richard A. Lang[1,2]

Myeloid cells are a feature of most tissues. Here we show that during development, retinal myeloid cells (RMCs) produce Wnt ligands to regulate blood vessel branching. In the mouse retina, where angiogenesis occurs postnatally[1], somatic deletion in RMCs of the Wnt ligand transporter *Wntless*[2,3] results in increased angiogenesis in the deeper layers. We also show that mutation of *Wnt5a* and *Wnt11* results in increased angiogenesis and that these ligands elicit RMC responses via a non-canonical Wnt pathway. Using cultured myeloid-like cells and RMC somatic deletion of *Flt1*, we show that an effector of Wnt-dependent suppression of angiogenesis by RMCs is Flt1, a naturally occurring inhibitor of vascular endothelial growth factor (VEGF)[4–6]. These findings indicate that resident myeloid cells can use a non-canonical, Wnt–Flt1 pathway to suppress angiogenic branching.

Myeloid cells have a wide array of biological activities that include immune activation, arteriogenesis[7], and regulation of salt balance and blood pressure[8]. Myeloid cells also regulate vascularity. Tumour-associated macrophages influence the growth of blood vessels[9] in part because they are a source of VEGFA[10]. Myeloid cells also promote angiogenic branching[11] and anastamosis[12]. Depending on the context, myeloid cells can be either anti-angiogenic[13] or pro-angiogenic[14]. Here we show that RMCs suppress retinal angiogenesis via a Wnt–Flt1 pathway (Supplementary Fig. 1).

Retinal angiogenesis begins on the day of birth in the mouse with the formation of a superficial vascular plexus (Fig. 1a) that lies within the ganglion cell layer[15]. After formation of this superficial plexus by postpartum day 7 (P7), angiogenic sprouts descend vertically through the retinal layers from P8 to P14 (Fig. 1a). At the outer edge of the inner nuclear layer (INL) the vertical angiogenic sprouts turn and simultaneously branch to form a deep vascular plexus (Fig. 1a). Using antibodies to the vascular endothelial cell (VEC) marker endomucin and to the green fluorescent protein (GFP) of the *c-fms–EGFP* (also known as *Tg(Csf1r-EGFP)1Hume*) transgene that marks RMCs, we show that myeloid cells have a unique spatial relationship with angiogenic tip cells. At the point of turning and branching in the outer INL, myeloid cells and angiogenic sprouts are in close contact (Supplementary Fig. 2a). This was confirmed by labelling with isolectin and F4/80 (F4/80 is also known as Emr1; Supplementary Fig. 2b). We then took advantage of high-intensity isolectin labelling of VECs and RMCs and performed a three-dimensional reconstruction with false colouring to illustrate the overall topology of the angiogenic tip cell–RMC interaction (Fig. 1b). This showed close contact between the two cell types throughout turn-and-branch angiogenesis in the deep retinal layer. Furthermore, after turning, angiogenic tip cells extend within the plane of the deep retinal layer and remain RMC-associated (Fig. 1c). In further defining RMCs,

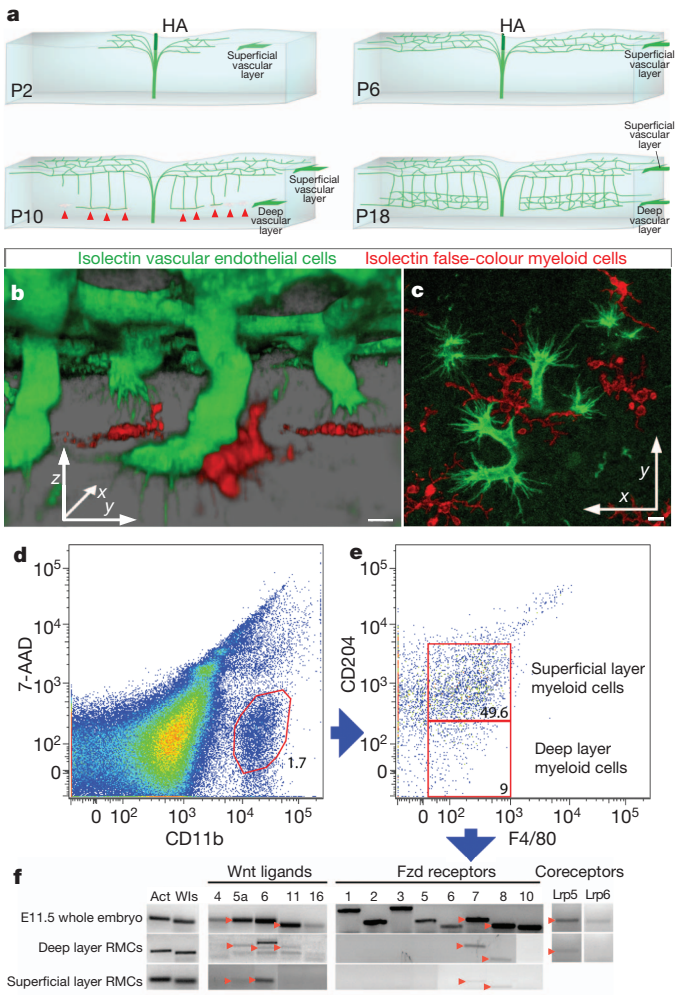


Figure 1 | RMCs interact with VECs and express Wnt components. **a**, Schematic of retina at postnatal days (P) 2, 6, 10, and 18. RMCs interacting with descending vertical sprouts are labelled with red arrowheads. Adapted from ref. 1. **b**, Isolectin-labelled three-dimensional reconstruction of vertical angiogenic sprouts (green) and RMCs (false-colour red). **c**, As in **b** but a two-dimensional image in the deep vascular layer. Scale bars, 5 μm. **d**, **e**, Flow cytometry of deep layer RMCs based on surface markers. 7-AAD, 7-aminoactinomycin D. **f**, PCR for Wnt pathway components on flow-sorted RMCs. Red arrowheads indicate expected sizes. Act, β-actin.

[1]The Visual Systems Group, Divisions of Pediatric Ophthalmology and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. [2]Department of Ophthalmology, University of Cincinnati, Cincinnati, Ohio 45229, USA. [3]Division of Immunobiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. [4]Vascular Biology Laboratory, London Research Institute, Cancer Research UK, London WC2 3PX, UK. [5]Division of Molecular Cardiovascular Biology, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Ohio 45229, USA. [6]Cancer and Developmental Biology Laboratory, National Cancer Institute, Frederick, Maryland 21701, USA. [7]Howard Hughes Medical Institute, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Ohio 45229, USA. [8]Center for Skeletal Disease Research, Van Andel Research Institute, 333 Bostwick NE, Grand Rapids, Michigan 49503, USA. [9]Albert Einstein College of Medicine of Yeshiva University, Jack and Pearl Resnick Campus, 1300 Morris Park Avenue, Bronx, New York 10461, USA. [10]Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. [11]Consultant Group Leader, Vascular Patterning Laboratory, Vesalius Research Center, VIB, Campus Gasthuisberg, B-3000 Leuven, Belgium.

we showed that CD204 (also known as Msr1) was unique to RMCs with amoeboid morphology in the superficial retinal layer (Supplementary Fig. 2c, d). By contrast, deep retinal layer RMCs with extended morphology were CD204-negative (Supplementary Fig. 2c, e). Both layers of RMCs expressed Iba1 (ref. 16) but at different levels (Supplementary Fig. 2c–e). This information allowed us to sort distinct populations of superficial (CD11b+, F4/80+, CD204+) and deep layer (CD11b+, F4/80+, CD204-) RMCs by flow cytometry (Fig. 1d, e).

On the basis of previous work showing vascular regulation by myeloid Wnt ligands[17] we proposed that RMCs might use Wnt ligands to regulate retinal angiogenesis. First, we examined the expression of Wnt ligands and receptors in superficial and deep RMCs isolated by flow cytometry. PCR with reverse transcription (RT–PCR) analysis of the deep RMC population showed expression of Wnt5a, Wnt6 and Wnt11, Fzd7 and Fzd8 as well as the co-receptor Lrp5 (Fig. 1f). With the exception of Wnt5b, which was inconsistent, no other Wnt ligand was detected in deep RMCs. Superficial RMCs expressed similar Wnt and Fzd proteins (Fig. 1f), but also expressed Wnt2b, 3 and 3a (data not shown).

The challenge of genetic analysis when RMCs express many Wnt ligands was addressed by the generation of a *loxP*-flanked conditional allele for the essential Wnt ligand transporter Wls[2,3]. Both superficial and deep RMCs expressed *Wls* (Fig. 1f). *Wls* was deleted using the myeloid cre driver *cfms–icre*[18], which we confirmed was functional in RMCs (Supplementary Fig. 3a–c). To analyse retinal vasculature, we imaged superficial and deep retinal vessels at P18 (Fig. 2a). Quantification of vessel branch points showed that *cfms–icre* alone had no effect (Fig. 3b). Furthermore, compared with control $Wls^{fl/+}$ mice, $Wls^{fl/+}$; *cfms–icre* animals had a normal superficial vascular plexus (Fig. 2a c). By contrast, the deep vascular layer (Fig. 2a, c) showed an overgrowth. Interestingly, no further enhancement of vascular overgrowth was apparent when the myeloid *Wls* deletion was homozygous as in $Wls^{-/fl}$; *cfms–icre* mice (data not shown). Because myeloid cells are positioned below descending vessels at P10, we assessed vessel branching at the base of these sprouts. $Wls^{fl/+}$; *cfms–icre* mice showed reduced simple turning (no branches) and single branching, but significantly more multi-branch events (Fig. 2d). As a weighted mean, the branch index was 2.0 in control and 3.2 in the mutant ($P < 0.0001$).

The higher vascular density of the deep layer in $Wls^{fl/+}$; *cfms–icre* mice could reflect enhanced angiogenesis or defective remodelling. To

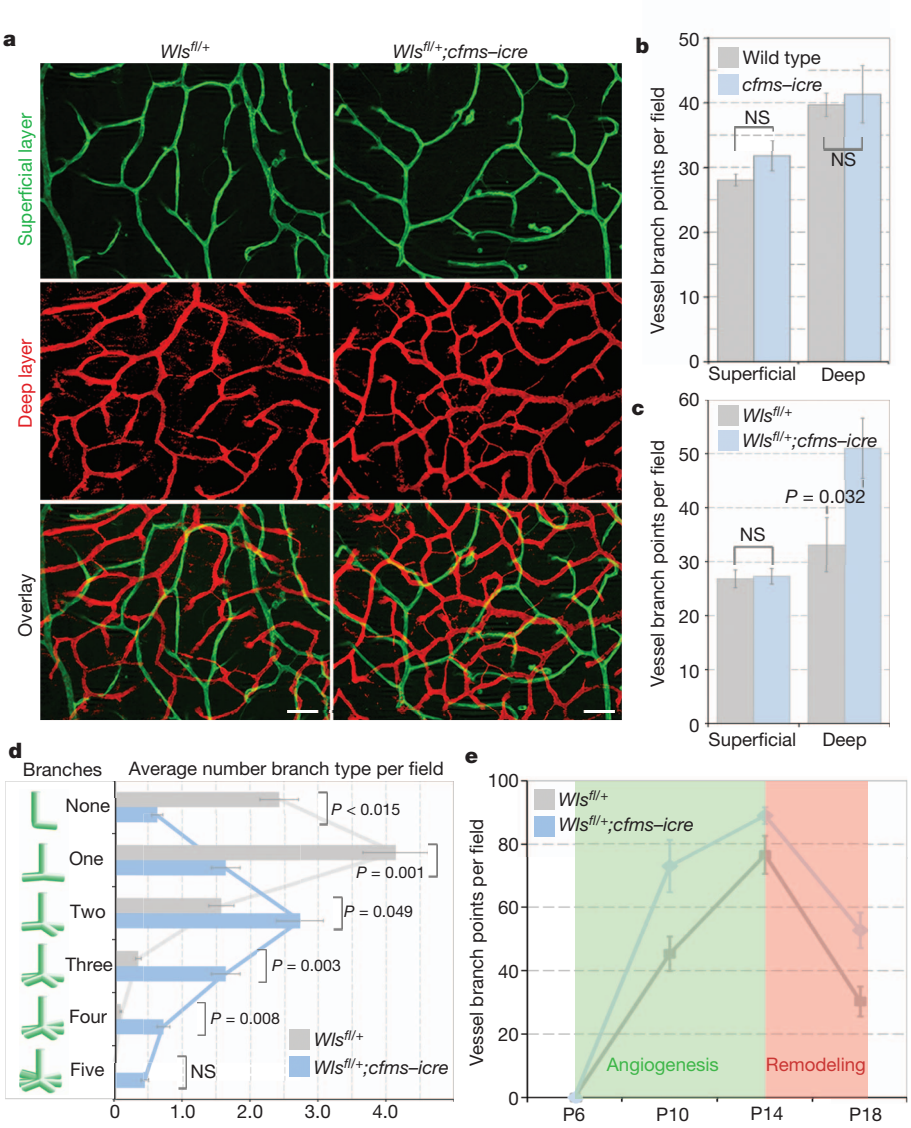


Figure 2 | RMC *Wls* is required for suppression of deep angiogenic branching. **a**, Isolectin labelling of superficial and deep retinal vasculature in $Wls^{fl/+}$ and $Wls^{fl/+}$; *cfms–icre* mice. Scale bars, 50 µm. **b**, **c**, P18 vessel branch points in labelled genotypes. $n = 4$ (**b**), $n = 8$ (**c**). **d**, Branches emanating from the base of vertical sprouts in the P10 deep vascular layer. $n = 8$. **b**–**d** used Student's *t*-test (two-tailed). **e**, Time-course of deep layer branches in indicated genotypes. Shading shows when angiogenesis (green) and remodelling (red) predominate. One-way ANOVA showed $P = 0.0021$. $n = 4$ for each point. Error bars are s.e.m., NS, not significant.

assess this, we counted deep layer branch points in control ($Wls^{fl/+}$) and experimental ($Wls^{fl/+}$; *cfms–icre*) mice over a time-course from P6 to P18 (Fig. 2e). Between P14 and P18 when remodelling predominates (Fig. 2e, red zone) control and mutant graph slopes were nearly identical. By contrast, between P6 and P14 when vessel growth predominates (Fig. 2e, green zone), the slope of the graph is greater in the mutant. There was no difference between control and mutant animals in the number of vertical sprouts connecting superficial and deep layers (Supplementary Fig. 3d). Combined, the data in Fig. 2 are consistent with a model in which myeloid Wnt ligands suppress branch formation as angiogenic sprouts make contact with deep layer RMCs (Fig. 1b, c). These data were corroborated by the vascular overgrowth phenotype of *Wnt5a* (ref. 19) and *Wnt11* (ref. 20) heterozygotes (Supplementary Figs 3e, f and 4).

Wnt5a and Wnt11 are most often associated with non-canonical Wnt signalling[21]. We thus investigated the possibility that Wnt-dependent suppression of deep retinal angiogenesis was a non-canonical response. Wnt3a, but not Wnt5a, elicited a canonical response in SuperTopflash (STF) reporter cells (Supplementary Fig. 5a). Some non-canonical Wnt pathways activate a Ca^{2+} flux[21]. To determine whether myeloid Wnt proteins elicited a Ca^{2+} flux, we added Wnt5a to myeloid-like RAW264.7 cells and measured Ca^{2+}-dependent Fura-2 dye emission. Recombinant Wnt5a increased intracellular Ca^{2+} compared to untreated controls (Fig. 3a and Supplementary Fig. 5b, c). Furthermore, to assess the requirement for Wls in Wnt ligand secretion, we exposed RAW264.7 cells to medium from $Wls^{fl/fl}$ mouse embryonic fibroblasts (MEFs) transfected with *Wnt5a* or *Wnt5a* and *cre* recombinase plasmids. RAW264.7 cells had increased Ca^{2+} flux in response to *Wnt5a*-transfected MEF medium relative to *Wnt5a*- and *cre*-transfected medium (Fig. 3b). This validates the role of Wls in secretion of Wnt5a. Combined, these data show that one myeloid Wnt, Wnt5a, does not stimulate a canonical Wnt response, but can elicit a Ca^{2+} flux characteristic of some non-canonical Wnt responses.

In recent work, it has been shown that Wnt5a loss-of-function can rescue the defects associated with deletion of the canonical Wnt pathway co-receptors Lrp5 and Lrp6 (ref. 22). This finding has suggested that Lrp5/6 deletion actually represents a non-canonical pathway gain-of-function and that in deleting the non-canonical ligand Wnt5a, there is a re-balancing of pathway activity. This hypothesis is supported by biochemical analysis showing that Wnt5a can bind Lrp6, but does not elicit the phosphorylation required for canonical signalling[22]. Canonical, non-canonical reciprocal pathway inhibition has also been demonstrated[23]. These findings argue that a deletion of Lrp5/6 can define whether a Wnt signalling pathway is canonical or non-canonical. If it is canonical, an Lrp5/6 deletion will give the same phenotype as ligand deletion. By contrast, if the signalling pathway is non-canonical, the consequence of Lrp5/6 deletion would be opposite to that of ligand mutation. Thus, to determine whether the RMC response was canonical or non-canonical, we generated a *cfms–icre* somatic mutant of the *Lrp5* coreceptor that is expressed in RMCs (Fig. 1f). Because the result was significantly diminished deep vascular layer density in somatic homozygotes (Fig. 3c–f), a response opposite to ligand deletion, this provides *in vivo* evidence that the Wnt response is non-canonical. Even though the superficial vascular layer was slightly deficient in somatic homozygotes (Fig. 3e) this was not the reason for reduced density in the deep vascular layer as the descending sprout number was unchanged (Fig. 3f).

The VEGF receptor Flt1 (also known as VEGFR1) can suppress angiogenesis because it has limited signalling capacity, a higher affinity for VEGF than Flk1 (also known as VEGFR2) and can sequester VEGF[5,6]. Alternative splicing produces both membrane-tethered and soluble forms[5]. Flt1 is known to contribute to corneal avascularity[4] as well as the selection of angiogenic tip cells when expressed regionally in an existing vessel[24]. Flt1 is also known to be expressed in some myeloid populations[6]. Because angiogenesis in the deep retinal layers is VEGF-dependent[25], Flt1 was a good candidate to mediate Wnt-dependent angiogenic suppression by deep RMCs.

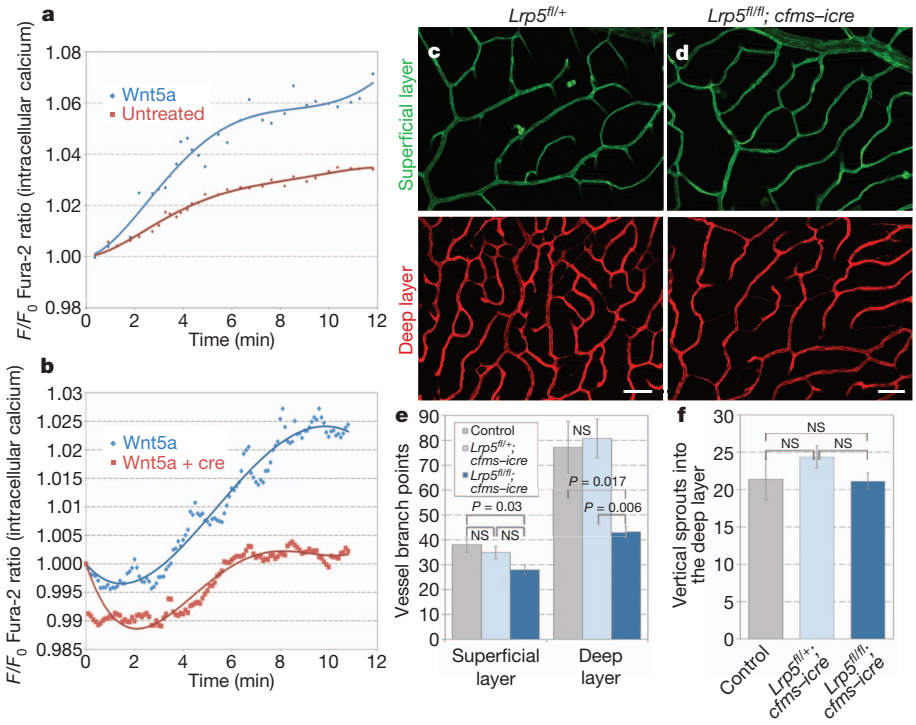


Figure 3 | Angiogenic suppression by RMCs is a non-canonical Wnt response. **a**, **b**, Intracellular Ca^{2+} in RAW264.7 cells treated with Wnt5a (**a**) or supernatant from $Wls^{fl/fl}$ MEFs expressing *Wnt5a* or *Wnt5a* and *cre* (**b**). One Way ANOVA showed $P \leq 0.0001$ for both. **c**, **d**, Isolectin labelling of superficial and deep retinal vasculature in $Lrp5^{fl/+}$ (**c**) and $Lrp^{fl/fl}$; *cfms–icre* (**d**) mice. 50 μm scale bars. **e**, P18 vessel branch points in labelled genotypes. **f**, Vertical vessels connecting to the deep layer in labelled genotypes. **e**, **f**, Used one-way ANOVA with Tukey's post-hoc. $n \geq 8$ per genotype. Errors are s.e.m., NS, not significant.

We thus determined whether *Flt1* was expressed in P12 RMCs by performing end-point RT–PCR on flow-sorted cells using primers that detected both transcripts. This showed that deep RMCs expressed *Flt1* (Fig. 4a). By contrast, superficial RMCs did not (Fig. 4a). We also showed that Iba1-labelled RMCs positioned at the outer edge of the INL labelled with anti Flt1 antibodies (Fig. 4b). We then generated an RMC *Flt1* loss-of-function using *cfms–icre* and the $Flt1^{fl}$ conditional allele[26]. Quantification of retinal vascular branch points in $Flt1^{fl/+}$; *cfms–icre* mice at P18 showed that the superficial layer was unaffected but that the deep layer showed an increase in density (Fig. 4c, d). As with *Wls* somatic mutants, conditional homozygosity for *Flt1* ($Flt1^{fl/fl}$; *cfms–icre*) did not give a further significant increase in vascular density (data not shown). Furthermore, the number of deep RMCs was unchanged (Supplementary Fig. 2g).

These data showed that conditional deletion of *Flt1* produced an enhancement of angiogenesis similar to that observed with conditional deletion of *Wls*. One implication was that Wls and Flt1 might function in the same angiogenesis suppression pathway. To test the possibility that *Flt1* might be regulated by Wnt ligands, we used myeloid-like RAW264.7 cells that, like RMCs, express *Fzd7* and *Fzd8* (Fig. 4e). When stimulated with the 'canonical' ligand Wnt3a, RAW264.7 did not change the level of *Flt1* transcript according to quantitative PCR (Fig. 4f). However, stimulation with Wnt5a produced a threefold increase in the level of both Flt1 isoform transcripts (Fig. 4f). Furthermore, an enzyme-linked immunosorbent assay (ELISA) showed that there was an increased level of soluble Flt1 in conditioned media from Wnt5a-stimulated RAW264.7 cells (Fig. 4g). Wnt3a stimulation again had no effect (Fig. 4g).

As a stringent test of the possibility that Wnt ligands stimulated expression of *Flt1* in deep RMCs, we flow-sorted the $CD11b^{+}$, $F4/80^{+}$, $CD204^{-}$ population from both control ($Wls^{fl/+}$) and experimental ($Wls^{fl/+}$; *cfms–icre*) mice at P12 and performed quantitative PCR for

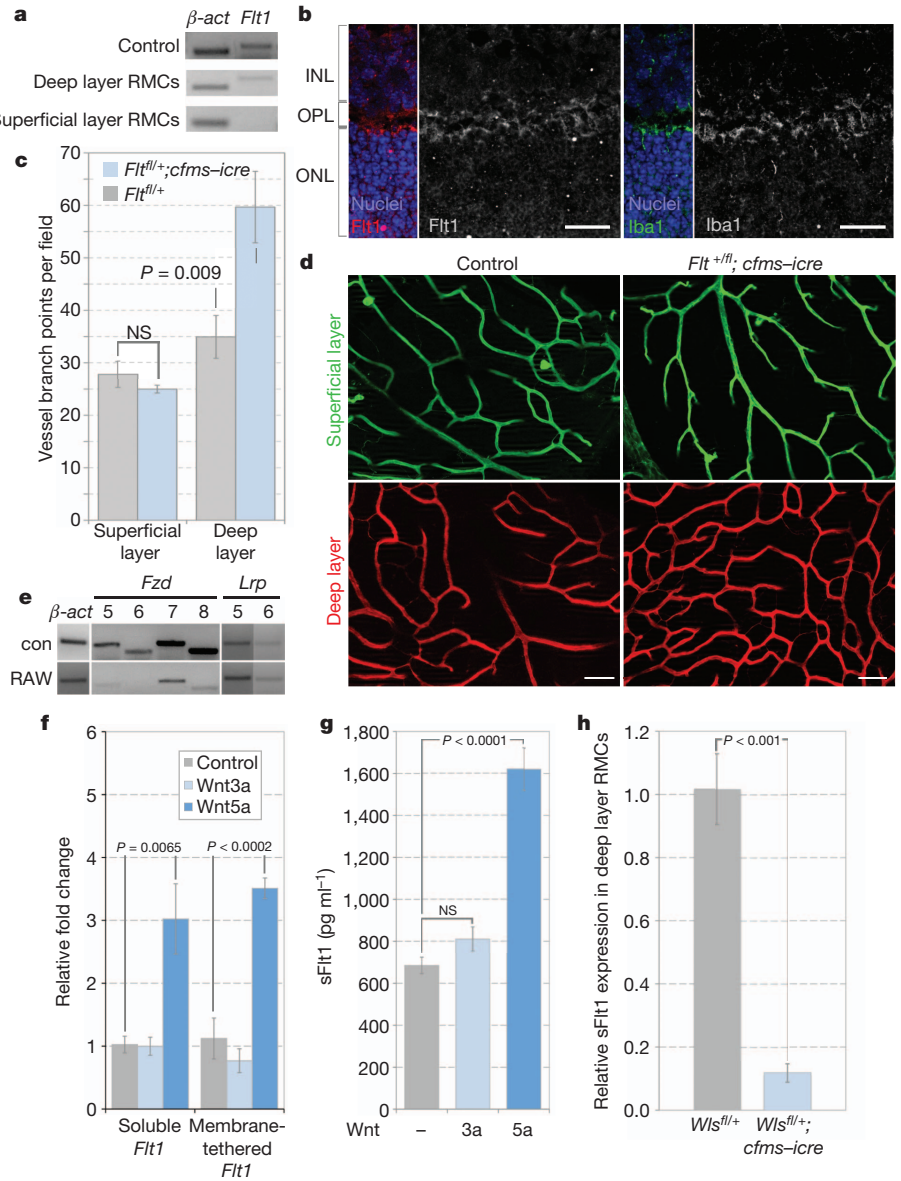


Figure 4 | Flt1 expression in myeloid cells is regulated by a Wnt pathway. **a**, PCR for *Flt1* transcript in flow-sorted RMCs. *β-act*, β-actin transcript. **b**, P14 retinal sections labelled for Flt1 and Iba1. Scale bars, 25 μm. **c**, P18 vessel branch points in the superficial and deep vasculature in $Flt1^{fl/+}$ and $Flt1^{fl/+}$; *cfms–icre* mice ($n = 7$, Student's *t*-test). $n = 7$. **d**, Isolectin labelling of the retinal vasculature in labelled genotypes. Scale bars, 50 μm. **e**, PCR for indicated transcripts in E11.5 whole embryo (con) and RAW264.7 cells. **f**, **g**, quantitative PCR (**f**) for soluble *Flt1* and membrane-tethered *Flt1* in Wnt-treated RAW264.7 cells and ELISA (**g**) for soluble *Flt1* on medium from Wnt-treated RAW264.7 cells ($n = 4$, one-way ANOVA). **h**, Quantitative PCR for soluble *Flt1* in flow-sorted deep RMCs from $Wls^{fl/+}$ and $Wls^{fl/+}$; *cfms–icre* mice ($n = 4$, Student's *t*-test). Errors are s.e.m.; NS, not significant.

Flt1 transcripts. We were never able to amplify the membrane-tethered *Flt1* transcript but showed that there was a 90% reduction in transcript level for soluble *Flt1* (Fig. 4h). This was consistent with the maximal phenotype observed in the heterozygous conditional *Wls* mutants. Although a heterozygous phenotype is not unusual, it is perhaps surprising that conditional homozygosity gave no further significant change. This kind of response might be explained by a low signalling sensitivity threshold for the Wnt–Flt1 pathway.

The data we describe show that RMCs can modulate angiogenic responses directly by producing the VEGF inhibitory receptor Flt1. Unexpectedly, we also show that the production of Flt1 depends on myeloid non-canonical Wnt ligands. In this setting, myeloid Wnt ligands might function in autocrine stimulation as has been documented for cultured macrophages[27] or might operate via a more complex pathway involving another cell type as an intermediate. The Wnt–Flt1 response represents a new pathway for the regulation of VEGF-stimulated angiogenesis. The targeting of appropriate Fzd receptors or other Wnt–Flt1 pathway components may offer new opportunities to modulate the production of Flt1 and thus, the VEGF-stimulated angiogenic response. Because macrophage-related cells are ubiquitous and highly mobile, it is possible that the Wnt–Flt1 pathway will be a general means to suppress VEGF locally and thus to restrain vascular responses. In future studies it will be interesting to determine, for example, if this pathway is active in the suppression of wound angiogenesis by macrophages[13] or is inactivated in macrophage-dependent tumour angiogenesis[28].

Additionally, our observations are consistent with a string of recent papers describing the role of myeloid cells beyond their well-documented functions in innate immunity[7,8,17,29]. These findings build on an idea articulated nearly a century ago[30]. It was suggested then that phagocytes originally evolved to regulate developmental processes, and that their immune functions were a later evolutionary adaptation. Macrophages were proposed to be the "policemen" of multi-cellular organisms, and that they could establish "harmony from chaos". Here we show, in the setting of the retina, that myeloid cells do just that by fine-tuning vascular density and directing vascular traffic.

METHODS SUMMARY

We prepared and stained retinas as reported previously[1]. To isolate RMCs, we digested retinas, pre-enriched with CD11b beads, and sorted for surface makers with the FACSAria II. To obtain conditioned medium from MEFs, we performed transient transfections with Wnt5a, Thy1.1 and cre plasmids, sorted Thy1.1-positive MEFS, and re-plated transfected cells. We performed calcium imaging on RAW264.7 cells loaded in Ringer's solution with 5 μM Fura-2 AM and imaged at 510 nm at 1 Hz after excitation at 340 nm and 380 nm. All animal experiments were performed in accordance with IACUC-approved guidelines and regulations.

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

Received 20 July 2010; accepted 30 March 2011.
Published online 29 May 2011.

1. Gerhardt, H. *et al.* VEGF guides angiogenic sprouting utilizing endothelial tip cell filopodia. *J. Cell Biol.* **161,** 1163–1177 (2003).
2. Ching, W. & Nusse, R. A dedicated Wnt secretion factor. *Cell* **125,** 432–433 (2006).
3. Carpenter, A. C., Rao, S., Wells, J. M., Campbell, K. & Lang, R. A. Generation of mice with a conditional null alelle for *Wntless*. *Genesis* **48,** 554–558 (2010).
4. Ambati, B. K. *et al.* Corneal avascularity is due to soluble VEGF receptor-1. *Nature* **443,** 993–997 (2006).
5. Kendall, R. L. & Thomas, K. A. Inhibition of vascular endothelial cell growth factor activity by an endogenously encoded soluble receptor. *Proc. Natl Acad. Sci. USA* **90,** 10705–10709 (1993).
6. Shibuya, M. Structure and dual function of vascular endothelial growth factor receptor-1 (Flt-1). *Int. J. Biochem. Cell Biol.* **33,** 409–420 (2001).
7. Pipp, F. *et al.* VEGFR-1-selective VEGF homologue PlGF is arteriogenic: evidence for a monocyte-mediated mechanism. *Circ. Res.* **92,** 378–385 (2003).
8. Machnik, A. *et al.* Macrophages regulate salt-dependent volume and blood pressure by a vascular endothelial growth factor-C-dependent buffering mechanism. *Nature Med.* **15,** 545–552 (2009).
9. Lin, E. Y. & Pollard, J. W. Tumor-associated macrophages press the angiogenic switch in breast cancer. *Cancer Res.* **67,** 5064–5066 (2007).
10. Stockmann, C. *et al.* Deletion of vascular endothelial growth factor in myeloid cells accelerates tumorigenesis. *Nature* **456,** 814–818 (2008).
11. Kubota, Y. *et al.* M-CSF inhibition selectively targets pathological angiogenesis and lymphangiogenesis. *J. Exp. Med.* **206,** 1089–1102 (2009).
12. Fantin, A. *et al.* Tissue macrophages act as cellular chaperones for vascular anastomosis downstream of VEGF-mediated endothelial tip cell induction. *Blood* **116,** 829–840 (2010).
13. Martin, P. *et al.* Wound healing in the PU.1 null mouse–tissue repair is not dependent on inflammatory cells. *Curr. Biol.* **13,** 1122–1128 (2003).
14. Grunewald, M. *et al.* VEGF-induced adult neovascularization: recruitment, retention, and role of accessory cells. *Cell* **124,** 175–189 (2006).
15. Saint-Geniez, M. & D'Amore, P. A. Development and pathology of the hyaloid, choroidal and retinal vasculature. *Int. J. Dev. Biol.* **48,** 1045–1058 (2004).
16. Mendes-Jorge, L. *et al.* Scavenger function of resident autofluorescent perivascular macrophages and their contribution to the maintenance of the blood-retinal barrier. *Invest. Ophthalmol. Vis. Sci.* **50,** 5997–6005 (2009).
17. Lobov, I. B. *et al.* WNT7b mediates macrophage-induced programmed cell death in patterning of the vasculature. *Nature* **437,** 417–421 (2005).
18. Deng, L. *et al.* A novel mouse model of inflammatory bowel disease links mammalian target of rapamycin-dependent hyperproliferation of colonic epithelium to inflammation-associated tumorigenesis. *Am. J. Pathol.* **176,** 952–967 (2010).
19. Yamaguchi, T. P., Bradley, A., McMahon, A. P. & Jones, S. A. Wnt5a pathway underlies outgrowth of multiple structures in the vertebrate embryo. *Development* **126,** 1211–1223 (1999).
20. Majumdar, A., Vainio, S., Kispert, A., McMahon, J. & McMahon, A. P. Wnt11 and Ret/Gdnf pathways cooperate in regulating ureteric branching during metanephric kidney development. *Development* **130,** 3175–3185 (2003).
21. Seifert, J. R. & Mlodzik, M. Frizzled/PCP signalling: a conserved mechanism regulating cell polarity and directed motility. *Nature Rev. Genet.* **8,** 126–138 (2007).
22. Bryja, V. *et al.* The extracellular domain of Lrp5/6 inhibits noncanonical Wnt signaling *in vivo*. *Mol. Biol. Cell* **20,** 924–936 (2009).
23. Grumolato, L. *et al.* Canonical and noncanonical Wnts use a common mechanism to activate completely unrelated coreceptors. *Genes Dev.* **24,** 2517–2530 (2010).
24. Chappell, J. C., Taylor, S. M., Ferrara, N. & Bautch, V. L. Local guidance of emerging vessel sprouts requires soluble Flt-1. *Dev. Cell* **17,** 377–386 (2009).
25. Haigh, J. J. *et al.* Cortical and retinal defects caused by dosage-dependent reductions in VEGF-A paracrine signaling. *Dev. Biol.* **262,** 225–241 (2003).
26. Lichtenberger, B. M. *et al.* Autocrine VEGF signaling synergizes with EGFR in tumor cells to promote epithelial cancer development. *Cell* **140,** 268–279 (2010).
27. Blumenthal, A. *et al.* The Wingless homolog WNT5A and its receptor Frizzled-5 regulate inflammatory responses of human mononuclear cells induced by microbial stimulation. *Blood* **108,** 965–973 (2006).
28. Stockmann, C. *et al.* Deletion of vascular endothelial growth factor in myeloid cells accelerates tumorigenesis. *Nature* **456,** 814–818 (2008).
29. Lin, S. L. *et al.* Macrophage Wnt7b is critical for kidney repair and regeneration. *Proc. Natl Acad. Sci. USA* **107,** 4194–4199 (2010).
30. Tauber, A. I. Metchnikoff and the phagocytosis theory. *Nature Rev. Mol. Cell Biol.* **4,** 897–901 (2003).

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

Acknowledgements We thank P. Speeg for technical assistance and A. P. McMahon for the Wnt11 mice. This work was supported by the NIH (J.A.S., M.W-K., J.W.P., J.D.M., T.Y., B.O.W., R.A.L.) by the HHMI (J.D.M.) and Cancer Research UK (H.G.).

Author Contributions R.A.L provided project leadership and wrote the manuscript with J.A.S. J.A.S., I.L., S.R., H.G., and R.A.L. designed the experiments. J.A.S, I.L., S.R., G.M., A.C.C., A.R.B., J.F., and R.A. performed the experiments. S.R., J.W.P., T.Y., N.F. and B.O.W developed critical reagents. Experimental supervision and helpful discussions were provided by M.W-K., J.D.M., S.R., J.W.P., and H.G.

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. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.A.L. (richard.lang@cchmc.org).

METHODS

Quantification of retinal vasculature and isolation of RMCs. Retinas were prepared and imaged as reported[1] except that Alexa Fluor 488 isolectin GS-IB_4 (Invitrogen) was used to label retinal vessels and RMCs. Other antibodies included anti-endomucin (Santa Cruz, V.7C7), anti-GFP (Abcam), anti-Iba1 (Wako), anti-CD204 (AbD Serotec), and anti-Flt1 (R&D). For quantification, we used 200× magnification images located at the retinal periphery between artery and vein. For each genotype, at least three fields were analysed from at least four animals from a minimum of two litters. Control and experimental animals were littermates. This minimized the effect of strain background in producing variation in vascular density[31–34]. For flow sorting, retinas were digested and sorted as reported[16] except that 0.5 mg ml^{-1} DNase II (Sigma Aldrich) was used. Furthermore, myeloid cells were pre-enriched using CD11b beads (Miltenyi Biotech). Cells were then incubated with anti-CD16/32 (clone 24.G2) for 30 min and labelled with monoclonal antibodies to phycoerythrin-Cy7 conjugated anti-CD11b (clone M1/70), allophycocyanin-Cy7 or peridinin chlorophyll A protein-Cy5.5 conjugated anti-F4/80 (clone BM8), Alexa-647 conjugated anti-CD204 (clone 2F8), and 7-aminoactinomycin D. Cells were sorted with a FACSAria II running DiVa software.

RNA isolation and RT–PCR. RNA was isolated using RNeasy (Qiagen). Quantitative PCR was performed with QuantiTect SYBR green (Qiagen). Primers are listed in Supplementary Table 1.

***In vitro* analysis.** MEFs were isolated as described[35] from $Wls^{fl/fl}$ mice[36] and transfected with combinations of 2.5 μg Wnt5a, Thy1.1 and cre plasmids using TransIT-2020 (Mirus). MEFs were sorted using magnetic beads for Thy1.1, replated and supernatant was collected after 24 h. RAW264.7 cells (ATCC) were grown in DMEM (10% FBS) and treated with recombinant Wnt3a (10 ng ml^{-1}, R&D) or Wnt5a (500 ng ml^{-1}, R&D). ELISA was performed using the sVEGFR1 Quantikine kit (R&D). For calcium imaging, RAW264.7 cells were loaded in Riger's solution with 5 μM Fura-2 AM and imaged at 510 nm at 1 Hz after excitation at 340 nm and 380 nm. Data was acquired using EasyRatioPro. Ca^{2+} traces are 50-cell averages from two experiments.

Statistics. All statistical tests used are stated in the figure legends. In analysing quantitative PCR data, the *P* values refer to a comparison of the $\Delta\Delta C_t$ values.

Animals. Breeding and genotyping of Wls^{fl} (ref. 36), *cfms–icre* (ref. 18), Z/EG (ref. 37), $Wnt5a^{+/-}$ (ref. 19), $Wnt11^{+/-}$ (ref. 20) and Flt^{fl} (ref. 26) was performed as previously described. $Lrp5^{fl}$ will be described in detail in a forthcoming publication. The allele is a conventional design where exon 2 of the *Lrp5* gene is flanked by LoxP sites. It has been confirmed that deletion between the LoxP sites produces a loss-of-function. All animal experimentation was carried out using protocols approved by the Institutional Animal Care and Use Committee.

31. Rohan, R. M., Fernandez, A., Udagawa, T., Yuan, J. & D'Amato, R. J. Genetic heterogeneity of angiogenesis in mice. *FASEB J.* **14,** 871–876 (2000).
32. Gao, G. *et al.* Difference in ischemic regulation of vascular endothelial growth factor and pigment epithelium–derived factor in Brown Norway and Sprague Dawley rats contributing to different susceptibilities to retinal neovascularization. *Diabetes* **51,** 1218–1225 (2002).
33. Chan, C. K. *et al.* Mouse strain-dependent heterogeneity of resting limbal vasculature. *Invest. Ophthalmol. Vis. Sci.* **45,** 441–447 (2004).
34. Chan, C. K. *et al.* Differential expression of pro- and antiangiogenic factors in mouse strain-dependent hypoxia-induced retinal neovascularization. *Lab. Invest.* **85,** 721–733 (2005).
35. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. *Manipulating the mouse embryo: a laboratory manual*. 3rd edn, 371–373 (Cold Spring Harbor Laboratory Press, 2003).
36. Carpenter, A. C., Rao, S., Wells, J. M., Campbell, K. & Lang, R. A. Generation of mice with a conditional null allele for *Wntless*. *Genesis* **48,** 554–558 (2010).
37. Novak, A., Guo, C., Yang, W., Nagy, A. & Lobe, C. G. Z/EG, a double reporter mouse line that expresses enhanced green fluorescent protein upon Cre-mediated excision. *Genesis* **28,** 147–155 (2000).

LETTER

doi:10.1038/nature10002

Determinants of nucleosome organization in primary human cells

Anton Valouev[1], Steven M. Johnson[2], Scott D. Boyd[1], Cheryl L. Smith[1], Andrew Z. Fire[1,3] & Arend Sidow[1,3]

Nucleosomes are the basic packaging units of chromatin, modulating accessibility of regulatory proteins to DNA and thus influencing eukaryotic gene regulation. Elaborate chromatin remodelling mechanisms have evolved that govern nucleosome organization at promoters, regulatory elements, and other functional regions in the genome[1]. Analyses of chromatin landscape have uncovered a variety of mechanisms, including DNA sequence preferences, that can influence nucleosome positions[2–4]. To identify major determinants of nucleosome organization in the human genome, we used deep sequencing to map nucleosome positions in three primary human cell types and *in vitro*. A majority of the genome showed substantial flexibility of nucleosome positions, whereas a small fraction showed reproducibly positioned nucleosomes. Certain sites that position *in vitro* can anchor the formation of nucleosomal arrays that have cell type-specific spacing *in vivo*. Our results unveil an interplay of sequence-based nucleosome preferences and non-nucleosomal factors in determining nucleosome organization within mammalian cells.

Previous studies in model organisms[3–7] as well as initial analyses in human cells[8] have identified fundamental aspects of nucleosome organization. Here we focus on the dynamic relationships between sequence-based nucleosome preferences and chromatin regulatory function in primary human cells. We mapped tissue-specific and DNA-encoded nucleosome organization across granulocytes and two types of T cells ($CD4^+$ and $CD8^+$) isolated from the blood of a single human donor, by isolating cellular chromatin and treating it with micrococcal nuclease (MNase) followed by deep sequencing of the resulting nucleosome-protected fragments (Methods, Supplementary Fig. 1). To provide sufficient depth for both local and global analyses, we used high-throughput SOLiD technology, generating 584, 342 and 343 million mapped reads for granulocytes, $CD4^+$ and $CD8^+$ T cells, respectively. These are equivalent to 16–28× genome coverage by 147 bp nucleosome footprints (cores; see Methods). The depth of sequence was critical for our subsequent analysis: although shallower coverage can illuminate features of nucleosome positions through statistical analysis (for example, refs 6, 8), any definitive map and thus comparison of static and dynamic positioning requires high sequence coverage throughout the genome.

To provide complementary data on purely sequence-driven nucleosome positioning in the absence of cellular influences, we reconstituted genomic DNA *in vitro* with recombinantly derived histone octamers to produce *in vitro* nucleosomes (Methods, Supplementary Fig. 2), and generated over 669 million mapped reads, representing 32× core coverage of the genome. To identify primary nucleosome positioning sites in DNA, the reconstitution was performed under conditions of DNA excess (see methods). We also generated a control data set of 321 million mapped reads from MNase-digested naked DNA. In the population of granulocytes (our deepest *in vivo* data set), over 99.5% of the mappable genome is engaged by nucleosomes (Methods), and 50 percent of nucleosome-depleted bases occur in regions shorter than 160 bp.

We first focused on global patterns of nucleosome positioning and spacing by calculating fragment distograms and phasograms[6,7,9]. Distograms (histograms of distances between mapped reads' start positions aligning in opposing orientation, Supplementary Fig. 3a) reveal the average core fragment size as a peak if there are many sites in the genome that contain consistently positioned nucleosomes. A positioning signal that is strongly amplified by conditioning the analysis on sites with three or more read starts (reflecting a positioning preference; 3-pile subset), is present not only *in vivo* (Fig. 1a), but also *in vitro* (Fig. 1b), demonstrating that many genomic sites bear intrinsic, sequence-driven, positioning signals. Phasograms (histograms of distances between mapped reads' start positions aligning in the same orientation, Supplementary Fig. 3b) reveal consistent spacing of positioned nucleosomes by exhibiting a wave-like pattern with a period that represents genome-average internucleosome spacing. In granulocytes, the wave peaks are 193 bp apart (Fig. 1c, adjusted $R^2 = 1$, P-value $< 10^{-15}$), which, given a core fragment length of 147 bp, indicates an internucleosome linker length of 46 bp. By contrast, the phasograms of both types of T cells have spacing that is wider by 10 bp (Fig. 1d), equivalent to a 56 bp average linker length. These results are consistent with classical observations of varying nucleosome phases in different cell types[10,11]. Linker length differences have been tied to differences in linker histone gene expression[12,13], which we found to be 2.4 times higher in T cells compared to granulocytes (84 reads per kilobase of mature transcript per million mapped reads (RPKM)[14] vs 35 RPKM). The *in vitro* phasogram (Fig. 1e) reveals no detectable stereotypic spacing of positioned nucleosomes, demonstrating a lack of intrinsic phasing among DNA-encoded nucleosome positioning sites.

Using a positioning stringency metric (Methods; Supplementary Fig. 4) that quantifies the fraction of defined nucleosome positions within a given segment, we calculated the fraction of the genome that is occupied by preferentially positioned nucleosomes at different stringency thresholds. The maximum number of sites at which some positioning preference can be detected statistically is 120 million, covering just over 20% of the genome (Supplementary Fig. 5) at the low stringency of 23%. Thus, the majority of nucleosome positioning preferences is weak, and nucleosomes across the majority of the human genome are not preferentially positioned, either by sequence or by cellular function.

Next we focused on how transcription and chromatin functions affect nucleosome organization regionally. For each cell type, we generated deep RNA-seq data and binned genes into groups according to their expression levels. The average spacing of nucleosomes was greatest within silent genes ($CD4^+$ T cells, 206 bp, Fig. 2a) and decreased by as much as 11 bp as the expression levels went up (t-statistic P-value $= 6.5 \times 10^{-34}$). This suggests that transcription-induced cycles of nucleosome eviction and reoccupation cause denser packing of nucleosomes and slight reduction in nucleosome occupancy (Supplementary Fig. 6). On the basis of this result, we hypothesized that higher-order chromatin organization as implied by specific

[1]Department of Pathology, Stanford University School of Medicine, 300 Pasteur Drive, Stanford, California 94305, USA. [2]Department of Microbiology and Molecular Biology, Brigham Young University, 757 WIDB, Provo, Utah 84602-5253, USA. [3]Department of Genetics, Stanford University School of Medicine, Pasteur Drive, Stanford, California 94305, USA.

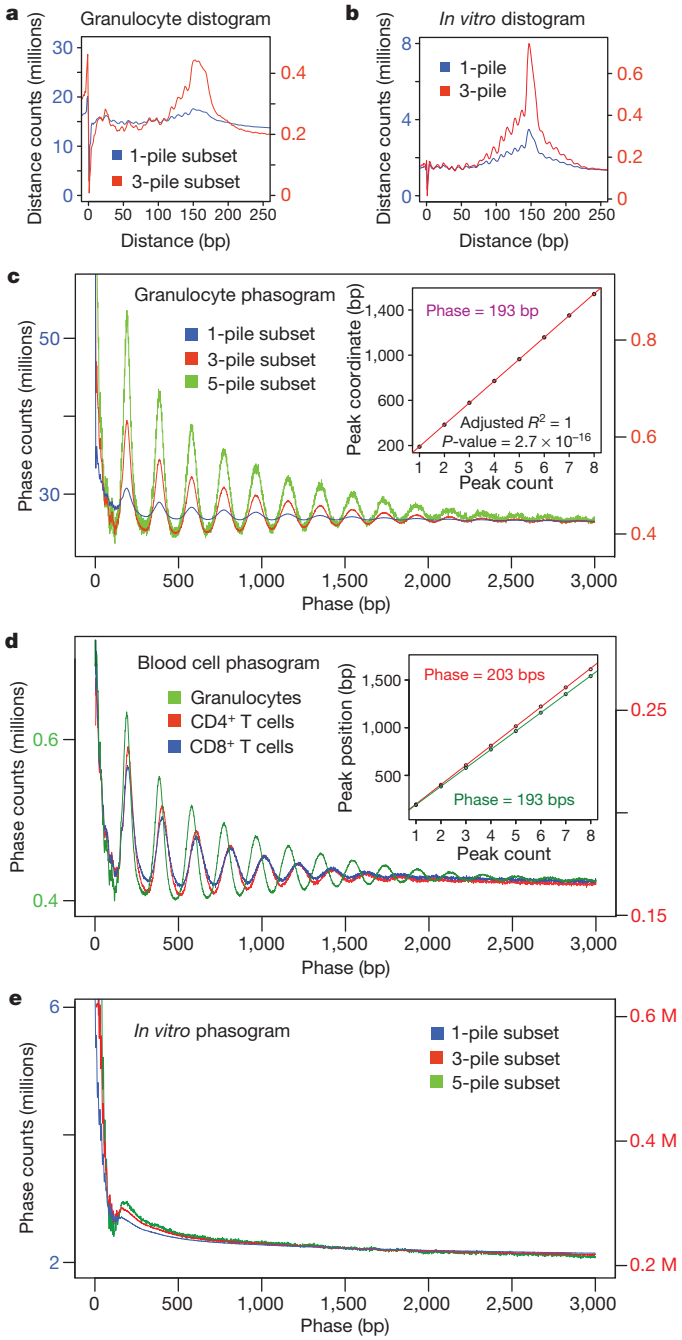


Figure 1 | Global parameters of cell-specific nucleosome phasing and positioning in human. **a**, *In vivo* granulocyte distogram (calculation explained in Supplementary Fig. 3a). *x*-axis represents the range of recorded distances. *y*-axis represents frequencies of observed distances within 1-pile (blue) and 3-pile (red) subsets. 1-pile subset represents the entire data set, 3-pile subset represents a subset of sites containing three or more coincident read starts. **b**, Distogram of the *in vitro* reconstituted nucleosomes showing 1-pile and 3-pile subsets as in (**a**). **c**, *In vivo* granulocyte phasogram (calculation explained in Supplementary Fig. 3b). *x*-axis shows the range of recorded phases. *y*-axis shows frequencies of corresponding phases. Phasograms of 1-pile, 3-pile and 5-pile subsets are plotted. Inset, linear fit to the positions of the phase peaks within 3-pile subsets (slope = 193 bp). **d**, Phasograms of blood cell types. Inset, linear fits in CD4$^+$ T cells (203 bp) and granulocytes (193 bp). **e**, Phasograms of 1-pile, 3-pile and 5-pile subsets in the *in vitro* data.

chromatin modifications might be associated with specific spacing patterns. Using previously published ChIP-seq data, we identified regions of enrichment[15] for histone modifications that are found within heterochromatin (H3K27me3, H3K9me3)[16], gene-body euchromatin

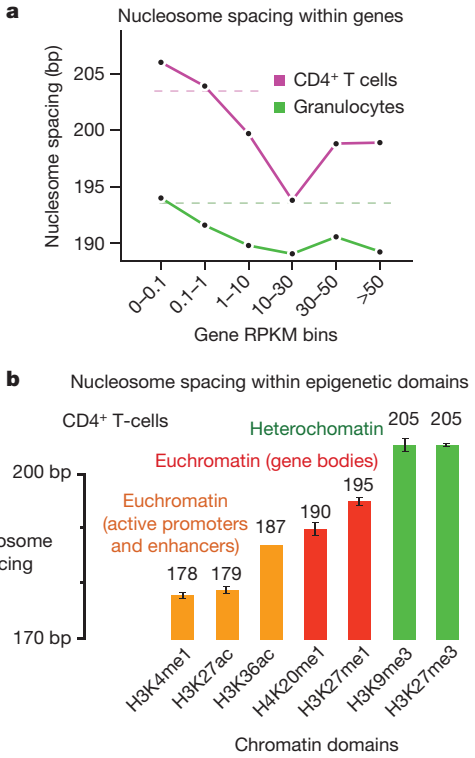


Figure 2 | Transcription and chromatin modification-dependent nucleosome spacing. **a**, Nucleosome spacing as a function of transcriptional activity. *x*- axis represents gene expression values binned according to RPKM values. Internucleosome spacing is plotted along the *y*-axis. Dashed lines represent genome-wide average spacing for each cell type. **b**, Nucleosome spacing within genomic regions marked by specific histone marks in CD4$^+$ T cells. Bar height plots estimated nucleosome spacing for each histone modification. Bar colours differentiate chromatin types (euchromatin vs heterochromatin).

(H4K20me1, H3K27me1)[16], or euchromatin associated with promoters and enhancers (H3K4me1, H3K27ac, H3K36ac)[17], and estimated spacing of nucleosomes for each of these epigenetic domains. We found that active promoter-associated domains contained the shortest spacing of 178–187 bp, followed by a larger spacing of 190–195 bp within the body of active genes, whereas heterochromatin spacing was largest at 205 bp (Fig. 2b). These results reveal striking heterogeneity in nucleosome organization across the genome that depends on global cellular identity, metabolic state, regional regulatory state, and local gene activity.

To characterize DNA signals responsible for consistent positioning of nucleosomes, we identified 0.3 million sites occupied *in vitro* by nucleosomes at high stringency (>0.5; Methods). The region occupied by the centre of the nucleosome (dyad) exhibits a significant increase in G/C usage (Poisson *P*-value $< 10^{-100}$; Fig. 3a). Flanking regions increase in A/T usage as the positioning strength increases (Fig. 3b). A subset of *in vitro* positioned nucleosomes (stringency > 0.5) which are also strongly positioned *in vivo* (stringency > 0.4) revealed increased A/T usage within the flanks (Fig. 3c) compared to *in vitro*-only positioning sites (Fig. 3a), which underscores the importance of flanking repelling elements for positioning *in vivo*. We term such elements with strong G/C cores and A/T flanks 'container sites' to emphasize the proposed positioning mechanism (Fig. 3d). This positioning signal is different from a 10-bp dinucleotide periodicity observed in populations of nucleosome core segments isolated from a variety of species[18,19] and proposed to contribute to precise positioning and/or rotational setting of DNA on nucleosomes[19] on a fine scale (Supplementary Fig. 7). G/C-rich signals are known to promote nucleosome occupancy[20,21], whereas AA-rich sequences repel nucleosomes[4], and our data demonstrate that precise arrangement of a core-length attractive segment flanked by repelling sequences can produce a strongly positioned nucleosome (Fig. 3d).

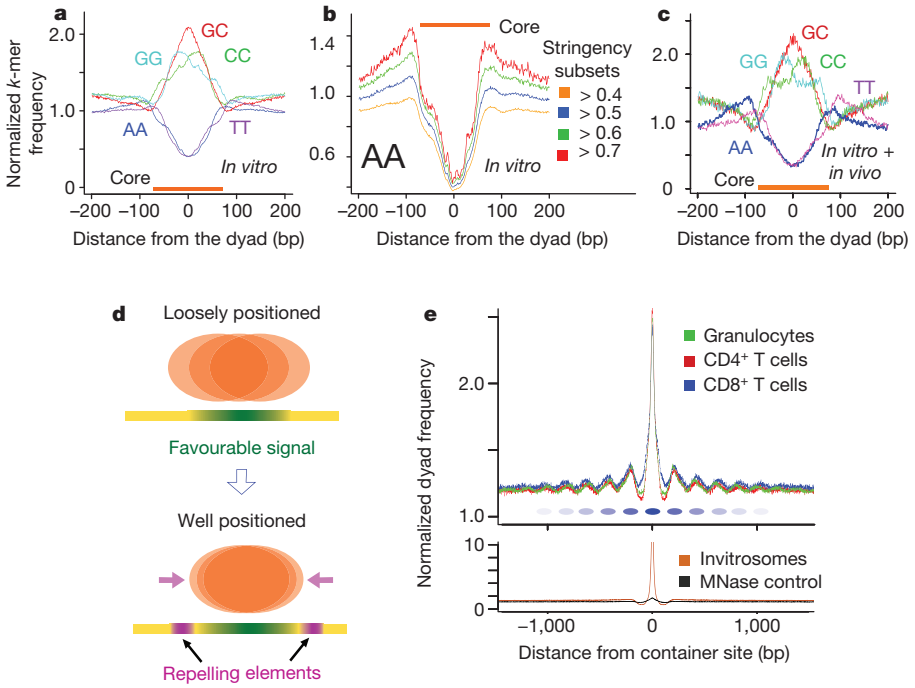


Figure 3 | Sequence signals that drive nucleosome positioning. **a**, Sequence signals within sites containing moderately positioned *in vitro* nucleosomes (stringency > 0.5). Distance from the positioned dyad to a given dinucleotide is plotted along the *x*-axis; *y*-axis represents frequency of a given *k*-mer divided by its genome-wide expectation. The 147-bp footprint of a nucleosome is indicated by an orange band. **b**, Changes in AA dinucleotide usage with increasing positioning stringency. *x* and *y* axes same as in (**a**). Curves of AA usage within the sites of increasingly positioned dyads are shown (stringency cutoffs of 0.4, 0.5, 0.6, 0.7). **c**, Sequence signals within sites containing *in-vitro*-positioned nucleosomes (stringency > 0.5) that also have high *in vivo* stringency (stringency > 0.4). *x* and *y* axes same as in (**a**). **d**, Schematic depiction of the container site positioning mechanism. The C/G-rich core area (green) favours occupancy, but does not precisely position the nucleosome (top). Adding flanking A/T-rich repelling elements (purple, bottom) restricts the position of the nucleosome. **e**, Nucleosome organization around container sites *in vivo* and *in vitro*. *x*-axis represents distances from the dyads to container sites (based on 300,000 container sites). Frequencies of nucleosome dyads around those sites are plotted along the *y*-axis. The upper plot shows distribution of *in vivo* dyads across $CD4^+$ cells, $CD8^+$ cells and granulocytes. The ovals depict hypothetical nucleosome positions across the site with colour intensities reflecting their positioning strength. The lower plot shows distribution of dyads *in vitro* and in MNase control.

Dyad frequencies around container sites (Fig. 3e) show a strong peak of enrichment *in vivo*, confirming that DNA positions nucleosomes *in vivo* over these sites. Additionally, wave-like patterns emanate from these sites *in vivo* (but not *in vitro*), reflecting the nucleation of phased arrays by positioned cellular nucleosomes. Viewing these results in light of the nucleosome barrier model[22], which proposes that nucleosomes are packed into positioned and phased arrays against a chromatin barrier, we conclude that sequence-positioned nucleosome can initiate propagation of adjacent stereotypically positioned nucleosomes. Importantly, wave periods around container sites are shorter in granulocytes than in T cells, allowing tissue-specific variation in linker length (Fig. 1d) to alter placement of nucleosomes over distances of as much as 1 kilobase from an initial container site. Functional consequences of such rearrangements might include global shifts in regulatory properties that could contribute to distinct transcription factor accessibility profiles in different cell types.

The cellular environment can drive nucleosomes to sequences not intrinsically favourable to being occupied, as is evident in a genome-wide comparison of observed nucleosome coverage of all possible tetranucleotides between the granulocyte and the *in vitro* data (Fig. 4a). *In vitro*, nucleosome occupancy is strongly associated with AT/GC content, but this preference is abolished *in vivo*; the exception are C/G rich tetramers that contain CpG dinucleotides, which show a 30% reduction in apparent nucleosome occupancy despite having high core coverage *in vitro*. Consistent with this, CpG islands are fivefold depleted for observed nucleosome coverage *in vivo* (Fig. 4b). No such decrease is observed in the *in vitro* data set.

The decreased nucleosome occupancy of promoters could be due to promoter-related functions of mammalian CpG islands, similar to promoter-associated nucleosome-free regions observed in flies[23] and yeast[5], which do not have CpG islands. We therefore analysed transcription-dependent nucleosome packaging around promoters. As in other organisms[23–27], promoters of active genes have a nucleosome-free region (NFR) of about 150 bp overlapping the transcriptional start site and arrays of well-positioned and phased nucleosomes that radiate from the NFR (Fig. 4c). A notable reduction in apparent nucleosome occupancy extends up to 1 kb into the gene body. We also observed consistent nucleosome coordinates in an independent data set of H3K4me3-bearing nucleosomes[16] (Fig. 4d). Comparison of the nucleosome data (Fig. 4d) with binding patterns of RNA polymerase II[16] (Fig. 4d) around active promoters indicates that phasing of positioned nucleosomes can be explained by packing of nucleosomes against Pol II stalled at the promoter, with Pol II potentially acting as the 'barrier'. The set of inactive promoters, by contrast, exhibits neither a pronounced depletion of nucleosomes, nor a positioning and phasing signal (Fig. 4c). The transition of an inactive promoter to an active one is therefore likely to involve eviction of nucleosomes, coupled with positioning and phasing of nucleosomes neighbouring RNA Pol II (Fig. 4e). These results indicate that CpG-rich segments in mammalian promoters override intrinsic signals of high nucleosome affinity (Supplementary Fig. 8) to become active; this would be in contrast to fly and yeast, where AT-rich promoters may comprise intrinsic sequence signals that are particularly prone to nucleosome eviction[28].

To explore how regulatory factors interact with sequence signals to influence nucleosome organization outside of promoters, we focused on binding sites of the NRSF/REST repressor protein[15] and the insulator protein CTCF. NRSF and CTCF sites are flanked by arrays of positioned nucleosomes (Fig. 4f and Supplementary Fig. 9), consistent with barrier-driven packing previously reported for CTCF[29,30]. Both proteins occupy additional linker space, with NRSF taking up an extra

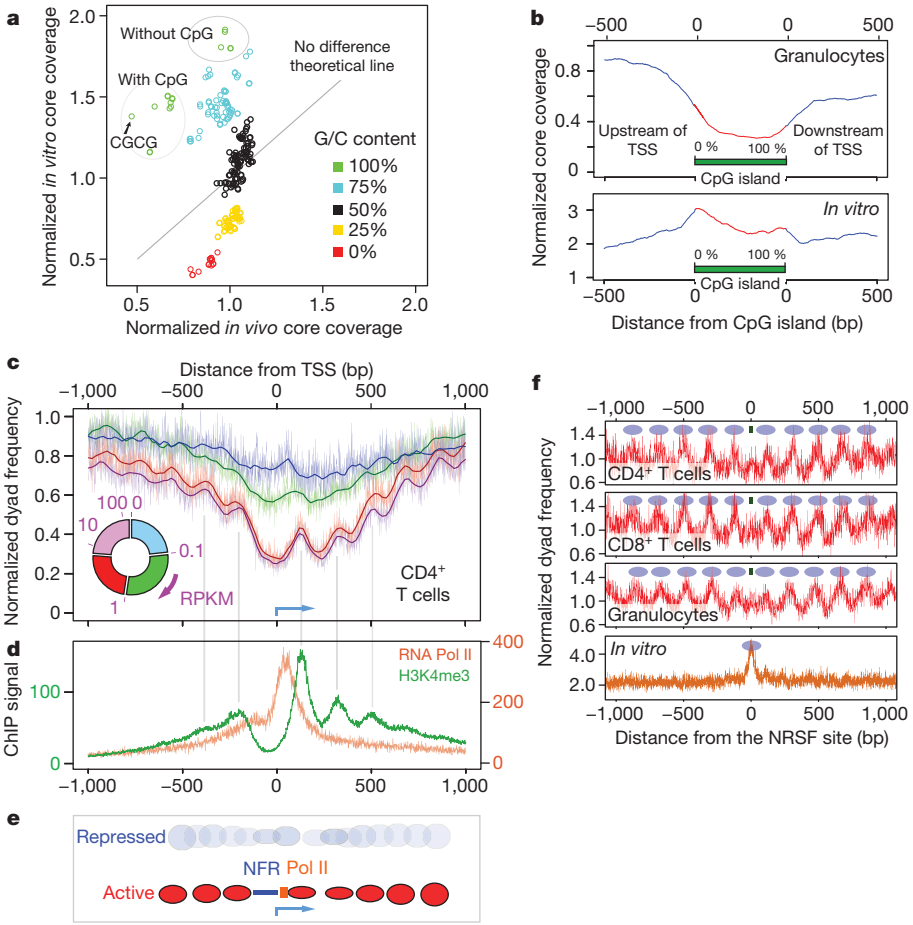


Figure 4 | Influence of gene regulatory function on nucleosome positioning. **a**, Comparison of sequence preferences of nucleosomes *in vivo* and *in vitro*. Normalized nucleosome core coverage *in vivo* (granulocytes) for a given sequence 4-mer is plotted along the *x*-axis. *In vitro* core coverage is plotted along the *y*-axis. Each data point on the plot represents one of the 256 possible 4-mers (coloured according to their G/C content). The diagonal line depicts the positions in the plot for which sequence-based preferences of nucleosomes would be the same *in vivo* and *in vitro*. **b**, Nucleosome core coverage over CpG islands *in vivo* and *in vitro*. *x*-axis represents coordinates within CpG islands (0–100%) and flanking upstream of the transcriptional start sites (TSS) (left) and downstream of the TSS (right). Normalized frequencies of nucleosome cores *in vivo* (upper plot) and *in vitro* (lower plot) are plotted along the *y*-axis. **c**, *In vivo* $CD4^+$ T-cell nucleosome organization around promoters. *x*-axis represents distance from the TSS (blue arrow). Normalized frequencies of nucleosome dyads are plotted along the *y*-axis. Nucleosome arrangements within four gene groups are shown (not expressed 0–0.1 RPKM, low expressed 0.1–1 RPKM, moderately expressed 1–8 RPKM, highly expressed > 8 RPKM). Pie chart depicts distribution of RPKM values across gene groups. **d**, RNA Pol II binding signal within highly expressed genes (orange curve) and H3K4me3-marked nucleosome dyad frequency (green curve) within highly expressed genes (>8 RPKM). Nucleosomes show consistent positions, indicated by grey lines pointing to nucleosome centres. **e**, Schematic depiction of nucleosome organization around promoters of repressed and active genes. Promoters of repressed genes do not have a well-defined nucleosome organization, whereas promoters of active genes have a nucleosome-free region (NFR, blue), RNA Pol II (orange) localized at the NFR boundary, and positioned nucleosomes (red) radiating from the NFR. Height of the ovals represents nucleosome frequency (inferred from **c**). **f**, Nucleosome distribution around the top 1,000 NRSF sites *in vivo* and *in vitro*. Distances from the NRSF binding sites are plotted along the *x*-axis. *y*-axis represents the normalized frequency of nucleosome dyads. Blue ovals depict hypothetical nucleosome positions. NRSF binding site is shown by the green rectangle.

37 bp and CTCF 74 bp. In agreement with sequence-based predictions[21], both CTCF and NRSF sites intrinsically encode high nucleosome occupancy as can be seen from the *in vitro* data (Fig. 4f and Supplementary Fig. 9), but this signal is overridden *in vivo* by occlusion of these sites from associating with nucleosomes. Additionally, phasing of nucleosomes around these regulatory sites is more compact in granulocytes compared to T cells (Supplementary Fig. 9), again exemplifying the importance of cellular parameters for placement of nucleosomes.

Our genome-wide, deep sequence data of nucleosome positions facilitated an initial characterization of the determinants of nucleosome organization in primary human cells. Spacing of nucleosomes differs between cell types and between distinct epigenetic domains in the same cell type, and is influenced by transcriptional activity. We confirm positioning preferences in regulatory elements such as promoters and chromatin regulator binding sites, but find that the majority of the human genome exhibits little if any detectable positioning. The influence of sequence on positioning of nucleosomes *in vivo* is modest but detectable. Despite DNA sequence being a potent driver of nucleosome organization at certain sites, the cellular environment often overrides sequence signals and can drive nucleosomes to occupy intrinsically unfavourable DNA elements or evict nucleosomes from intrinsically favourable sites. We find evidence for the barrier model for nucleosome organization, and that barriers can be nucleosomes (positioned by container sites), RNA polymerase II (stalled at the promoter), or sequence-specific regulatory factors. Our nucleosome maps should be useful for investigating how nucleosome organization affects gene regulation and vice versa, as well as for pinpointing the mechanisms driving regional heterogeneity of nucleosome spacing.

METHODS SUMMARY

Neutrophil granulocytes, $CD4^+$ and $CD8^+$ T cells were isolated from donor blood using Histopaque density gradients and Ig-coupled beads against blood cell surface makers (pan T and $CD4^+$ microbeads, Miltenyi Biotec). Nucleosome cores

were prepared as described previously[7]; cells were snap-frozen and crushed to release chromatin, followed by micrococcal nuclease treatment. *In vitro* nucleosomes were prepared by combining human genomic DNA with recombinantly-derived histone octamers at an average ratio of 1 octamer per 850 bp. Unbound DNA was then digested using micrococcal nuclease. After digestion, reactions were stopped with EDTA, samples were treated with proteinase K, and nucleosome-bound DNA was extracted with phenol-chloroform and precipitated with ethanol (Supplementary methods). Purified DNA was size-selected (120–180 bp) on agarose to obtain mononucleosome cores, followed by sequencing library construction. RNA was isolated by homogenizing purified cells in TRIzol, poly-A RNA was purified using a Qiagen Oligotex kit and RNA-seq libraries were constructed using a SOLiD Whole Transcriptome Analysis kit. All sequence data was obtained using the SOLiD 35 bp protocol and aligned using the SOLiD pipeline against the human hg18 reference genome. Downstream analyses were all conducted using custom scripts (Methods).

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

Received 11 August 2010; accepted 18 March 2011.
Published online 22 May; corrected 23 June 2011 (see full-text HTML version for details).

1. Mellor, J. The dynamics of chromatin remodeling at promoters. *Mol. Cell* **19,** 147–157 (2005).
2. Radman-Livaja, M. & Rando, O. J. Nucleosome positioning: how is it established, and why does it matter? *Dev. Biol.* **339,** 258–266 (2010).
3. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458,** 362–366 (2009).
4. Berstein, B. E., Liu, C. L., Humphrey, E. L., Perlstein, E. O. & Schreiber, S. L. Global nucleosome occupancy in yeast. *Genome Biol.* **5,** R62 (2004).
5. Yuan, G.-C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309,** 626–630 (2005).
6. Johnson, S. M., Tan, F. J., McCullough, H. L., Riordan, D. P. & Fire, A. Z. Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin. *Genome Res.* **16,** 1505–1516 (2006).
7. Valouev, A. *et al.* A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.* **18,** 1051–1063 (2008).
8. Schones, D. E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132,** 887–898 (2008).
9. Trifonov, E. N. & Sussman, J. L. The pitch of chromatin DNA is reflected in it its nucleotide sequence. *Proc. Natl Acad. Sci. USA* **77,** 3816–3820 (1980).
10. Kornberg, R. D. Structure of chromatin. *Ann. Rev. Biochem.* **46,** 931–954 (1977).
11. Widom, J. A relationship between the helical twist of DNA and the ordered positioning of nucleosomes in all eukaryotic cells. *Proc. Natl Acad. Sci. USA* **89,** 1095–1099 (1992).
12. Schlegel, R. A., Haye, K. R., Litwack, A. H. & Phelps, B. M. Nucleosome repeat lengths in the definitive erythroid series of the adult chicken. *Biochim. Biophys. Acta* **606,** 316–330 (1980).
13. Fan, Y. *et al.* Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* **29,** 1199–1212 (2005).
14. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5,** 621–628 (2008).
15. Valouev, A. *et al.* Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nature Methods* **5,** 829–834 (2008).
16. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823–837 (2007).
17. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genet.* **40,** 897–903 (2008).
18. Satchwell, S. C., Drew, H. R. & Travers, A. A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191,** 659–675 (1986).
19. Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442,** 772–778 (2006).
20. Hughes, A. & Rando, O. J. Chromatin 'programming' by sequence - is there more to the nucleosome code than %GC? *J. Biol.* **8,** 96 (2009).
21. Tillo, D. *et al.* High nucleosome occupancy is encoded at human regulatory sequences. *PLoS ONE* **5,** e9129 (2010).
22. Mavrich, T. N. *et al.* A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* **18,** 1073–1083 (2008).
23. Mavrich, T. N. *et al.* Nucleosome organization in the *Drosophila* genome. *Nature* **453,** 358–362 (2008).
24. Lee, W. *et al.* A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genet.* **39,** 1235–1244 (2007).
25. Gu, S. G. & Fire, A. Partitioning the *C. elegans* genome by nucleosome modification, occupancy, and positioning. *Chromosoma* **119,** 73–87 (2010).
26. Sasaki, S. *et al.* Chromatin-associated periodicity in genetic variation downstream of transcriptional start sites. *Science* **323,** 401–404 (2009).
27. Zhang, Y. *et al.* Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions *in vivo*. *Nature Struct. Mol. Biol.* **16,** 847–852 (2009).
28. Field, Y. *et al.* Gene expression divergence in yeast is coupled to evolution of DNA-encoded nucleosome organization. *Nature Genet.* **41,** 438–445 (2009).
29. Chuddapah, S. *et al.* Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res.* **19,** 24–32 (2009).
30. Fu, Y., Sinha, M., Peterson, C. L. & Weng, Z. The insulator binding protein CTCF positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet.* **4,** e1000138 (2008).
31. Albert, I. *et al.* Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* **446,** 572–576 (2007).
32. Wellinger, R. E. & Thoma, F. Nucleosome structure and positioning modulate nucleotide excision repair in the non-transcribed strand of an active gene. *EMBO J.* **16,** 5046–5056 (1997).
33. Sha, K. *et al.* Distributed probing of chromatin structure *in vivo* reveals pervasive chromatin accessibility for expressed and non-expressed genes during tissue differentiation in *C. elegans*. *BMC Genomics* **11,** 465 (2010).

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

Acknowledgements This work was supported by the Stanford Genetics/Pathology Sequencing Initiative. We thank G. Narlikar for help with *in vitro* experiments, Life Technologies, especially J. Briggs, for help with generating sequencing data, P. Lacroute for help with sequence alignment, S. Galli for valuable discussions, L. Gracey for critical reading of the manuscript, and members of the Sidow and Fire labs for valuable feedback and discussions. Work in the Fire lab was partially supported by NIGMS (R01GM37706). A.V. was partially supported by an ENCODE subcontract to A.S. (NHGRI U01HG004695). S.M.J. was partially supported by the Stanford Genome Training program (NHGRI T32HG00044).

Author Contributions A.V., S.M.J., A.S. and A.Z.F. designed the experiments. S.M.J., A.V., C.L.S. and S.D.B. performed the experiments. A.V. designed and carried out analyses with input from A.S., A.Z.F. and S.M.J.; A.V., A.S. and A.Z.F. wrote the manuscript.

Author Information All sequence data were submitted to Sequence Read Archive (accession number GSE25133). Sites containing strongly positioned *in vitro* nucleosomes are available as a supplementary data file. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.S. (arend@stanford.edu) or A.Z.F. (afire@stanford.edu).

METHODS

Cell purification. Blood samples were obtained from the Stanford Blood Center. Samples were screened for any medical history of malignancy or signs of infectious disease, and tested for serologic evidence of viral infections to ensure that samples came from healthy donors. The Stanford Blood Center procedures used for the cells in this study are the same as those used for transfusion of patients and are routinely inspected by the FDA, the American Association of Blood Banks, and the College of American Pathologists. The blood for the experiments was processed immediately upon donation to avoid any change in quality as a result of sample storage.

Buffy coat (36 ml) from a blood donor was diluted in PBS to a total volume of 200 ml. The cells were layered on a Histopaque gradient with densities 1.119 and 1.077 g ml^{-1} according to manufacturer's instructions (Sigma HISTOPAQUE-1119 and 1077) and separated by centrifugation to yield granulocytes and mononuclear fractions. T cells were isolated from mononuclear cells using a Pan T isolation Kit (Miltenyi Biotec), followed by separation into CD4$^+$ and CD8$^+$ fractions using CD4$^+$ microbeads (Miltenyi Biotec).

Isolation of mononucleosome core DNA fragments from human cells. To isolate mononucleosome core DNA from human cells, neutrophil granulocytes, CD4$^+$ lymphocytes and CD8$^+$ lymphocytes were flash-frozen in liquid nitrogen in 0.34 M sucrose Buffer A and ground, digested on different days, and isolated as described in ref 7. By carrying out an MNase digestion in a short time frame (12 min at 16 °C) following grinding of the samples, we minimize the potential for nucleosome mobility. To maximize uniformity of representation, we use an extraction protocol after MNase digestion that does not rely on solubility of the individual core particles; this resulted in recovery of the bulk of input DNA as a mono-nucleosome band (Supplementary Fig. 10), limiting the degree to which the protocol might select for specific (for example, accessible) chromosomal regions. The mean nucleosome core length obtained for analysis (153 nucleotides) indicates an average overhang of 3 nucleotides on each side of individual cores (147 bp + 2 × 3 bp = 153 bp). Subsequent analyses assign nucleosome positions accounting for this mean overhang and making use of the ability to define location based on interpolation between values calculated from plus-oriented and minus-oriented reads (see below).

Preparation of *in vitro* nucleosomes. Naked genomic DNA isolated from neutrophil granulocytes from our *in vivo* studies was sheared by sonication using a Covaris sonicator and separated on a 1% UltraPure Agarose (Invitrogen) gel run at 100 V for 1 h. A smear of fragments with lengths from 850–2,000 bp (the bulk of the sheared DNA) was isolated and extracted from the gel using the QIAquick Gel Extraction Kit (Qiagen). DNA fragment lengths several-fold larger than nucleosome cores were chosen for this analysis to minimize any end-effects that could have contributed an end-based signal at shorter fragment sizes. Lack of end-preference in the reconstitutions was then confirmed under the conditions of these assays using a series of defined restriction fragments as templates for assembly (S.M.J. and A.F., results not shown).

The ends of the sheared DNA fragments were repaired as described below and then were assembled with recombinant *Xenopus* histones into nucleosomes as described previously[34] at a 1.1:1 molar ratio of DNA to histone octamer such that on average one nucleosome would occupy 850 bp of DNA. Specifically, 4.9 μg of DNA and 0.80 μg of octamer were reconstituted in a total volume of 200 μl.

The ref. 34 conditions (in which DNA was not limiting) were used for our analysis in order to focus specifically on primary sequence effects on nucleosome position. We note that two recent studies in yeast use somewhat different conditions, with a higher ratio of nucleosomes to DNA[3,27]. Assays at high nucleosome:DNA ratio provide a composite readout reflecting both (1) primary preferences of nucleosomes (caused by sequence signals within the nucleosome-bound DNA) and (2) secondary effects due to steric hindrance as a result of dense packing of nucleosomes. Although such data are certainly valuable in modelling chromosome dynamics, the goals of our study (definition of individual sequence elements that can initiate positioning) were best served with the lower nucleosome:DNA assay conditions[34].

Isolation of *in vitro* nucleosome core DNA fragments. *In vitro* nucleosome core DNAs were isolated by diluting 70 μl of the reconstituted *in vitro* nucleosome into a total volume of 200 μl containing 5 mM $MgCl_2$, 5 mM $CaCl_2$, 70 mM KCl and 10 mM Hepes at pH 7.9 (final concentrations) and digesting with 20 units of micrococcal nuclease (Roche) resuspended at 1 U μl^{-1} for 15 min at room temperature. The digestion was stopped by adding an equal volume of 3% SDS, 100 mM EDTA and 50 mM Tris. Octamer proteins were removed by treating with one-tenth volume proteinase K (20 mg ml^{-1} in TE at pH 7.4) for 30 min at 50 °C followed by phenol/chloroform and chloroform extractions and ethanol precipitation. This procedure was repeated twice to process the entire *in vitro* sample, and then *in vitro* DNA cores were isolated on a 2% UltraPure Agarose (Invitrogen) gel run at 100 V for 1 h followed by DNA extraction from the gel using a QIAquick Gel Extraction Kit (Qiagen) following the standard protocol with the exception of allowing the isolated gel sample to incubate in Buffer QG at room temperature until dissolved.

Genomic MNase digest control library preparation. For control libraries, genomic DNA (20 μg) from human neutrophil granulocytes in 0.34 M sucrose Buffer A with 1× BSA (New England Biolabs) and 1 mM $CaCl_2$ was digested with 200 units of micrococcal nuclease (Roche) (0.4 U μl^{-1} final concentration) in a total volume of 500 μl for 10 min at 23 °C. The digestion was stopped by addition of 10 μl 0.5 M EDTA, followed by ethanol precipitation. The digested DNA was run on a 2.5% agarose gel and the smear of DNA fragments from 135–225 bp was excised from the gel and purified using a QIAquick Gel Extraction Kit (Qiagen) as noted above.

End repair, linker ligation and library amplification. The ends of isolated mononucleosome core DNAs (granulocytes, CD4$^+$ lymphocytes and CD8$^+$ lymphocytes), *in vitro* core DNAs and genomic control DNAs were processed by treating 0.3–0.5 μg of the DNA samples with T4 polynucleotide kinase (New England Biolabs) at 37 °C for 2.5 h followed by ethanol precipitation and subsequent treatment with T4 DNA polymerase (New England Biolabs) in the presence of dNTPs for 15 min at 12 °C. After purification using either a QIAquick Gel Extraction Kit as described above or a QIAquick PCR Purification Kit (Qiagen), linkering of previously annealed duplexes AF-SJ-47 (5′-OH-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT-3′)/AF-SJ-48 (5′-P-ATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGGTT-3′) and AF-SJ-49 (5′-OH-CTGCCCCGGGTTCCTCATTCTCT-3′)/AF-SJ-50 (5′-P-AGAGAATGAGGAACCCGGGGCAGTT-3′) to the samples was accomplished with T4 DNA ligase during a 6.5-h room-temperature incubation. The ligation reactions were separated on a 2% agarose gel, and the relevant band isolated as described above. Amplification of the linkered libraries was accomplished with 8 (granulocyte mononucleosome library), 10 (CD4$^+$ lymphocytes, CD8$^+$ lymphocytes and genomic control libraries) or 12 (*in vitro* library) cycles of polymerase chair reaction (PCR) using primers AF-SJ-47 (SOLiD P1 primer) and AF-SJ-49 (SOLiD P2 primer) with subsequent separation and purification using a 2% agarose gel and the QIAquick Gel Extraction Kit as described above. The number of cycles used in the PCR amplification were monitored and selected as described in ref. 25.

RNA-seq library preparation. Cells were homogenized in TRIzol using an 18G needle, followed by total RNA extraction using phenol-chloroform-isoamyl alcohol. Poly-A RNA was isolated from total RNA using a Qiagen Oligotex kit according to the manufacturer's instructions. The RNA-seq SOLiD sequencing library was built from 100 ng of poly-A RNA according to the manufacturer's instructions (SOLiD whole transcriptome analysis kit).

DNA sequencing and mapping. Both nucleosome fragment and RNA-seq libraries were sequenced using the SOLiD DNA sequencing platform to produce 35 bp reads. All sequence data was mapped using SOLiD software pipeline against the human hg18 assembly using the first 25 bp from each read. This was done to maximize the number of the reference-mapped reads, as the higher error rate in read positions 26–35 of that version of the SOLiD chemistry prevented a substantial fraction of reads from mapping to the genome. For the genome-wide analysis we retained only unambiguously mapped reads.

Genome coverage by nucleosome cores was calculated as: core coverage = (number of mapped reads) × (147)/(genome size)

mRNA sequencing and data analysis. RNA-seq libraries were sequenced on the SOLiD platform to produce 35 bp reads and then the first 25 bp of each read were mapped to hg18 using the SOLiD mapping pipeline which resulted between 77 and 99 million mapped reads for each cell type. RPKM values were calculated as in ref. 14, with a modification that adjusted for transcript length, which was calculated according to the formula $L' = L - 50 \times (E - 1)$, where L is the actual transcript length, and E is the number of exons in the gene. This modification is needed because of the lack of mappings across splice junctions.

Mathematical notations. Start counts: $S_{+/-}(j)$ represent counts of 5′ coordinates of reads that map in + or – orientation at the j-th position of the reference strands. For example, if read maps to the interval $[x,y)$ on the + strand, then its 5′ coordinate is x, if it maps to – strand, then it's $y - 1$.

Indicator functions: I(condition) = 1 if condition is satisfied, 0 otherwise.

Nucleosome positioning stringency metric: nucleosome positioning stringency metric quantifies the fraction of nucleosomes covering a given position that are 'well positioned'. The stringency at position i of the genome is calculated according to the formula:

$$S(i, w=30) = \frac{D(i, w=30)}{\sum_{i-150}^{i+150} \frac{1.09}{w} D(j, w=30)},$$

where $D(i,w)$ is a kernel-smoothed dyad count calculated according to the formula:

$$D(i,w) = \sum_{j=0}^{L} K(i-j,w) \cdot d(j),$$

where L is the size of a given chromosome, and $K(u,w)$ is a smoothing kernel function of the form:

$$K(u,w) = (1-(u/w)^2)^3 I\{|u| < w\},$$

and

$$\int_{-1}^{1} (1-u^2)^3 du = 1/1.09,$$

and $d(j)$ represents the number of dyads that occurs at the position j:

$$d(j) = s_+(j-l/2) + s_-(j+l/2).$$

Here l is the average library size ($l = 153$ for *in vivo* data sets, 147 for *in vitro* data set). The core size is inferred from the 3-pile distogram peak position in the range of 100–200 bp.

The numerator of the stringency formula represents a kernel-smoothed count of nucleosome centres (dyads) at position i in the genome, whereas the denominator represents the count of nucleosome centres that infringe on the nucleosome centred at that position, which is inferred by integration of the dyad density estimate over an area of nucleosome infringement. The stringency is constructed in such a way that it would achieve a maximum of 1 if all nucleosomes were perfectly centred at that position (Supplementary Fig. 4). If two alternative, mutually exclusive, equally frequent nucleosome positions are observed in the data, then the stringency would be 0.5 or 50% for each alternative site (illustrated in Supplementary Fig. 4).

Application of the Kernel Density Estimation allowed obtaining smooth estimates of the stringency, which was useful for detection of nucleosome centres and robustly estimating the degree of positioning. We experimented with other smooth kernels and obtained highly consistent results. In principle, the kernel choice should not affect the results substantially as long as there is sufficient nucleosome core coverage (which follows from the convergence property of Kernel Density Estimation).

The kernel bandwidth w is an important parameter of the stringency formula and provides a means to control the smoothness of the stringency profile. Larger values of w provide higher smoothing but result in less accurate estimates of positioning centres, which is acceptable in cases of low core coverage. On the other hand, lower values of w result in less smoothing but more accurate estimation of the positioning centres, which is desirable in cases when nucleosome core coverage is high. We decided to use $w = 30$ in our calculation as it provided a sufficient amount smoothing across all of our data sets without sacrificing the sharpness of the positioning estimate.

Nucleosome positioning stringency was used for calculation of the fraction of the genome containing preferentially positioned nucleosomes (Supplementary Fig. 5). Positioned nucleosomes used in the container site analysis (Fig. 3a–c) were identified with the positioning stringency metric (as shown) and additional filters on nucleosome occupancy (*in vitro* occupancy > 30) to improve the statistical confidence of the positioning estimates.

Nucleosome dyad coordinates. Nucleosome dyads were inferred from 5′ coordinates of reads by shifting them by half the average nucleosome core size towards the 3′ end. The average nucleosome cores size was estimated by a maximum value of the 3-pile distogram in a size range of 100–200 bp.

Rotational positioning analysis. We examined oligonucleotide preferences of rotational positioning of nucleosomes, which is associated with 10-bp patterning of short k-mers within nucleosome cores[18,31]. Plotting the frequencies of dyads around specific oligomers within the genome showed that the strongest patterning was exhibited by C-polymers (CC,CCC) with an exact helical period of 10.15 bp (Supplementary Fig. 7a, P-value $< 2 \times 10^{-16}$), indicating that they are important for rotational positioning. *In vivo*, such rotational preferences are much less pronounced (Supplementary Fig. 7b), indicating that cellular factors or conditions often override the sequence-encoded rotational settings.

Characterization MNase cleavage patterns. MNase is known to have sequence preferences that can affect both individual and bulk analyses of chromatin structure. Previous studies comparing MNase with alternative probes in model systems, both at specific loci (for example, ref. 32) and genome wide (for example, ref. 33), support the correspondence between the patterns of nucleosomes inferred from MNase digestion of chromatin and the *in vivo* chromatin landscapes. Nonetheless, it remained important to characterize the patterns of MNase activity in our data.

We investigated the extent of cleavage bias by MNase by examining sequence preferences within the cleavage sites, which correspond to 5′ end read positions in our data (Supplementary Figure 7a–e). Consistent with previous observations, MNase exhibits a pronounced but imperfect tendency to cleave at A or T nucleotides in naked DNA (Supplementary Fig. 11a). This same bias is detectable but, importantly, weaker when nucleosomes occupy the DNA, both *in vivo* and *in vitro* (1-pile subsets, top row b–e). Sites of more frequent cleavage (3-pile and 5-pile subsets, middle and bottom rows) revealed preferences that were virtually indistinguishable from the single-site preference.

The fact that the cleavage bias does not extend beyond 1–2 base pairs suggests that our analyses of nucleosome positioning preferences, which have substantially less than single-base resolution, should be robust to biases introduced by the MNase digestion. A case in point is the above-discussed rotational positioning analysis, whose resolution is on the order of 10 bp and which involves oligonucleotides that do not resemble the MNase cleavage site (Supplementary Fig. 7a).

To investigate whether the sequence-driven nucleosome positioning element identified by the *in vitro* reconstitution experiment (Fig. 3) was a result of particularly pronounced MNase digestion bias within specific sites, we examined nucleotide preferences of nucleosome fragments overlapping sites of medium (>0.5) and high (>0.7) positioning stringency (Supplementary Fig. 11f, g). Preferences within these sites are identical to genome-wide preferences, ruling out the possibility that their positioning is an artefact of MNase digestion. In addition, we observe wave-like patterns *in vivo* around these sites (Fig. 3e) consistent with existence of a chromatin barrier in the form of a well-positioned nucleosome.

The lack of systematic differences in cleavage bias in our experimental data sets, in conjunction with the fact that naked DNA is affected most by the cleavage bias, suggests that our conclusions are robust to the use of MNase.

Analyses of independent data sets. We conducted additional analyses on independent data not generated by us to address any lingering concerns about biases or reproducibility. First, we sought to confirm independently that MNase cuts the linker DNA separating nucleosomes. In our data, CTCF sites (Supplementary Fig. 9) are surrounded by arrays of highly positioned and phased nucleosomes extending at least 1 kb in each direction. We investigated the frequency of cleavage by DNase I, a nuclease with preferences different from those of MNase, around CTCF sites within lymphoblastoid cell lines, using publicly available data from the ENCODE project. In agreement with our MNase results, we observed strongly phased peaks in the DNase I ENCODE data that align with linker DNA sites in our nucleosome data (Supplementary Fig. 12).

The estimates of spacing between nucleosomes as depicted in Fig. 1d are consistent between the two types of T cells we analysed. To ask whether these estimates were also reproducible by a different approach, we turned to a published data set that was generated for a different purpose, and by different means. Ref. 8 compared nucleosome distribution between resting and activated $CD4^+$ T cells using MNase treatment of the cellular chromatin. We analysed spacing of nucleosomes in their data and obtained a highly concordant estimate of 202 and 203 bp (Supplementary Fig. 13) which is in agreement with the 203 bp spacing we see in our data (Fig. 1d).

34. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Preparation of nucleosome core particle from recombinant histones. *Methods Enzymol.* **304,** 3–19 (1999).

LETTER

doi:10.1038/nature10136

Agonist-bound adenosine A_{2A} receptor structures reveal common features of GPCR activation

Guillaume Lebon[1], Tony Warne[1], Patricia C. Edwards[1], Kirstie Bennett[2], Christopher J. Langmead[2], Andrew G. W. Leslie[1] & Christopher G. Tate[1]

Adenosine receptors and β-adrenoceptors are G-protein-coupled receptors (GPCRs) that activate intracellular G proteins on binding the agonists adenosine[1] or noradrenaline[2], respectively. GPCRs have similar structures consisting of seven transmembrane helices that contain well-conserved sequence motifs, indicating that they are probably activated by a common mechanism[3,4]. Recent structures of β-adrenoceptors highlight residues in transmembrane region 5 that initially bind specifically to agonists rather than to antagonists, indicating that these residues have an important role in agonist-induced activation of receptors[5–7]. Here we present two crystal structures of the thermostabilized human adenosine A_{2A} receptor (A_{2A}R-GL31) bound to its endogenous agonist adenosine and the synthetic agonist NECA. The structures represent an intermediate conformation between the inactive and active states, because they share all the features of GPCRs that are thought to be in a fully activated state, except that the cytoplasmic end of transmembrane helix 6 partially occludes the G-protein-binding site. The adenine substituent of the agonists binds in a similar fashion to the chemically related region of the inverse agonist ZM241385 (ref. 8). Both agonists contain a ribose group, not found in ZM241385, which extends deep into the ligand-binding pocket where it makes polar interactions with conserved residues in H7 (Ser $277^{7.42}$ and His $278^{7.43}$; superscripts refer to Ballesteros–Weinstein numbering[9]) and non-polar interactions with residues in H3. In contrast, the inverse agonist ZM241385 does not interact with any of these residues and comparison with the agonist-bound structures indicates that ZM241385 sterically prevents the conformational change in H5 and therefore it acts as an inverse agonist. Comparison of the agonist-bound structures of A_{2A}R with the agonist-bound structures of β-adrenoceptors indicates that the contraction of the ligand-binding pocket caused by the inward motion of helices 3, 5 and 7 may be a common feature in the activation of all GPCRs.

In the simplest model for the conformational dynamics of GPCRs[10] there is an equilibrium between two states, R and R*. The inactive state R preferentially binds inverse agonists and the activated state R* preferentially binds agonists[11]. Only R* can couple and activate G proteins. Although there are far more complex schemes[12] describing intermediates between R and R*, studies on rhodopsin have indicated that there is only one major conformational change that significantly alters the structure of the receptor[3]. Thus the structures of dark-state rhodopsin[13,14] and of opsin[15,16] are considered to be representative structures for the R and R* state, respectively. Structures of six different GPCRs[8,13,17–21] in conformations closely approximating the R state have now been determined and it is clear that they are similar to each other, with root mean squared deviation (r.m.s.d.) between any pair of structures in the transmembrane domains being less than 3 Å. As observed in light activation of rhodopsin, the major structural difference between R and R* is the movement of the cytoplasmic ends of helices 5 and 6 away from the receptor core by 5–6 Å, opening up a cleft in the centre of the helix bundle where the carboxy terminus of a G protein can bind[16]. Recently, the structure of an agonist-bound β-adrenoceptor (β_2-AR) was determined in complex with an antibody fragment (nanobody Nb80)[5]. This structure of β_2-AR is very similar to the structure of opsin, which indicates that the nanobody mimicked the action of a G protein by maintaining the receptor structure in an activated state. Given the structural similarities between opsin and the β_2-AR–Nb80 complex, it is likely that the structures of the R* states of other GPCRs are also highly similar. This is consistent with the same heterotrimeric G proteins being able to couple to multiple different receptors[22]. However, it is unclear whether conserved structures of R and R* indicate that all agonists activate the receptors in an identical fashion. The recent structures of a thermostabilized β_1-AR bound to four different agonists indicated that a defining feature of agonist binding to this receptor is the formation of a hydrogen bond with $Ser^{5.46}$ on transmembrane helix 5 that accompanies the contraction of the ligand-binding pocket[7]. Here we describe two structures of the adenosine A_{2A} receptor (A_{2A}R) bound to two different agonists, which indicates that the initial action of agonist binding to A_{2A}R has both similarities and differences compared to agonist binding in β-ARs.

The native human A_{2A}R when bound to its endogenous agonist adenosine or to the high-affinity synthetic agonist NECA is unstable in detergent, so crystallization and structure determination relied on using a thermostabilized construct (A_{2A}R-GL31) that contained four point mutations, which markedly improved its thermostability. Pharmacological analysis showed that the mutant receptor bound the five antagonists tested with greatly reduced affinity (1.8–4.3 log units), whereas four agonists bound with similar affinity to the wild-type receptor (Supplementary Fig. 1). However, A_{2A}R-GL31 is only weakly activated by the agonist CGS21680 (Supplementary Fig. 2), which indicates that the thermostabilizing mutations might also decouple high-affinity agonist binding from the formation of R*. The conformation of A_{2A}R-GL31 is not consistent with it being in the fully activated G-protein-coupled state, because we do not observe a 42-fold increase in affinity for NECA binding measured for $G_{\alpha s}$-coupled A_{2A}R (ref. 23). These data all indicate that A_{2A}R-GL31 is in an intermediate conformation between R and R*, which is consistent with the structural analysis presented later.

The two structures we have determined are of A_{2A}R-GL31 bound to adenosine and NECA with resolutions of 3.0 Å and 2.6 Å, respectively (Supplementary Table 1). Global alignments of the A_{2A}R-GL31 structures with A_{2A}-T4L (A_{2A}R with T4 lysozyme inserted into inner loop 3) bound to the inverse agonist ZM241385 were performed based on those residues in the region of the ligand-binding pocket that show the closest structural homology (Fig. 1 and Supplementary Text). This gave an r.m.s.d. in Cα positions of 0.66 Å for the 96 atoms selected, which include all residues involved in binding either adenosine or NECA, with the exception of those in H3. Using this transformation, the adenine-like moiety of the two ligands superimposes almost exactly (r.m.s.d. 0.56 Å). The most significant differences between the two structures are seen in a distortion and a 2 Å shift primarily along the helical axis of H3, a bulge in H5 (resulting from non-helical backbone conformation angles of residues Cys 185 and Val 186) that

[1]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK. [2]Heptares Therapeutics, BioPark, Broadwater Road, Welwyn Garden City AL7 3AX, UK.

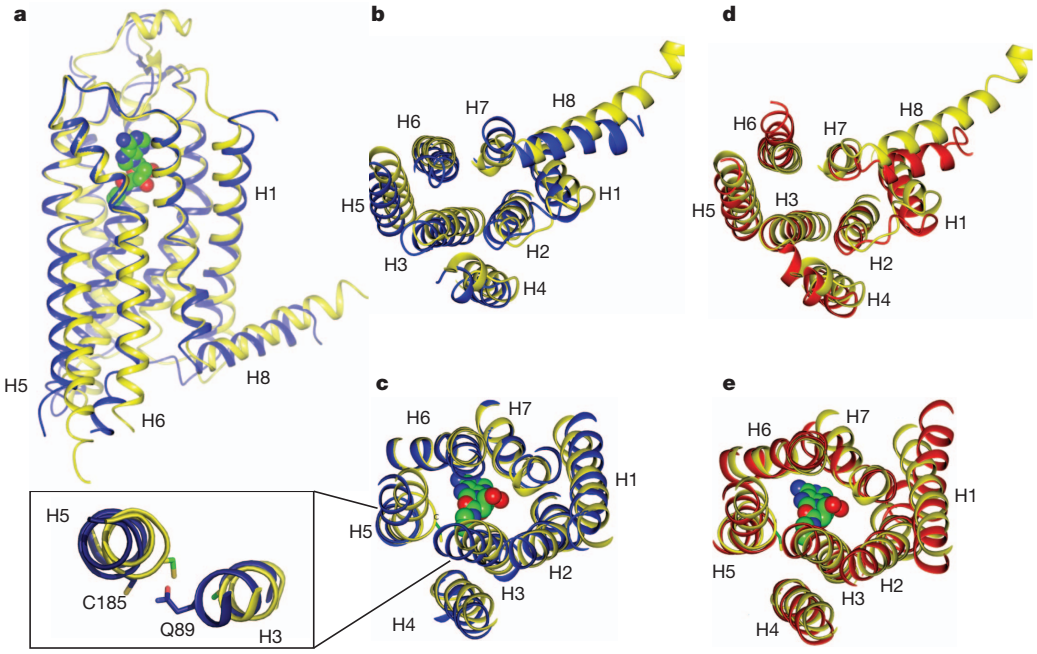


Figure 1 | Structure of the adenosine A_{2A} receptor bound to NECA compared to other GPCR structures. **a**, The structure of NECA-bound $A_{2A}R$ is shown as a cartoon (yellow) aligned with the structure of A_{2A}-T4L bound to the inverse agonist ZM241385 (blue; PDB code 3EML[8]). NECA is shown as a space-filling model (C, green; N, blue; O, red). **b**, **c**, Sections through the aligned receptors in **a** that highlight the differences in the intracellular face of the receptors (**b**) and in the ligand-binding pocket (**c**), with the bulge in H5 shown as an inset. **d**, **e**, Alignment of NECA-bound $A_{2A}R$ (yellow) with agonist-bound β_2-AR–Nb80 (red; PDB code 3P0G[5]) showing the intracellular face of the receptors (**d**) and the ligand-binding pocket (**e**). NECA is shown as a space-filling model in **c** and **e**. The figures were generated using CCP4mg[31]. Analogous alignments to opsin are depicted in Supplementary Fig. 7.

shifts residues into the binding pocket by up to 2 Å and also a change in conformation of the cytoplasmic ends of H5, H6 and H7 (Fig. 1). Comparison of the $A_{2A}R$-GL31 structure with the agonist-bound β_2-AR–Nb80 complex indicates that these differences are similar to the conformational changes in the β_2-AR that are proposed to be responsible for the formation of the R* state[5]. However, it is unlikely that the structure of $A_{2A}R$-GL31 represents the fully activated state, because comparison with opsin bound to the C-terminal peptide of the G protein transducin shows that there is insufficient space in $A_{2A}R$-GL31 for the C terminus of the G protein to bind (Supplementary Fig. 3). This is on the basis of the assumption that all G proteins bind and activate GPCRs in a similar fashion, but given the highly conserved structures of both G proteins and GPCRs this seems a reasonable hypothesis.

The fact that the structure of $A_{2A}R$-GL31 represents an agonist-binding state is consistent with how $A_{2A}R$-GL31 was engineered. Thermostabilizing mutations were selected by heating the NECA-bound detergent-solubilised receptor, so the mutations are anticipated to stabilize the agonist-bound state either by stabilizing helix–helix interactions and/or biasing the conformational equilibrium between the agonist-bound R* state and the agonist-bound R state[24–26]. The two most thermostabilizing mutations, L48A and Q89A, are in regions of the receptor that are involved in transitions between R and R*, providing a possible explanation for their thermostabilizing effect (Supplementary Fig. 4). The other two mutations, A54L and T65A, are at the receptor–lipid interface and the reason for their thermostabilizing effect is unclear. Although the overall shape of the ligand-binding pockets of $A_{2A}R$ and β_2AR are different, the structural similarities with the β_2-AR–Nb80 (ref. 5) and the structural differences to ZM241385-bound A_{2A}-T4L[8] indicate that the structure of the binding pocket in $A_{2A}R$-GL31 is a good representation of the agonist-bound binding pocket of the wild-type receptor (Fig. 1).

Adenosine and NECA bind to $A_{2A}R$-GL31 in a virtually identical fashion; in addition, the adenine ring in the agonists interacts with $A_{2A}R$ in a similar way to the chemically related triazolotriazine ring of the inverse agonist ZM241385 (Fig. 2). Thus the hydrogen bonds between exocyclic adenosine N6 (Supplementary Fig. 5) with both Glu 169 in extracellular loop 2 (EL2) and Asn $253^{6.55}$ in H6 are similar, with the significant π-stacking interaction with Phe 168 in EL2 also conserved. One of the major structural differences between ZM241385 and the agonists is the presence of a furan substituent on C20 of triazolotriazine in the inverse agonist, whereas agonists contain a ribose substituent linked to N9 of adenine (Fig. 2 and Supplementary Fig. 5). In ZM241385, the furan group forms a hydrogen bond with Asn $253^{6.55}$ in H6 and van der Waals contacts with other residues in H3, H5 and H6 (ref. 8). In contrast, the ribose moiety in agonists forms hydrogen bonds with Ser $277^{7.42}$ and His $278^{7.43}$ in H7, in addition to van der Waals interactions with other residues in H3 and H6 (Fig. 2). In particular, Val $84^{3.32}$ has to shift its position upon agonist binding owing to a steric clash with the ribose ring, which may contribute to the 2 Å shift observed in H3 (Fig. 3). These differences in binding between ZM241385 and either adenosine or NECA indicate that the residues that bind uniquely to agonists (Ser $277^{7.42}$ and His $278^{7.43}$) have a key role in the activation of the receptor, as previously shown by mutagenesis studies[27,28]. This is analogous to the situation in the activation of β_1-AR, where only full agonists cause the rotamer conformation changes of Ser$^{5.46}$ in H5, whereas the inverse agonist ICI118551 prevents receptor activation by sterically blocking the rotamer change[7,29]. However, the details of the activation differ in that the critical residues that bind agonists and not antagonists are in H5 in the β_1-AR, but in H7 in the $A_{2A}R$ (Fig. 4).

Adenosine and NECA activate the $A_{2A}R$ through interactions with H3 and H7 that are absent in the interactions between the receptor and the inverse agonist ZM241385 (Fig. 2). The inward shift of H7, the movement of H3 and the consequent formation of a bulge in H5 are all observed in the structures of agonist-bound $A_{2A}R$-GL31 and β_2-AR–Nb80 (Fig. 1). The formation of the bulge in H5 of the β_2-AR–Nb80 structure was linked to a series of conformational changes that generate the 60° rotation of H6 about Phe $282^{6.44}$, resulting in the cytoplasmic end of H6 moving out from the receptor centre and opening the cleft where the C terminus of a G protein is predicted to bind as observed in opsin[5,6]. There are analogous side-chain movements in $A_{2A}R$-GL31

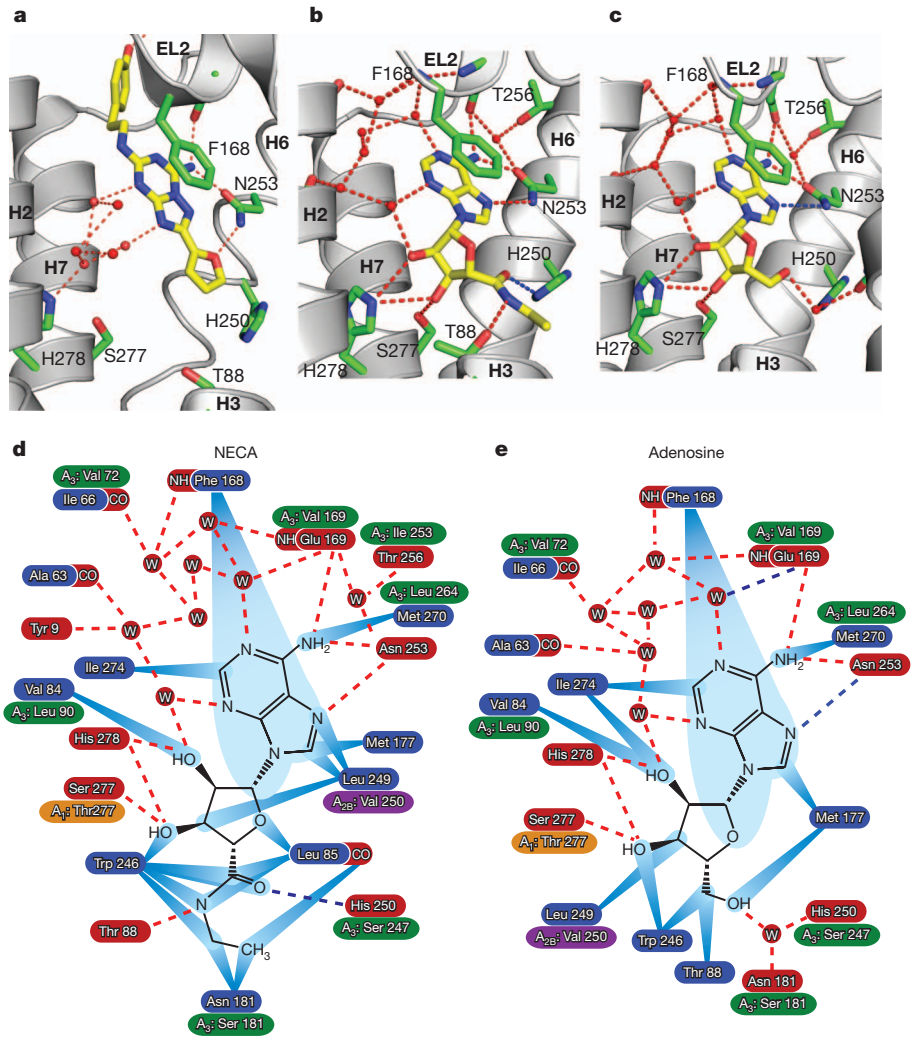


Figure 2 | Comparison of receptor–ligand interactions for $A_{2A}R$ bound to the inverse agonist ZM241385 and the agonists NECA and adenosine. **a**–**c**, Structures of the human $A_{2A}R$ in cartoon representation are shown bound to the following ligands: **a**, ZM241385 (PDB code 3EML[8]); **b**, NECA; and **c**, adenosine. **d**, **e**, Polar and non-polar interactions involved in agonist binding to $A_{2A}R$ are shown for NECA (**d**) and adenosine (**e**). Amino acid residues within 3.9 Å of the ligands are depicted, with residues highlighted in blue making van der Waals contacts (blue rays) and residues highlighted in red making potential hydrogen bonds with favourable geometry (red dashed lines, as identified by HBPLUS, see Methods) or hydrogen bonds with unfavourable geometry (blue dashed lines, donor acceptor distance more than 3.6 Å). Where the amino acid residue differs between the human $A_{2A}R$ and the human A_1R, $A_{2B}R$ and A_3R, the equivalent residue is shown highlighted in orange, purple or green, respectively. Panels **a**–**c** were generated using PyMOL (http://www.pymol.org/). Omit densities for the ligands are shown in Supplementary Fig. 6 and densities for water molecules in Supplementary Fig. 8.

that result in a 40° rotation of H6, but the cytoplasmic end of H6 remains partially occluding the G-protein-binding cleft (Supplementary Fig. 3), perhaps because the fully active conformation requires the binding of G proteins to stabilize it. Interestingly, the structure of β_2-AR[6] with a covalently bound agonist is also not in the fully activated R* conformation, which is only seen after the nanobody Nb80 is bound[5]. The importance of the bulge in H5 in the activation of $A_{2A}R$ is highlighted by how inverse agonists bind. Formation of the H5 bulge results in the inward movement of Cys $185^{5.46}$ (Cβ moves by 4 Å), which in turn causes the movement of Val 186 and ultimately a shift of His $250^{6.52}$ by 2 Å into the ligand-binding pocket, thereby sterically blocking the binding of ZM241385 (Supplementary Fig. 4). Hence, when the inverse agonist binds, it is anticipated that the H5 bulge is unlikely to form owing to the opposite series of events and hence the formation of the R* state is inhibited.

Thus, in both β-ARs and $A_{2A}R$, the formation of the H5 bulge seems to be a common action of agonists, whereas inverse agonists seem to prevent its formation. However, the energetic contributions to its formation may be different between the two receptors. In β-ARs there is a major contribution from direct interaction between the agonist and $Ser^{5.46}$, whereas in $A_{2A}Rs$, the major interaction seems to come from

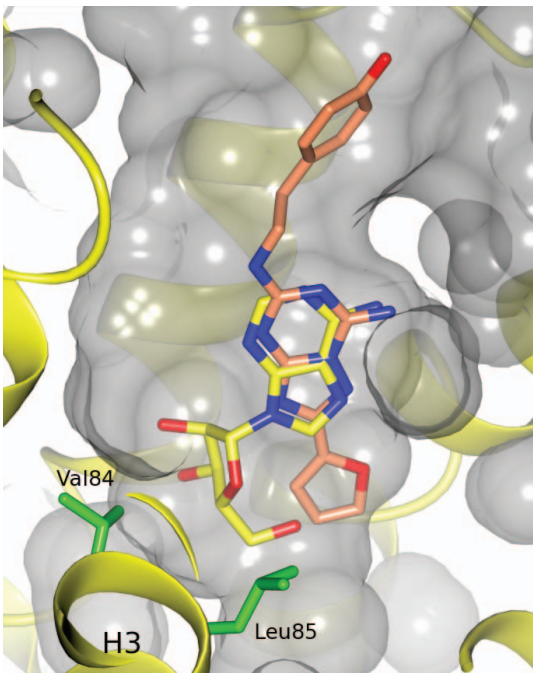


Figure 3 | Positions of adenosine and ZM241385 in the $A_{2A}R$ ligand-binding pocket. The structures of adenosine-bound $A_{2A}R$-GL31 and ZM241385-bound A_{2A}-T4L were aligned using only atoms from the protein to allow the ligand positions to be compared, with adenosine in yellow and ZM241385 in pink (N, blue; O, red). The ligands are shown in the context of the binding pocket of $A_{2A}R$-GL31, with transmembrane helices of $A_{2A}R$-GL31 shown in yellow and the surfaces of the receptor, including the cavity of the ligand binding pocket, shown in grey. The side chains of Val 84 and Leu 85 that interact with the ribose moiety of the agonist are shown in green.

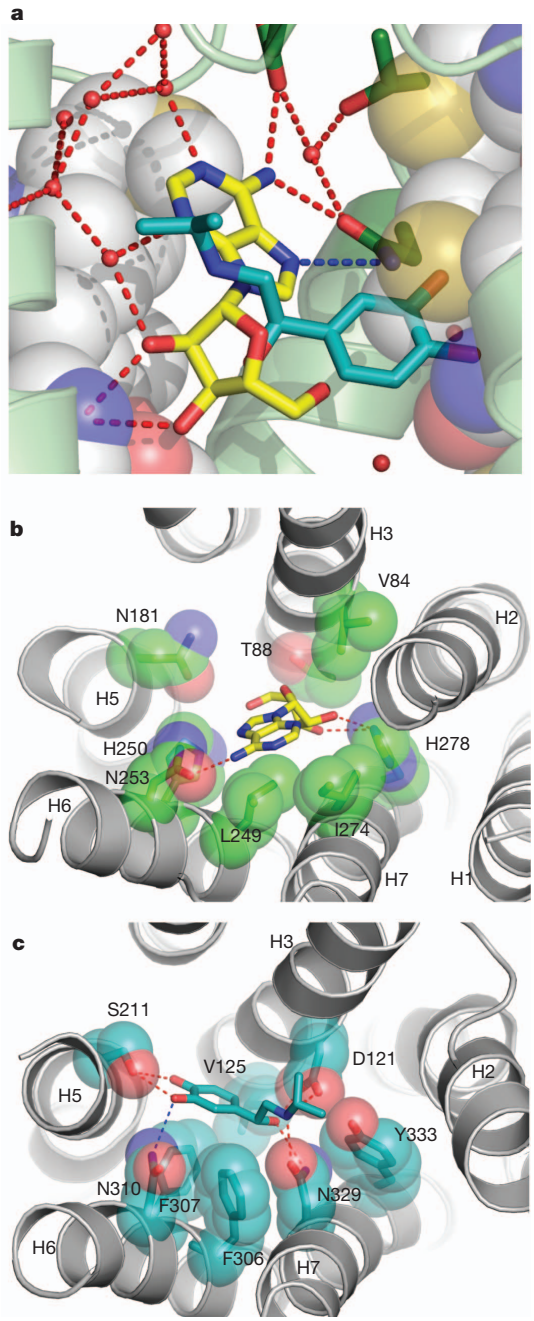


Figure 4 | Comparison of the positions of agonists in the binding pockets of the A_{2A}R and β_1-AR. a, The structures of the A_{2A}R bound to adenosine and the β_1-AR bound to isoprenaline (PDB code 2Y03)[7] were aligned by superimposing equivalent atoms in the protein structure and the positions of both ligands are shown as stick models with the carbon atoms in blue/green (isoprenaline) or yellow (adenosine); N, blue; O, red. The A_{2A}R structure is shown, with H5 and H7 as space-filling models (C, grey; N, blue; O, red) and the remainder of the structure as a cartoon (pale green). Some water molecules are shown as red spheres, hydrogen bonds as red dashed lines and the polar contacts as blue dashed lines. The orientation of the figure is identical to that shown in Fig. 2. **b**, Structure of the A_{2A}R bound to adenosine viewed from the extracellular surface. **c**, Structure of β_1-AR bound to isoprenaline (PDB code 2Y03)[7] viewed from the extracellular surface. In panels **b** and **c**, equivalent side chains in the respective structures that make contacts to both isoprenaline and adenosine in their respective receptors are shown as space-filling models and they have the following Ballesteros–Weinstein numbers (amino acid side chains are shown in parentheses for the A_{2A}R and β_1-AR, respectively): 3.32 (V84, D121); 3.36 (T88, V125); 5.42 (N181*, S211); 6.51 (L249, F306); 6.55 (N253, N310); 6.52 (H250*, F307); 7.39 (I274, N329); 7.43 (H278, Y333). An asterisk indicates residues that only make indirect contacts to the agonists via a water molecule.

interactions between the agonist and H3, combined with polar interactions involving residues in H7. Despite these differences, agonist binding to both receptors involves strong attractive non-covalent interactions that pull the extracellular ends of H3, H5 and H7 together, which is the necessary prerequisite to receptor activation.

While this manuscript was in review, a related manuscript appeared[30], describing the structure of the A_{2A}-T4L chimaera bound to the agonist UK432097, which is identical to NECA except for two large substituents on the adenine ring. The structure of UK432097-bound A_{2A}-T4L is very similar to the structures presented here in the transmembrane regions (r.m.s.d. 0.6 Å), although there are differences in the extracellular surface due to the bulky extensions of UK432097 interacting with the extracellular loops and the absence of density for residues 149–157. Xu *et al.*[30] conclude that the structure of UK432097-bound A_{2A}-T4L is in an "active state configuration", whereas we conclude that the NECA- and adenosine-bound structures are best defined as representing an intermediate state between R and R*.

METHODS SUMMARY

Expression, purification and crystallization. The thermostabilized A_{2A}R-GL31 construct contains amino acid residues 1–316 of the human A_{2A}R, four thermostabilizing point mutations (L48A$^{2.46}$, A54L$^{2.52}$, T65A$^{2.63}$ and Q89A$^{3.37}$) and the mutation N154A to remove a potential N-glycosylation site. A_{2A}R-GL31 was expressed in insect cells using the baculovirus expression system and purified in the detergent octylthioglucoside using Ni^{2+}-NTA affinity chromatography and size exclusion chromatography (see Methods). The purified receptor was crystallized in the presence of cholesteryl hemisuccinate by vapour diffusion, under conditions described in Methods.

Data collection, structure solution and refinement. Diffraction data were collected in multiple wedges (20° per wedge) from a single cryo-cooled crystal (100 K) for the GL31–NECA complex at beamline ID23-2 at the European Synchrotron Radiation Facility and from four crystals for the GL31–adenosine complex, at beamline I24 at the Diamond Light Source. The structures were solved by molecular replacement using the ZM241385-bound A_{2A}-T4L structure (PDB code 3EML)[8] as a model (see Methods). Data collection and refinement statistics are presented in Supplementary Table 1 and omit densities for the ligands are shown in Supplementary Fig. 6.

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

Received 22 February; accepted 21 April 2011.
Published online 18 May 2011.

1. Fredholm, B. B. *et al.* International Union of Basic and Clinical Pharmacology. LXXXI. Nomenclature and classification of adenosine receptors—an update. *Pharmacol. Rev.* **63,** 1–34 (2011).
2. Evans, B. A. *et al.* Ligand-directed signalling at β-adrenoceptors. *Br. J. Pharmacol.* **159,** 1022–1038 (2010).
3. Hofmann, K. P. *et al.* A G protein-coupled receptor at work: the rhodopsin model. *Trends Biochem. Sci.* **34,** 540–552 (2009).
4. Rosenbaum, D. M., Rasmussen, S. G. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459,** 356–363 (2009).
5. Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* **469,** 175–180 (2011).
6. Rosenbaum, D. M. *et al.* Structure and function of an irreversible agonist–β_2 adrenoceptor complex. *Nature* **469,** 236–240 (2011).
7. Warne, T. *et al.* The structural basis for agonist and partial agonist action on a β_1-adrenergic receptor. *Nature* **469,** 241–244 (2011).
8. Jaakola, V. P. *et al.* The 2.6 Ångstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. *Science* **322,** 1211–1217 (2008).
9. Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three dimensional models and computational probing of structure function relations in G protein-coupled receptors. *Methods Neurosci.* **25,** 366–428 (1995).
10. Kobilka, B. K. & Deupi, X. Conformational complexity of G-protein-coupled receptors. *Trends Pharmacol. Sci.* **28,** 397–406 (2007).
11. Yao, X. J. *et al.* The effect of ligand efficacy on the formation and stability of a GPCR-G protein complex. *Proc. Natl Acad. Sci. USA* **106,** 9501–9506 (2009).
12. Vauquelin, G. & Van Liefde, I. G protein-coupled receptors: a count of 1001 conformations. *Fundam. Clin. Pharmacol.* **19,** 45–56 (2005).
13. Palczewski, K. *et al.* Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289,** 739–745 (2000).
14. Li, J. *et al.* Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **343,** 1409–1438 (2004).
15. Park, J. H. *et al.* Crystal structure of the ligand-free G-protein-coupled receptor opsin. *Nature* **454,** 183–187 (2008).
16. Scheerer, P. *et al.* Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455,** 497–502 (2008).

17. Cherezov, V. *et al.* High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor. *Science* **318,** 1258–1265 (2007).
18. Rasmussen, S. G. *et al.* Crystal structure of the human β_2 adrenergic G-protein-coupled receptor. *Nature* **450,** 383–387 (2007).
19. Warne, T. *et al.* Structure of a β_1-adrenergic G-protein-coupled receptor. *Nature* **454,** 486–491 (2008).
20. Wu, B. *et al.* Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330,** 1066–1071 (2010).
21. Chien, E. Y. *et al.* Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **330,** 1091–1095 (2010).
22. Oldham, W. M. & Hamm, H. E. Heterotrimeric G protein activation by G-protein-coupled receptors. *Nature Rev. Mol. Cell Biol.* **9,** 60–71 (2008).
23. Murphree, L. J. *et al.* Human A_{2A} adenosine receptors: high-affinity agonist binding to receptor-G protein complexes containing $G\beta_4$. *Mol. Pharmacol.* **61,** 455–462 (2002).
24. Serrano-Vega, M. J., Magnani, F., Shibata, Y. & Tate, C. G. Conformational thermostabilization of the β_1-adrenergic receptor in a detergent-resistant form. *Proc. Natl Acad. Sci. USA* **105,** 877–882 (2008).
25. Magnani, F., Shibata, Y., Serrano-Vega, M. J. & Tate, C. G. Co-evolving stability and conformational homogeneity of the human adenosine A_{2A} receptor. *Proc. Natl Acad. Sci. USA* **105,** 10744–10749 (2008).
26. Shibata, Y. *et al.* Thermostabilization of the neurotensin receptor NTS1. *J. Mol. Biol.* **390,** 262–277 (2009).
27. Kim, S. K. *et al.* Modeling the adenosine receptors: comparison of the binding domains of A_{2A} agonists and antagonists. *J. Med. Chem.* **46,** 4847–4859 (2003).
28. Dal Ben, D. *et al.* Adenosine receptor modeling: what does the A_{2A} crystal structure tell us? *Curr. Top. Med. Chem.* **10,** 993–1018 (2010).
29. Wacker, D. *et al.* Conserved binding mode of human β2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J. Am. Chem. Soc.* **132,** 11443–11445 (2010).
30. Xu, F. *et al.* Structure of an agonist-bound human A_{2A} adenosine receptor. *Science* (2011).
31. Potterton, L. *et al.* Developments in the CCP4 molecular-graphics project. *Acta Crystallogr. D* **60,** 2288–2294 (2004).

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

Acknowledgements This work was supported by core funding from the Medical Research Council, and grants from Heptares Therapeutics Ltd and from the Biotechnology and Biological Sciences Research Council (BB/G003653/1). We would like to thank F. Magnani for technical help at the start of the project and F. Gorrec for developing the crystallization screen. We also thank the beamline staff at the European Synchrotron Radiation Facility, particularly at (beamline ID23-2; D. Flot and A. Popov), the Swiss Light Source (beamline X06SA) and at the Diamond Light Source (beamline I24; G. Evans, D. Axford and R. Owen). F. Marshall, M. Weir, M. Congreve and R. Henderson are thanked for their comments on the manuscript.

Author Contributions G.L. devised and performed receptor expression, purification, crystallization, cryo-cooling of the crystals, data collection, data processing and structure refinement. T.W. and P.C.E. helped with expression, crystal cryo-cooling and data collection. K.B. performed the radioligand binding assays and pharmacological analyses on receptor mutants in whole cells and C.J.L. was involved in data analysis and experimental design. A.G.W.L. was involved in data processing and structure refinement. Manuscript preparation was performed by G.L., A.G.W.L. and C.G.T. Overall project management was by C.G.T.

Author Information Co-ordinates and structure factors have been submitted to the PDB database under accession codes 2YDO and 2YDV for A_{2A}R-GL31 bound to adenosine or NECA, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.G.T. (cgt@mrc-lmb.cam.ac.uk).

METHODS

Expression, purification and crystallization. The human A_{2A} construct, GL31, contains four thermostabilizing point mutations (L48A$^{2.46}$, A54L$^{2.52}$, T65A$^{2.63}$ and Q89A$^{3.37}$), the mutation N154A to remove the potential N-glycosylation site and a truncation at the C terminus after Ala 316 (ref. 32). A polyhistidine tag (His_{10}) was engineered at the C terminus, separated from the receptor by a TEV protease cleavage site. Baculovirus expression and membrane preparation were performed as described previously for the β_1-AR[33].

Membranes were thawed at room temperature (20–25 °C), diluted with 25 mM HEPES pH 7.4, in the presence of protease inhibitors (Complete; Boehringer). Membranes were pre-incubated with NECA at 100 μM for 45 min at room temperature. The receptor–ligand complexes were then solubilised by adding decylmaltoside (DM) and NaCl to give final concentrations of 1.5% and 0.3 M, respectively, stirred for 30 min (4 °C) and insoluble material removed by ultracentrifugation (120,000*g*, 45 min, 4 °C). All protein purification steps were performed at 4 °C. The solubilised receptor sample was filtered through a 0.22 μm filter (Millipore) and applied to a 5 ml Ni-NTA superflow cartridge (Qiagen) pre-equilibrated with buffer (25 mM HEPES, pH 7.4, 0.1 M NaCl, 100 μM NECA, 0.15% DM, 2.5 mM imidazole). The column was washed sequentially with the same buffer supplemented with either 10, 40 or 80 mM imidazole, and then eluted with 250 mM imidazole. The eluted receptor–ligand complex was mixed with His_6-tagged TEV protease to cleave the tag for 4–6 h, 4 °C, concentrated to 2 ml using an Amicon-ultra spin concentrator (Ultracel-50K, Millipore) and then imidazole was removed using a PD-10 column (GE Healthcare). Eluted fractions were further purified by binding the TEV and other contaminants to Ni-NTA (QIAGEN) pre-equilibrated in 25 mM HEPES pH 7.4, 0.1 M NaCl, 100 μM NECA, 0.15% DM, 40 mM imidazole, incubating for 30 min and then collecting the flow-through. For detergent exchange into 0.35% octylthioglucoside (OTG), the sample was concentrated using an Amicon-ultra concentrator (Ultracel-50K, Millipore), diluted tenfold in 25 mM HEPES pH 7.4, 0.1 M NaCl, 100 μM NECA, 0.35% OTG, and concentrated again to 0.3 ml. The protein sample was applied to a Superdex 200 10/300 GL size-exclusion column pre-equilibrated in 25 mM HEPES pH 7.4, 0.1 M NaCl, 100 μM NECA, 0.35% OTG and run at 0.5 ml min^{-1}. Eluted receptor fractions (2–2.5 ml) were concentrated to 50–60 μl. Protein determination was performed using the amido black[34] assay.

Before crystallization, cholesteryl hemisuccinate (CHS) and OTG were added to 1 mg ml^{-1} and 0.5% respectively and the protein concentration adjusted to 10–12.5 mg ml^{-1}. NECA and adenosine A_{2A}-GL31 crystal hits were obtained using a new PEG-based crystallization screen developed in house[35]. Crystals were grown at 4 °C in 100 nl sitting drops using 0.05 M ADA NaOH, pH 6.4, 23.6% PEG 400, 4% v/v 2-propanol for the NECA complex. Crystals were cryoprotected by soaking in 0.05 M ADA NaOH, pH 6.4, 45% PEG 400. For the adenosine complex, crystals were initially grown in 0.05 M TrisHCl, pH 7.6, 9.6% PEG 200, 22.9%. PEG 300. Crystals were cryoprotected by soaking in 0.05 M TrisHCl, pH 7.5, 15% PEG 200, 30% PEG 300. The crystals were mounted on Hampton CrystalCap HT loops and cryo-cooled in liquid nitrogen.

Data collection, structure solution and refinement. Diffraction data for the NECA complex were collected at the European Synchrotron Radiation Facility with a Mar 225 CCD detector on the microfocus beamline ID23-2 (wavelength, 0.8726 Å) using a 10 μm focused beam and for the adenosine complex on beamline I24 at the Diamond Light Source with a Pilatus 6M detector and a 10 μm microfocus beam (wavelength 0.9778 Å). The microfocus beam was essential for the location of the best diffracting parts of single crystals, as well as allowing several wedges to be collected from different positions. Images were processed with MOSFLM[36] and SCALA[37]. The NECA complex was solved by molecular replacement with PHASER[38] using the A_{2A}-T4L structure (PDB code 3EML)[8] as a model after removal of the coordinates for T4L, all solvent molecules and the inverse agonist ZM241384. This structure was then used as a starting model for the structure solution of the adenosine complex. Refinement and rebuilding were carried out with REFMAC5[39] and COOT[40], respectively. In the final models, 98.1% of residues were in the favoured region of the Ramachandran plot with one outlier for the NECA complex, and 97.7% with no outliers for the adenosine complex. Smile strings for NECA and adenosine were created using Sketcher and dictionary entries using Libcheck. Hydrogen bond assignments for the ligands were determined using HBPLUS[41].

To facilitate a structural comparison between ZM241385-bound A_{2A}-T4L and the thermostabilized A_{2A}-GL31 with bound agonist, the structures were superimposed based on those residues in the region of the ligand-binding pocket that show the closest structural homology. This was achieved using the lsq_improve option of program O[42] and an initial transformation based on residues at the C terminus of helix 6 and the N terminus of helix 7. The final superposition, based on residues 16–21 in H1, 51–70 in H2 and ECL1, 132–140 in H4 and ECL2, 142–146 in ECL2, 166–182 in ECL2 and H5 and 245–283 in H6, ECL3 and H7, gave an r.m.s.d. in Cα positions of 0.66 Å for the 96 atoms and includes almost all residues involved in binding either ligand with the exception of those in H3. Using this transformation, the adenine moiety of the agonist superimposes well with the equivalent atoms of the triazolotriazene bicyclic ring of ZM241385 (r.m.s.d. 0.56 Å). Validation of the final refined models was carried out using Molprobity[43]. Omit densities for the ligands are shown in Supplementary Fig. 6. All figures in the manuscript were generated using either Pymol (DeLano Scientific) or CCPmg[31].

Binding of agonists and antagonist to A_{2A}R-GL31 expressed in CHO cells. Chinese hamster ovary (CHO) cells were maintained in culture in DMEM HAMs F12 media containing 10% FBS. Cells were transfected with plasmids expressing either wild-type adenosine A_{2A}R or A_{2A}R-GL31 using GeneJuice according to manufacturer's instructions (EMD Biosciences). Forty-eight hours after transfection, cells were harvested, centrifuged at 200*g* for 5 min at 4 °C and the pellet re-suspended in 20 mM HEPES, 10 mM EDTA buffer (pH 7.4). The membrane suspension was homogenized and centrifuged at 200*g* for 15 min at 4 °C. The supernatant was collected, the pellet re-suspended in 20 mM HEPES, 10 mM EDTA (pH 7.4) buffer and the solution homogenized and centrifuged as described previously[44]. The collected supernatant was centrifuged for 30 min at 40,000*g* at 4 °C. Pellets were re-suspended in 20 mM HEPES, 0.1 mM EDTA to a protein concentration of 1 mg ml^{-1} and stored at −80 °C .

Membranes from CHO cells transiently expressing wild-type or A_{2A}R-GL31 (10–15 μg per well) were assessed using competition [^{3}H]NECA binding in buffer containing 50 mM Tris-HCl (pH 7.4) as described previously[44]. Inhibition curves were fitted to a four-parameter logistic equation to determine IC_{50} values, which were converted to K_i values using K_d values determined by saturation binding and the [^{3}H]NECA concentration of 10 nM.

G-protein-coupling activity of A_{2A}R-GL31 measured in whole cells. A_{2A}R–His_6 and A_{2A}R-GL31–His_6 (amino acid residues 1–316 of human A_{2A}R) were subcloned into plasmid pcDNA5/FRT/TO using KpnI and NotI restriction sites. Flp-in T-Rex HEK293 cells were maintained at 37 °C in a humidified atmosphere in Dulbecco's modified Eagle's medium without sodium pyruvate, supplemented with 4,500 mg l^{-1} glucose, L-glutamine, 10% (v/v) FBS, 1% penicillin/streptomycin mixture and 10 μg ml^{-1} blasticidin. To generate stable cell lines, the cells were transfected with a ratio of 1:9 receptor cDNA in pcDNA5/FRT/TO vector and pOG44 vector using Genejuice as per manufacturer's instructions (EMD Biosciences). After 48 h, media were replaced with fresh medium supplemented with 200 μg ml^{-1} hygromycin B to select for stably expressing clones. Colonies were combined and tested for doxycycline-induced receptor expression. To induce receptor expression clones were treated with either 1 ng ml^{-1} or 3 ng ml^{-1} doxycyline for 16 h.

Cells were seeded at a density of 25,000 per well in a poly-L-lysine coated 96-well half area plate. Cells were induced with doxycyline (3 or 1 ng ml^{-1}) for 16 h. After 16 h media were removed and replaced with fresh media containing 100 μM Ro-201724 and 2 U ml^{-1} adenosine deaminase. Cells were incubated at 37 °C for 30 min before addition of varying concentrations of agonist (25 °C, 30 min). As a control cells were also incubated for 30 min (25 °C) with 10 μM forskolin. Cells were then lysed and cAMP produced detected using the CisBio cAMP kit according to manufacturer's instructions before plates were read on a PolarStar fluorescence plate reader.

32. Lebon, G. Bennett, K. Jazayeri, A. & Tate, C. G. Thermostabilization of an agonist-bound conformation of the human adenosine A_{2A} receptor. *J. Mol. Biol.* doi:10.1016/j.jmb.2011.03.075 (in the press).
33. Warne, T., Chirnside, J. & Schertler, G. F. Expression and purification of truncated, non-glycosylated turkey β-adrenergic receptors for crystallization. *Biochim. Biophys. Acta* **1610,** 133–140 (2003).
34. Schaffner, W. & Weissmann, C. A rapid, sensitive, and specific method for the determination of protein in dilute solution. *Anal. Biochem.* **56,** 502–514 (1973).
35. Gorrec, F., Palmer, C., Lebon, G. & Warne, T. Pi sampling: a methodical and flexible approach to macromolecular crystallization initial screening. *Acta Crystallogr. D* **67,** 463–470 (2011).
36. Leslie, A. G. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62,** 48–57 (2006).
37. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62,** 72–82 (2006).
38. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
39. 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).
40. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
41. McDonald, I. K. & Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238,** 777–793 (1994).
42. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
43. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–W383 (2007).
44. Robertson, N. *et al.* The properties of thermostabilised G protein-coupled receptors (StaRs) and their use in drug discovery. *Neuropharmacology* **60,** 36–44 (2011).

LETTER

doi:10.1038/nature10015

Tunable pK_a values and the basis of opposite charge selectivities in nicotinic-type receptors

Gisela D. Cymes[1] & Claudio Grosman[1]

Among ion channels, only the nicotinic-receptor superfamily has evolved to generate both cation- and anion-selective members. Although other, structurally unrelated, neurotransmitter-gated cation channels exist, no other type of neurotransmitter-gated anion channel, and thus no other source of fast synaptic inhibitory signals, has been described so far. In addition to the seemingly straightforward electrostatic effect of the presence (in the cation-selective members) or absence (in the anion-selective ones) of a ring of pore-facing carboxylates, mutational studies have identified other features of the amino-acid sequence near the intracellular end of the pore-lining transmembrane segments (M2) that are also required to achieve the high charge selectivity shown by native channels[1–10]. However, the mechanism underlying this more subtle effect has remained elusive[11] and a subject of speculation. Here we show, using single-channel electrophysiological recordings to estimate the protonation state of native ionizable side chains, that anion-selective-type sequences favour whereas cation-selective-type sequences prevent the protonation of the conserved, buried basic residues at the intracellular entrance of the pore (the M2 0′ position). We conclude that the previously unrecognized tunable charge state of the 0′ ring of buried basic side chains is an essential feature of these channels' versatile charge-selectivity filter.

The amino-acid differences that underlie the opposite charge selectivities of the members of the nicotinic-receptor superfamily have been known for several years now[1–10] (see Supplementary Fig. 1 and Supplementary Text for a brief introduction to this group of ion channels). Reversal-potential measurements have revealed that the cation-selective members of the superfamily become anion selective upon both insertion of a proline into the loop that connects transmembrane segments M1 and M2 (between positions −2′ and −1′), and mutation of the pore-lining glutamate at position −1′ to alanine, two changes that bring the sequence of the cation channels closer to that of their anion-selective counterparts (Fig. 1a). Similarly, the reciprocal changes (that is, deletion of the −2′ proline and mutation of the −1′ alanine to glutamate) engineered on (natively) anion-selective members of the superfamily have been shown to confer high selectivity for cations. Whereas the effect of the presence or absence of pore-exposed carboxylates may seem unsurprising, the basis for the effect of the insertion or deletion of a proline from the M1–M2 loop on charge selectivity (Fig. 1b) is much more subtle and has remained, largely, a mystery[11].

As a first step towards understanding the effect of the proline insertion, we engineered a proline between positions −2′ and −1′ of the α1, β1 and δ subunits of the (cation-selective) $(\alpha1)_2\beta1\delta\varepsilon$ acetylcholine receptor (muscle AChR), one subunit at a time. Because the ε subunit already has an extra residue at this position (a glycine; Fig. 1a), a proline was introduced in this subunit by a residue-to-residue mutation rather than by insertion. In the presence of a typical, divalent cation-containing solution in the pipette of cell-attached patches ($[Ca^{2+}] = 1.8$ mM; $[Mg^{2+}] = 1.7$ mM; solution 1 in Supplementary Table 1), single-channel recordings from the proline-insertion mutant in, for example, the δ subunit show an unusually noisy open-channel

a

		Intracellular end of M1			M1–M2 loop								Intracellular end of M2		
Subunit	Origin	−10′	−9′	−8′	−7′	−6′	−5′	−4′	−3′	−2′	−1′	0′	1′	2′	3′
AChRα1	Mouse	F	Y	L	P	T	D	S	-	G	E	K	M	T	L
AChRα1	Zebrafish	F	Y	L	P	T	D	S	-	G	E	K	M	T	L
AChRα1	*A. californica*	F	Y	L	P	S	D	S	-	G	E	K	I	T	L
AChRα1	Fruit fly	F	Y	L	P	S	D	S	-	G	E	K	I	S	L
AChRα4	Mouse	F	Y	L	P	S	E	C	-	G	E	K	V	T	L
AChRα7	Mouse	F	L	L	P	A	D	S	-	G	E	K	I	S	L
AChRβ1	Mouse	F	Y	L	P	P	D	A	-	G	E	K	M	G	L
AChRδ	Mouse	F	Y	L	P	G	D	C	-	G	E	K	T	S	V
AChRε	Mouse	Y	F	L	P	A	Q	A	G	G	Q	K	C	T	V
5-HT_{3A}R	Mouse	F	C	L	P	P	D	S	-	G	E	R	V	S	F
GlyRα1	Mouse	F	W	I	N	M	D	A	A	P	A	R	V	G	L
$GABA_A$Rα1	Mouse	F	W	L	N	R	E	S	V	P	A	R	T	V	F
GluClα1	*C. elegans*	F	W	F	D	R	T	A	I	P	A	R	V	T	L
HisClα1	Fruit fly	F	W	I	K	P	E	A	A	P	A	R	V	T	L
GlyRβ	Mouse	F	W	I	N	P	D	A	S	A	A	R	V	P	L
$GABA_A$Rβ1	Mouse	F	W	I	N	Y	D	A	S	A	A	R	V	A	L
GluClβ	*C. elegans*	F	W	I	D	L	H	S	T	A	G	R	V	A	L
AChR B	*L. stagnalis*	F	W	L	P	P	E	S	-	P	A	K	M	Q	L
Capca1_92948	*C. capitata*	F	W	L	P	P	E	S	-	P	A	K	M	M	L
Smp_176310	*S. mansoni*	F	W	L	P	P	E	T	-	P	A	K	M	V	L
AChR2β	*S. haematobium*	F	W	L	P	P	D	S	-	A	A	K	M	M	L
Helro1_192409	*H. robusta*	F	L	L	P	P	D	T	-	N	A	K	I	S	L

b

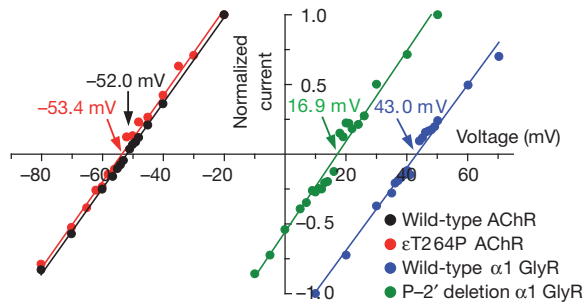


Figure 1 | The versatile charge selectivity of nicotinic-type receptors. **a**, Sequence alignment of residues in and flanking the M1–M2 loop. The broken horizontal line separates the sequences that are known or predicted (on the basis of their sequences) to form cation-selective channels (top) from those that are known or predicted to form anion-selective ones (bottom). Included in this alignment are subunits from receptors to acetylcholine (ACh), serotonin (5-HT), glycine (Gly), γ-aminobutyric acid (GABA), glutamate (Glu), histamine (His), and from receptors with as yet unidentified ligands. The invertebrate organisms in this list are: *Aplysia californica* (a mollusc), the fruit fly *Drosophila melanogaster* (an arthropod), *Caenorhabditis elegans* (a nematode), *Lymnaea stagnalis* (a mollusc), *Capitella capitata* (an annelid), *Schistosoma mansoni* and *S. haematobium* (two human parasitic platyhelminths) and *Helobdella robusta* (an annelid). **b**, Macroscopic current–voltage (*I*–*V*) relationships recorded under KCl-dilution conditions (solutions 8 and 9 in Supplementary Table 1; pH 7.4, both sides) in the outside-out configuration, as indicated in Supplementary Fig. 9 and in Methods. The equilibrium (Nernst) potentials at 22 °C, using ion concentration values, are −55.0 mV for K^+ and +50.6 mV for Cl^-. Reversal potentials are indicated. T264 denotes the threonine occupying position 12′, near the middle of M2, of the ε subunit of the AChR. P−2′ denotes the proline occupying position −2′ of the α1 subunit of the GlyR.

[1]Department of Molecular and Integrative Physiology, Center for Biophysics and Computational Biology, and Neuroscience Program, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

level and frequent sojourns of brief duration in the zero-current ('shut') level (Fig. 2a; openings are downward deflections). This behaviour, which differs markedly from that of the wild-type channel under identical experimental conditions (Supplementary Fig. 2), suggested the occurrence of channel block by some of the components of the patch-bathing solutions. Indeed, omission of both Ca^{2+} and Mg^{2+} from the solution in the pipette eliminated this block and, at the same time, uncovered an unanticipated phenomenon involving two interconverting open-channel conductance levels (Fig. 2a). Unless otherwise indicated, all the results reported here correspond to recordings obtained in the absence of extracellular Ca^{2+} or Mg^{2+}.

As illustrated in Fig. 2b, c, the proline mutation in the four types of AChR subunit leads to the appearance of current fluctuations between two levels of open-channel current with the higher level (the 'main level') having roughly the same conductance as the single level observed in the wild-type channel, at least in the case of the δ- and ε-subunit mutants, where the fluctuations were most clearly resolved. Although the conductance of the lower level (the 'sublevel') and the occupancy probabilities of the two alternative open-channel states differ among mutants, the underlying phenomenon is undoubtedly the same, as expected from the nearly symmetrical arrangement of the five AChR M2 segments around the channel pore (Supplementary Fig. 3). Moreover, the kinetics of the fluctuations were found to depend on the pH of the external and internal solutions (Fig. 2d) in a manner that is fully consistent with these current oscillations reflecting the alternate protonation and deprotonation of an ionizable side chain. Thus, the effect of these proline mutations on ion conduction is highly reminiscent of the effect of lysine, arginine or histidine substitutions along the M1, M2 or M3 transmembrane segments of the muscle AChR[12,13]; the remarkable difference, however, is that the mutations reported here do not introduce any new protonatable group in the protein's amino-acid sequence.

To identify the residue(s) responsible for this phenomenon, we mutated each of the (native) ionizable amino acids in and flanking the M2 segment to non-ionizable residues while keeping the extra proline inserted between positions −2′ and −1′. Combining the results of mutations in the β1 and δ subunits (Fig. 3), we conclude that the observed main-level ⇌ sublevel transitions reflect the protonation and deprotonation of the 0′-lysine side chain of the subunit containing the proline mutation. Furthermore, as protonation reduces

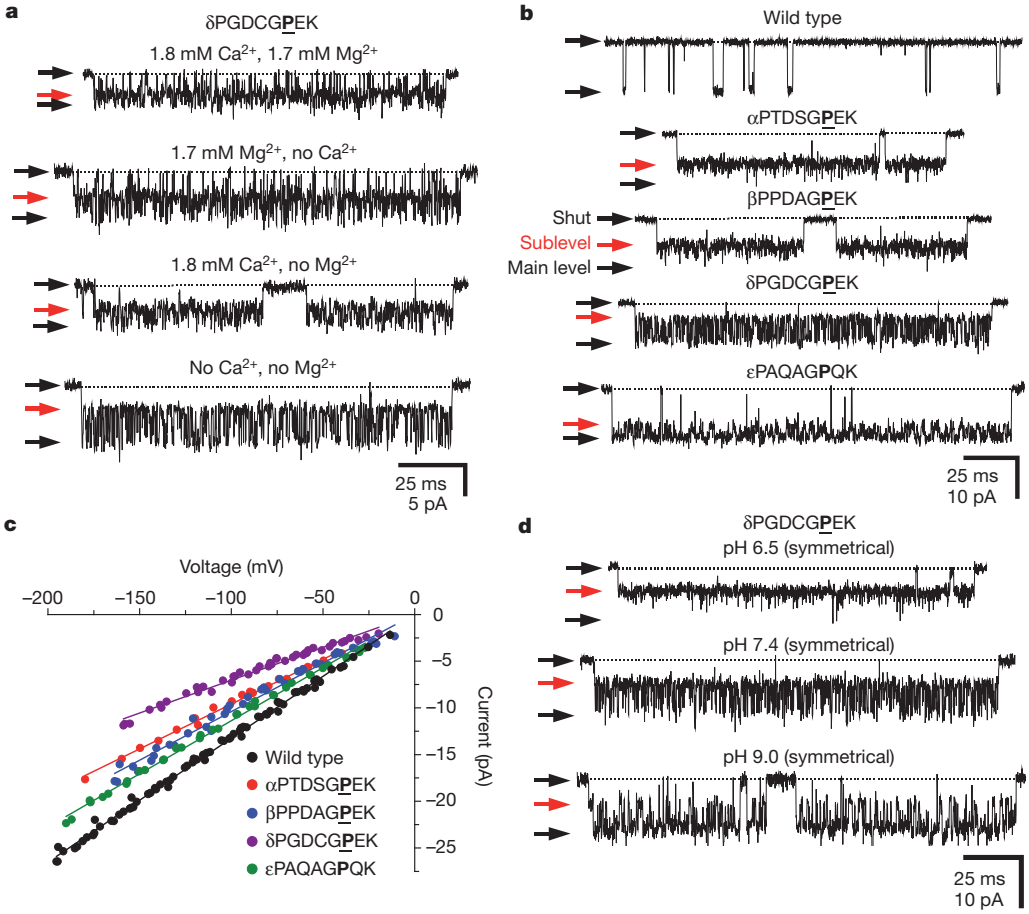


Figure 2 | A proline mutation unveils a proton-binding site. **a**, Single-channel inward currents (cell-attached configuration; approximately −100 mV; 1 μM ACh; pH of pipette solution ($pH_{pipette}$) 7.4) recorded from a mutant AChR having a proline inserted between positions −2′ and −1′ of the δ subunit. To increase the number of main-level ⇄ sublevel interconversions, a mutation that prolongs the mean duration of bursts of openings (εT264P) was also engineered. Solution compositions are indicated in Supplementary Table 1 (solutions 1–3). Mutations are indicated on the M1–M2 loop sequences; underlined bold symbols denote insertions whereas bold symbols (without the underline) denote substitutions. **b**, Inward currents (cell-attached configuration; approximately −100 mV; 1 μM ACh; $pH_{pipette}$ 7.4; solutions 2 and 3) recorded from the indicated AChR constructs. The burst-prolonging mutation was εT264P (in the case of AChRs with a proline inserted in the α1, β1 or δ subunit) or δS268Q (in the case of the glycine-to-proline substitution mutant at position −2′ of the ε subunit). In the case of the α1-subunit insertion, the trace shown corresponds to the construct having only one of the two α subunits mutated. **c**, Single-channel *I*–*V* relationships (cell-attached configuration; 1 μM ACh; $pH_{pipette}$ 7.4; solutions 2 and 3) recorded from the five constructs in **b**. For clarity, only the *I*–*V* curves corresponding to the sublevel are shown for the mutants. To facilitate the visual comparison of the slopes, each curve was displaced along the voltage axis so that it extrapolates exactly to the origin. **d**, pH dependence of the main-level ⇄ sublevel current fluctuations (outside-out configuration; −100 mV; 1 μM ACh; solutions 4 and 5).

the current amplitude (Fig. 2d), we also conclude that AChR mutants having a proline inserted (or substituted) into only one of the five subunits still conduct mostly cations.

A number of observations indicate that the lysines at position 0′ of the different AChR wild-type subunits reside on the stripe of M2 that faces away from the pore's lumen and that their εNH_2 groups are largely deprotonated, even at pH 6.0 (hence, $pK_a < 5.0$; Supplementary Text, Supplementary Figs 4, 5 and Supplementary Table 2). On introducing a proline, however, the affinity of these lysines for protons increases, probably as a result of a rearrangement of this portion of M2. This rearrangement does not seem to be a major one, however, because the extent to which the single-channel conductance is attenuated upon protonation of the 0′ lysines (a measure of the distance between the εNH_3^+ group and the long axis of the pore[12,13]; Supplementary Table 3) does not differ much from that caused by lysines engineered on the back or the sides of M2 (Supplementary Fig. 6) or the front of M1 or M3 (ref. 13). Moreover, although higher than in the wild-type AChR, the pK_a values of the εNH_3^+ group in the δ- and ε-subunit mutants (~7.58 and ~7.15, respectively; Supplementary Table 3) are still lower than the value expected for this group when fully exposed to bulk water (~10.4) by ~3 units (1 pK_a unit ≅ 1.36 kcal mol^{-1}). This further confirms the notion that these side chains do not face the aqueous lumen of the pore directly. For comparison, the pK_as of lysines engineered on the back of M2 (ref. 12) or on the front of M1 or M3 (ref. 13) are also, at least, ~3 units lower than the bulk-water value of ~10.4 (in Supplementary Table 4, we show that the pK_a values of substituted lysines are rather insensitive to the presence or absence of millimolar concentrations of external Ca^{2+} or Mg^{2+}). In addition, we conclude that the effect of these mutations is not highly position-specific because proline insertions at the five other possible positions along the M1–M2 loop of the δ subunit give rise, essentially, to the same pH-dependent phenotype (Supplementary Fig. 7).

The exact nature of the reorganization of the M1–M2 loop upon mutation, and how this change lowers the hydrophobicity of the microenvironment around the 0′ basic side chain, remains unknown. However, an increased exposure to water (through an increase in solvent penetration and/or a slight repositioning of the side chain) is expected to be an important factor in the stabilization of a positive charge buried in a region of the protein that lacks properly oriented acidic side chains. Indeed, note that the side chains of the nearby −5′ aspartate or the −1′ glutamate do not contribute to the observed main-level ⇄ sublevel current fluctuations (Fig. 3a), either by providing the proton-binding site itself or by electrostatically stabilizing the 0′ εNH_3^+ group. The idea of a structural rearrangement around position 0′ receives further support from the finding of a complex interaction between the proline mutants and extracellular Ca^{2+} and Mg^{2+} (compare Fig. 2a with Supplementary Fig. 2). Certainly, as elaborated in the Supplementary Text, it seems reasonable to ascribe the anomalous nature of this interaction to the probable concomitant rearrangement of the ring of glutamates at the neighbouring position −1′.

However, not all of the anion-selective members of the superfamily contain a full ring of 'inserted' prolines at position −2′ of the M1–M2 loop. Instead, some β-subunit homomers (such as those formed by the β subunits of $GABA_A$ receptors[14] or of invertebrate GluCl receptors[15]; Fig. 1a) present a full ring of alanines at this position without sacrificing high selectivity for anions. And, even more divergently, some highly anion-selective AChRs from invertebrates do not contain any extra residues in the M1–M2 loop, but rather replace the −2′ glycine of the cation-selective counterparts with a proline[16] (Fig. 1a). Remarkably, we found that mutating the muscle AChR to mimic the 'atypical' features of these M1–M2 loops also gives rise to current fluctuations that closely resemble those caused by the—more common—proline insertions characterized above. In fact, we found that the insertion or substitution of a number of amino acids (not only proline or alanine) in and around position −2′ have largely the same effect (Fig. 4 and Supplementary Text).

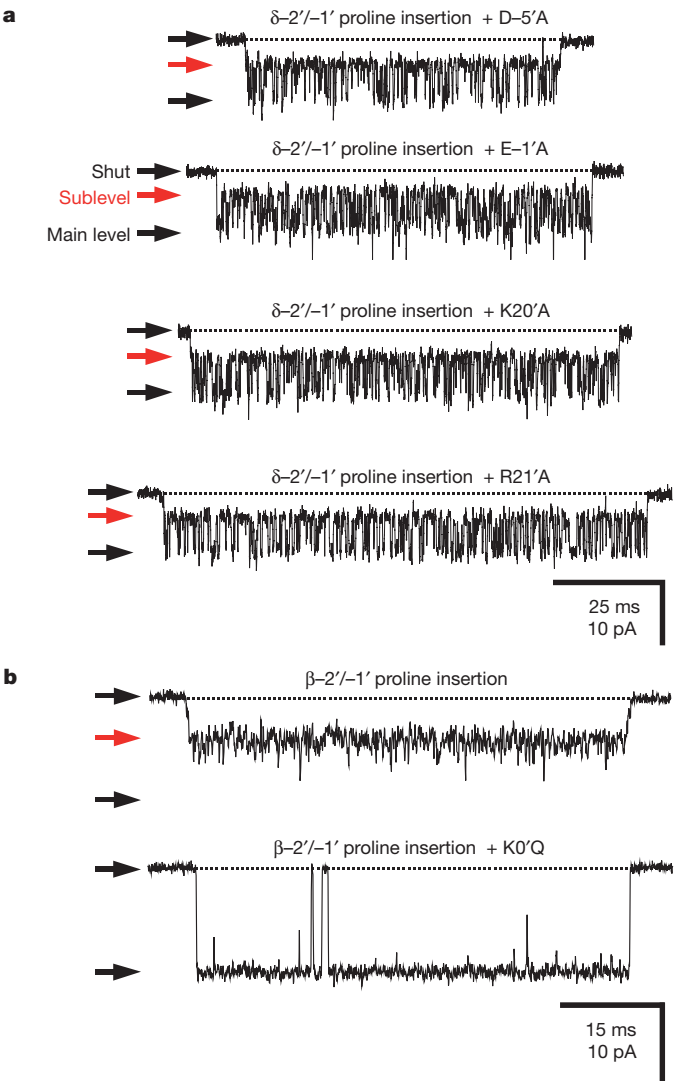


Figure 3 | The side chain of the 0′ basic residue is the proton-binding site. **a**, Single-channel inward currents (cell-attached configuration; 1 μM ACh; $pH_{pipette}$ 7.4; solutions 2 and 3) recorded from AChRs with a proline inserted between positions −2′ and −1′ of the δ subunit and having four of the five native ionizable residues that flank δM2 mutated to alanine, one at a time. The burst-prolonging mutation was εT264P. Mutation of the fifth residue (the 0′ lysine) to alanine, glutamine or valine (in the presence of the inserted proline) abolishes receptor expression in the plasma membrane, as revealed by the lack of specific α-bungarotoxin binding. The applied potential was approximately −100 mV for all constructs, with the exception of the receptor containing the glutamate-to-alanine mutation at position −1′, in which case the potential was approximately −150 mV (to compensate for its lower single-channel conductance). **b**, Inward currents recorded from a mutant AChR having a proline inserted between positions −2′ and −1′ of the β1 subunit and from the mutant having, in addition, a lysine-to-glutamine mutation at position 0′ of the same subunit. The applied potential was approximately −100 mV. All other experimental conditions were as in **a**.

Evidently, charge-selective permeation through members of the nicotinic-receptor superfamily has arisen during evolution as a result of several different amino-acid changes in the M1–M2 loop; yet, all these changes seem to act, at least in part, by tuning the proton affinity of the same basic side chain. Consistent with the functional relevance of the different protonation states of the ionizable group at 0′ (neutral in cation-selective members and, at least partly, positively charged in anion-selective members) we notice that some recently identified cation-selective members of the superfamily from nematodes[17] and bacteria[18] replace this lysine or arginine with non-ionizable residues. However, all known native anion-selective nicotinic-type receptors

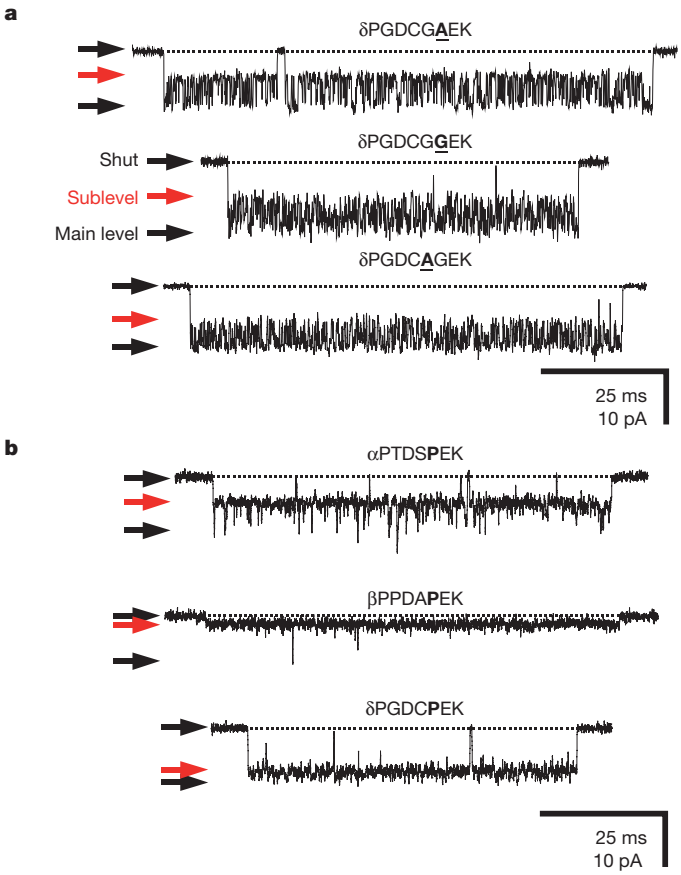


Figure 4 | Not only prolines, not only insertions. **a**, Single-channel inward currents (cell-attached configuration; approximately −100 mV; 1 μM ACh; $pH_{pipette}$ 7.4; solutions 2 and 3) recorded from the indicated AChR insertion mutants. The burst-prolonging mutation was εT264P. Threonine insertions have a similar effect. **b**, Inward currents recorded from the indicated AChR substitution mutants under the same experimental conditions as in **a**. Note that the insertion of a residue is not required to reveal a proton-binding site. Instead, replacing the conserved glycine at position −2′ with a variety of other residues (see Supplementary Text; only proline is shown here) also unveils a protonation site in the four types of subunit. In the case of the α1-subunit mutant, the trace shown corresponds to the construct having only one of the two α subunits mutated. The trace illustrating the effect of a glycine-to-proline mutation at this position of the ε subunit is shown in Fig. 2b.

present a basic residue at position 0′. Whether these additional positive charges are directly responsible for the anion selectivity or, rather, they act to increase the single-channel current amplitude of a channel that is highly selective for anions irrespective of the protonation state of the 0′ basic side chains (as a result, perhaps, of concomitant changes in pore size[6,11,19] or in the orientation of backbone groups[2,10]) remains unclear. What is clear, however, is that both high charge selectivity and high single-channel current amplitude are essential for proper electrical signalling at fast chemical synapses. What is also clear is that the differential tuning of side-chain pK_a values described here represents a novel mechanism for turning protein charges on or off without the need of replacing ionizable amino acids with non-ionizable ones (or vice versa).

It is worth noting that the finding of different protonation states for the 0′ basic side chain in cation-selective-type versus anion-selective-type charge-selectivity filters would have gone unnoticed by even such powerful approaches as X-ray or electron crystallography. Certainly, these methods do not typically reach the resolution of 1.0–1.2 Å (especially when applied to membrane proteins) that is needed to detect the presence of hydrogen atoms. Also, although a variety of structure-based computational algorithms for the prediction of protein side-chain pK_a values have been developed and could in principle be applied to structural models of members of the nicotinic-receptor superfamily, their accuracy in the case of large deviations from values in bulk water is still very limited[20–24].

Overall, our results provide a compelling example of the marked sensitivity of side-chain pK_a values to the details of the microenvironment, of the profound impact that differentially tuned proton affinities can have on protein function, and of the advantage evolution has taken of this physicochemical phenomenon. Lastly, our data also remind us that assuming default protonation states for the ionizable side chains in a protein may be highly misleading, and that the structural determinants of ion-conduction properties through ion channels need not face the lumen of the pore directly.

METHODS SUMMARY

Currents were recorded from HEK-293 cells transiently transfected with wild-type or mutant complementary DNAs (cDNAs) encoding the adult muscle-type AChR (mouse α1, β1, δ and ε subunits) or the α1 GlyR (human or rat isoform b). Single-channel currents were recorded at 22 °C from cell-attached patches with the exception of recordings that required access to both sides of the membrane in which case the outside-out configuration with a constant application of ligand was used. Ensemble ('macroscopic') currents were recorded at 22 °C from outside-out patches exposed to step changes in the concentration of ligand. The composition of all solutions used for electrophysiological recordings is given in Supplementary Table 1. Extent-of-channel-block and pK_a values were estimated from cell-attached, single-channel recordings as detailed in our previous work[12,13] and in Supplementary Fig. 8. All single-channel current traces are displayed at $f_c \approx 6$ kHz. Reversal potentials were estimated from macroscopic-current recordings elicited by 1- or 10-ms pulses of ligand applied to outside-out patches (Supplementary Fig. 9). The expression of mutant AChRs in the plasma membrane of transfected cells was estimated using an equilibrium [^{125}I]-α-bungarotoxin binding assay.

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

Received 5 April 2010; accepted 21 March 2011.
Published online 22 May; corrected 23 June 2011 (see full-text HTML version for details).

1. Galzi, J. L. *et al.* Mutations in the ion channel domain of a neuronal nicotinic receptor convert ion selectivity from cationic to anionic. *Nature* **359,** 500–505 (1992).
2. Corringer, P.-J. *et al.* Mutational analysis of the charge selectivity filter of the α7 nicotinic acetylcholine receptor. *Neuron* **22,** 831–843 (1999).
3. Keramidas, A., Moorhouse, A. J., French, C. R., Schofield, P. R. & Barry, P. H. M2 pore mutations convert the glycine receptor channel from being anion- to cation-selective. *Biophys. J.* **79,** 247–259 (2000).
4. Gunthorpe, M. J. & Lummis, S. C. R. Conversion of the ion selectivity of the 5-HT_{3A} receptor from cationic to anionic reveals a conserved feature of the ligand-gated ion channel superfamily. *J. Biol. Chem.* **276,** 10977–10983 (2001).
5. Jensen, M. L. *et al.* The β subunit determines the ion selectivity of the $GABA_A$ receptor. *J. Biol. Chem.* **277,** 41438–41447 (2002).
6. Keramidas, A., Moorhouse, A. J., Pierce, K. D., Schofield, P. R. & Barry, P. H. Cation-selective mutations in the M2 domain of the inhibitory glycine receptor channel reveal determinants of ion-charge selectivity. *J. Gen. Physiol.* **119,** 393–410 (2002).
7. Thompson, A. J. & Lummis, S. C. R. A single ring of charged amino acids at one end of the pore can control ion selectivity in the 5-HT_3 receptor. *Br. J. Pharmacol.* **140,** 359–365 (2003).
8. Wotring, V. E., Miller, T. S. & Weiss, D. S. Mutations at the GABA receptor selectivity filter: a possible role for effective charges. *J. Physiol. (Lond.)* **548,** 527–540 (2003).
9. Menard, C., Horvitz, H. R. & Cannon, S. Chimeric mutations in the M2 segment of the 5-hydroxytryptamine-gated chloride channel MOD-1 define a minimal determinant of anion/cation permeability. *J. Biol. Chem.* **280,** 27502–27507 (2005).
10. Sunesen, M. *et al.* Mechanism of Cl^- selection by a glutamate-gated chloride (GluCl) receptor revealed through mutations in the selectivity filter. *J. Biol. Chem.* **281,** 14875–14881 (2006).
11. Keramidas, A., Moorhouse, A. J., Schofield, P. R. & Barry, P. H. Ligand-gated ion channels: mechanisms underlying ion selectivity. *Prog. Biophys. Mol. Biol.* **86,** 161–204 (2004).
12. Cymes, G. D., Ni, Y. & Grosman, C. Probing ion-channel pores one proton at a time. *Nature* **438,** 975–980 (2005).
13. Cymes, G. D. & Grosman, C. Pore-opening mechanism of the nicotinic acetylcholine receptor evinced by proton transfer. *Nature Struct. Mol. Biol.* **15,** 389–396 (2008).
14. Krishek, B. J., Moss, S. J. & Smart, T. G. Homomeric β1 γ-aminobutyric $acid_A$ receptor-ion channels: evaluation of pharmacological and physiological properties. *Mol. Pharmacol.* **49,** 494–504 (1996).
15. Cully, D. F. *et al.* Cloning of an avermectin-sensitive glutamate-gated chloride channel from *Caenorhabditis elegans*. *Nature* **371,** 707–711 (1994).

16. van Nierop, P. *et al.* Identification of molluscan nicotinic acetylcholine receptor (nAChR) subunits involved in formation of cation- and anion-selective nAChRs. *J. Neurosci.* **25,** 10617–10626 (2005).
17. Beg, A. A. & Jorgensen, E. M. EXP-1 is an excitatory GABA-gated cation channel. *Nature Neurosci.* **6,** 1145–1152 (2003).
18. Bocquet, N. *et al.* A prokaryotic proton-gated ion channel from the nicotinic acetylcholine receptor family. *Nature* **445,** 116–119 (2007).
19. Lee, D. J.-S., Keramidas, A., Moorhouse, A. J., Schofield, P. R. & Barry, P. H. The contribution of proline 250 (P-2′) to pore diameter and ion selectivity in the human glycine receptor channel. *Neurosci. Lett.* **351,** 196–200 (2003).
20. Schutz, C. N. & Warshel, A. What are the dielectric 'constants' of proteins and how to validate electrostatic models? *Proteins* **44,** 400–417 (2001).
21. Harms, M. J. *et al.* The pK_a values of acidic and basic residues buried at the same internal location in a protein are governed by different factors. *J. Mol. Biol.* **389,** 34–47 (2009).
22. Kamerlin, S. C. L., Haranczyk, M. & Warshel, A. Progress in *ab initio* QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of p*K*, redox reactions and solvation free energies. *J. Phys. Chem. B* **113,** 1253–1272 (2009).
23. Karp, D. A., Stahley, M. R. & García-Moreno, E. B. Conformational consequences of ionization of Lys, Asp, and Glu buried at position 66 in staphylococcal nuclease. *Biochemistry* **49,** 4138–4146 (2010).
24. Chimenti, M. S., Castañeda, C. A., Majumdar, A. & García-Moreno, E. B. Structural origins of high apparent dielectric constants experienced by ionizable groups in the hydrophobic core of a protein. *J. Mol. Biol.* **405,** 361–377 (2011).

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

Acknowledgements We thank S. Sine for wild-type muscle AChR cDNA; M. Slaughter and D. Papke for wild-type α1 GlyR cDNA; S. Elenes for critical advice on fast-perfusion experiments; E. Jakobsson and H. Robertson for discussions; and G. Papke, M. Maybaum, J. Pizarek and C. Staehlin for technical assistance. This work was supported by a grant from the US National Institutes of Health (R01-NS042169 to C.G.).

Author Contributions G.D.C. and C.G. designed experiments, analysed data and wrote the manuscript; G.D.C performed experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.G. (grosman@illinois.edu).

METHODS

DNA clones, mutagenesis and transfection. HEK-293 cells were transiently transfected with cDNAs encoding the adult muscle-type AChR (mouse α1, β1, δ and ε subunits) or the α1 GlyR (human or rat isoform b prepared as indicated in ref. 25; no differences were found between the charge selectivities of these two orthologues) using a calcium-phosphate precipitation method. Mutations were engineered using the QuikChange site-directed mutagenesis kit (Stratagene) and were confirmed by dideoxy sequencing. When deemed necessary, mutations that prolong individual activations of the channel ('bursts of openings') were also introduced in the mutant AChR constructs to increase the number of proton-transfer events recorded. These mutations were βV266M (M2 position 13′; ref. 26), δS268Q (M2 12′; ref. 27) or εT264P (M2 12′; ref. 28), and their lack of appreciable effect on charge selectivity and single-channel conductance is shown in Fig. 1b and Supplementary Fig. 10, respectively.

Extent of channel block. Single-channel currents were digitized (at 100 kHz), filtered (cascaded $f_c \approx 30$ kHz) and idealized (using the SKM algorithm in QuB software[29]) to obtain the mean amplitudes of the different current levels and the sequences of dwell times. *I–V* curves were generated from patch-clamp recordings obtained in the cell-attached configuration, and these only include data on inward currents (for example, Fig. 2c). For most mutants studied here, the rectilinear portion of the sublevel's *I–V* curve extrapolates onto the voltage axis at a negative potential (whereas the main-level's curve extrapolates near zero), a probable result of the more pronounced inward rectification of the current sublevel and, for some mutants at least, the result perhaps of the diminished selectivity of the sublevel for cations (note that, under the conditions of our cell-attached experiments, a decrease in cation selectivity would shift the reversal potential to negative values). Because of these different intercepts, the ratio between the sublevel and the main-level single-channel current amplitudes becomes a function of the transmembrane potential, and the choice of any particular voltage value to calculate the extent of channel block from current amplitudes would be arbitrary. Hence, here (as in our previous work[12,13]) we chose to calculate the extent of block using single-channel conductances, instead. As a result, our extent-of-block values differ from those that could be inferred from a mere inspection of the single-channel traces at a single potential (for example, −100 mV, in the case of our figures). Different intercepts for the rectilinear portions of the main-level and sublevel *I–V* curves are not unique to the mutants studied here; rather, these differences were also observed for AChR mutants bearing engineered basic residues along M2 (ref. 12). The extent of channel block for each construct was calculated as the difference between the conductance values of the main level and the sublevel normalized by the conductance of the main level. In some cases, the conductance of the main level could not be estimated with confidence (for example, because the open-channel signal dwelled only briefly and infrequently in the main level). In these cases, the normalization was done relative to the conductance of the corresponding background construct (that is, the wild-type AChR with or without one of the burst-prolonging mutations).

pK_a values. Protonation and deprotonation rates (as well as all other transition rates) were estimated from maximum-likelihood fits of single-channel dwell-time sequences with kinetic models (Supplementary Fig. 8) as described in our previous work[12,13]. To this end, we used the MIL algorithm in QuB software[30] with a retrospectively imposed time resolution of 25 μs. The ratio between the proton-dissociation and proton-association rates thus estimated gives the ratio of the probabilities of the engineered ionizable side chain being deprotonated versus protonated while the channel is open. The reported pK_a values (Supplementary Tables 3 and 4) were calculated from the product of these ratios and the concentration of protons in the channel-bathing solution (Supplementary Fig. 8). For the calculation of the pK_as of lysine side chains engineered in M1 or M2 (Supplementary Table 4), we used the concentration of protons in the pipette solution of cell-attached patches (that is, pH 6.0 in the case of the mutant at position 11′ and 7.4 in all other cases). For the calculation of the pK_as of the 0′ side chain in the various M1–M2 loop mutants studied here (Supplementary Table 3), however, the choice of a pH value is not obvious because we found that the kinetics of protonation and deprotonation in this region of the channel are sensitive to the pHs of the two solutions bathing the membrane, behaving as if the protonatable group were exposed to a solution of intermediate pH. Hence, although probably not strictly correct, we decided to use a pH of 7.3, a value halfway between the pH of the pipette solution (~7.4) and that of the cytosol (~7.2). This uncertainty leads to a maximum systematic error of ± 0.1 units in the pK_a estimates shown in Supplementary Table 3.

Concentration jumps, reversal potentials and kinetics. Step changes in the concentration of ligand bathing the external aspect of outside-out patches were achieved by the rapid switching of two solutions (differing only in the presence or absence of ligand) flowing from either barrel of a piece of theta-type capillary glass mounted on a piezo-electric device (Burleigh-LSS-3100; Lumen Dynamics) as described previously[31] (solution-exchange time$_{10–90\%}$ < 150 μs). Reversal potentials were estimated from *I–V* relationships generated by plotting the peak-current responses to brief (1- or 10-ms) pulses of ligand applied to outside-out patches at concentrations that evoke nearly maximal responses (100 μM ACh for the AChR; 10 mM Gly for the α1 GlyR). Consecutive pulses were separated by 8-s intervals during which the patches were exposed to ligand-free solution and the applied voltage was changed. In these particular experiments, the reference Ag/AgCl wire was connected to the bath solution (the composition of which was the same as that of the solution flowing through the theta-type glass tubing; solution 9 in Supplementary Table 1) through an agar bridge containing 200 mM KCl, to minimize the liquid-junction potential. A new, fresh agar bridge was connected every <2 h. Liquid-junction potentials were calculated using the JPCalc module in pClamp 9.0 (ref. 32). To characterize the kinetics of AChR deactivation, entry into desensitization and recovery from desensitization, and the response to the repetitive application (25 Hz) of nearly saturating ACh, macroscopic currents were recorded from outside-out patches (at –80 mV) using various 100-μM ACh pulse protocols, as indicated in Supplementary Fig. 5. All macroscopic currents were analysed using a combination of pClamp 9.0 (Molecular Devices) and SigmaPlot 7.101 (Systat Software) software.

Plasma-membrane AChR expression. To estimate the number of wild-type or mutant AChRs in the plasma membrane, transfected HEK-293 cells were incubated with 20-nM [^{125}I]-α-bungarotoxin (PerkinElmer) in fresh DMEM culture medium at 4–5 °C for 2–3 h so as to saturate all toxin-binding sites. The associated radioactivity was measured in a γ-counter and was normalized to the corresponding mass of total protein, which was quantified using the bicinchoninic-acid method (Thermo Scientific) after solubilizing the cells with 0.1 N NaOH. The non-specific binding of radiolabelled toxin was estimated on cells transfected with cDNA encoding the β1, δ and ε subunits of the mouse-muscle AChR (but not the α1 subunit). The amount of [^{125}I]-α-bungarotoxin bound to these mock-transfected cells (normalized to total protein content) was never higher than 6% of that associated with the expression of the wild-type AChR.

25. Papke, D., Gonzalez-Gutierrez, G. & Grosman, C. Desensitization of neurotransmitter-gated ion channels during high-frequency stimulation: a comparative study of Cys-loop, AMPA and purinergic receptors. *J. Physiol. (Lond.)* **589,** 1571–1585 (2011).
26. Engel, A. G. *et al.* New mutations in acetylcholine receptor subunit genes reveal heterogeneity in the slow-channel congenital myasthenic syndrome. *Hum. Mol. Genet.* **5,** 1217–1227 (1996).
27. Grosman, C. & Auerbach, A. Asymmetric and independent contribution of the second transmembrane segment 12′ residues to diliganded gating of acetylcholine receptor channels. A single-channel study with choline as the agonist. *J. Gen. Physiol.* **115,** 637–651 (2000).
28. Ohno, K. *et al.* Congenital myasthenic syndrome caused by prolonged acetylcholine receptor channel openings due to a mutation in the M2 domain of the epsilon subunit. *Proc. Natl Acad. Sci. USA* **92,** 758–762 (1995).
29. Qin, F. Restoration of single-channel currents using the segmental k-means method based on hidden Markov modeling. *Biophys. J.* **86,** 1488–1501 (2004).
30. Qin, F., Auerbach, A. & Sachs, F. Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.* **70,** 264–280 (1996).
31. Elenes, S., Ni, Y., Cymes, G. D. & Grosman, C. Desensitization contributes to the synaptic response of gain-of-function mutants of the muscle nicotinic receptor. *J. Gen. Physiol.* **128,** 615–627 (2006).
32. Barry, P. H. & Lynch, J. W. Liquid junction potentials and small cell effects in patch-clamp analysis. *J. Membr. Biol.* **121,** 101–117 (1991).

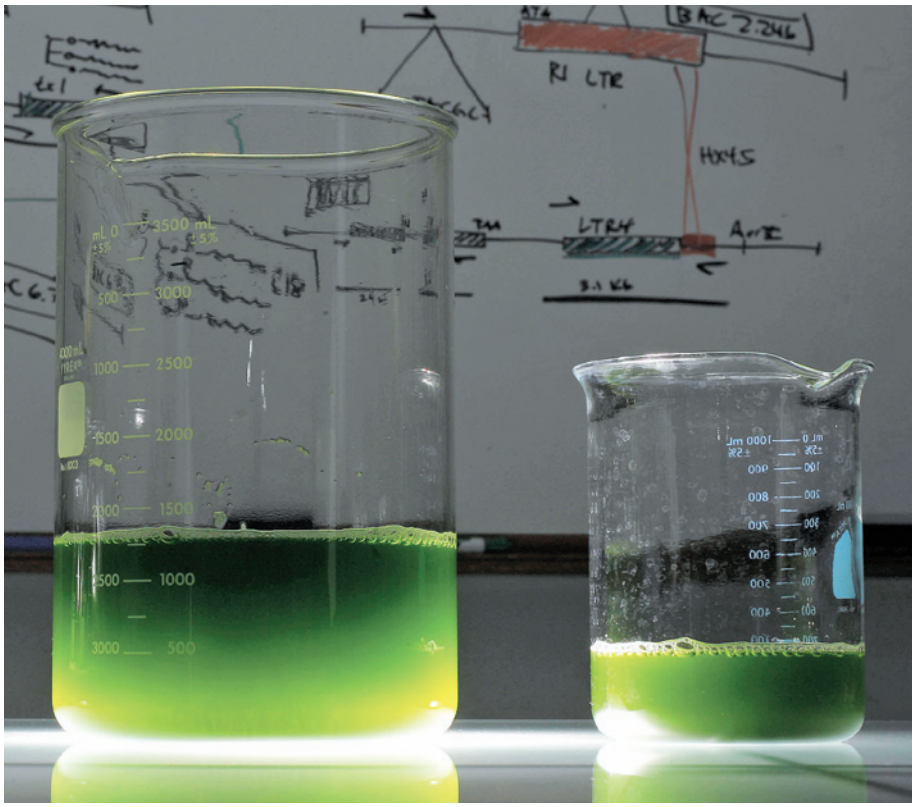
CAREERS

CANADA Postdoctoral researchers endure a heavy tax burden **p.533**

JOURNAL ENTRIES Read stories of young scientists' travails **go.nature.com/3fttcj**

NATUREJOBS For the latest career listings and advice **www.naturejobs.com**

M. MACOR/SAN FRANCISCO CHRONICLE/CORBIS



Early attempts to make algae-based biofuels are creating research opportunities.

RENEWABLE ENERGY

Making fuels for the future

The field of biofuels has suffered from cycles of interest and neglect. Still, enterprising researchers have myriad options.

BY CHARLOTTE SCHUBERT

Haiyu Ren completed her PhD in agricultural and biological engineering in 2006, just as biofuels research was heating up. Oil prices were spiking, and governments and investors worldwide were beginning to look at biofuel options beyond soya biodiesel and ethanol made from maize (corn). Money began to flow to research into fuels made from woodchips, algae and wheat straw.

Ren got only one job interview — with Novozymes, an enzymes company based in Bagsvaerd, Denmark. She was part of a surge of researchers newly enthusiastic about alternatives to petroleum, all competing for the job — but she got it.

Three years later, Ren was working in Beijing, leading a lab of 16 scientists — biochemists, molecular biologists and chemical engineers — collaborating with several Chinese companies on two pilot-scale facilities that used Novozymes enzymes to help process corn stalks. One facility turned the stalks into ethanol, and another turned them into ethylene glycol, a building block of plastic.

But last year, her career took a turn. Ren left research and development to try her hand at the business side of the industry, with the aim of working on international partnerships. She is now based in the Novozymes business-development office in Bagsvaerd, and later this year will begin earning a master's degree in business administration, courtesy of the company.

Ren's changing career parallels the evolution of an international field. Growing options stem, in part, from an expansion of research aims: many companies that once focused exclusively on turning biomass into fuel are, like Novozymes, now exploring how to turn plant-derived substances such as starch, sugar or cellulose into higher-value chemicals from which paints, plastics and other industrial products can be made. "The number of jobs has increased significantly since 2006, and there is more variety," says Tom Richard, director of the Penn State Institutes of Energy and the Environment in University Park, Pennsylvania.

Still, the fields of advanced biofuels and biochemicals are young, so uncertainty remains about long-term funding and investment. Hundreds of millions of dollars in government and private financing have been pumped into the industry, but no commercial-scale facility to turn biomass into fuel has yet launched successfully, owing to cooling investor interest, technology lagging behind initial promises, or both. But the field's immaturity also means that there is still a need for highly trained scientists who can make the technology work. For researchers flexible enough to weather uncertainty, biofuels offer an opportunity to help mould an emerging industry.

BREAKING THROUGH

Research in advanced biofuels focuses on reducing the cost of squeezing fuel out of biomass, aiming to make it competitive with the fluctuating price of oil. To achieve that goal, scientists must find efficient, elegant ways to extract molecules such as glucose — a crucial chemical from which ethanol and other products are made — from lignocellulose, the hard woody part of plants. Companies such as Novozymes use proprietary enzymes to help chew up the lignocellulose after a 'pretreatment' step, which often involves steam or acid. Other firms gasify the lignocellulose to break it apart, or use ▶

engineered, heat-loving bacteria to help with the task. Hundreds of small companies are working on converting all kinds of substances, from municipal waste to algae, into a range of products, including but not limited to ethanol, the industrial chemical polylactic acid, and 'drop-in' fuels such as butanol, that are close substitutes for petroleum, diesel or jet fuel.

After a serious lull during the recession, companies big and small have begun to advertise job openings. They are seeking yeast geneticists, enzyme biochemists, polymer chemists, fermentation scientists, agronomists and chemical engineers, among others. The field's technological challenges have created demand for researchers at all levels — from basic scientists who can engineer microbes and algae to produce fuel to chemical engineers and fermentation experts, whose expertise is needed in scaling up pilot facilities. The near-term demand is highest for researchers in biochemical engineering, biochemistry, microbiology, metabolic engineering and plant biology, says Bruce Dale, a chemical engineer at Michigan State University in East Lansing who has mentored students for more than 30 years.

"We are motivated by this big vision to get renewable energy on board. I dream about my work."
Haiyu Ren

Employers look for people with broad experience, says Dale, and many of the jobs use interchangeable skills. "This field rewards people who can work collaboratively across disciplines," he says. Dale advises chemical engineers to get training in microbiology or molecular biology, and biochemists to take an engineering class or two. Richard recommends that students undertake coursework in fuel chemistry, combustion science, emissions regulations and analysis of greenhouse-gas emissions. He also advises those interested in advanced biochemicals to take courses in organic chemistry (including green, or environmental, chemistry), materials science and engineering (especially polymer science), and product design.

Ren's education prepared her to work on all aspects of pilot-plant design: optimization of enzyme use, management of waste water and others. Her classes at university ranged from carbohydrate chemistry to one examining the machinery used in crop harvesting, and set her up for work with both biofuels and biochemicals. "I am not an expert in all these areas," she says. "But the education enabled me to apply the right method to solve the challenges step by step, or to find the right person to consult with."

More jobs are emerging for people with these skill sets, not only in corporate research labs, but also in manufacturing, supply-chain logistics, finance and marketing. "As this industry moves out of the laboratory and into the factories," says Richard, "this diversity and flexibility to absorb allied career tracks can only increase." Recent graduates mentored by Dale and Richard have gone on to work in research positions at DuPont — the chemicals giant based in Wilmington, Delaware — and Novozymes; others are doing postdoctoral work in academic labs, and one became an analyst at the US Energy Information Administration in Washington DC. In the future, opportunities will also emerge for experts in sustainable agriculture and crop-yield optimization, so graduates with degrees in plant physiology, forestry or agricultural engineering will become more sought after, says Dale.

Biofuels companies will take on researchers with basic experience in microbiology or yeast genetics, but it helps if applicants are educated at an institution with a reputation for putting a large number of scientists into the field. In Europe, that means programmes such as that at BE-Basic, a consortium of universities, research institutions and industry based at Delft University of Technology in the Netherlands. Brazil is a leading global manufacturer of sugar-cane ethanol, and trains experts at the Brazilian Bioethanol Science and Technology Laboratory in Campinas, among other institutions. In the United States, graduates with prospects are emerging from the Energy Biosciences Research Institute (EBI) at the University of California, Berkeley, which was launched with a $500-million ten-year grant from the oil giant BP in 2007. Other up-and-coming institutions include the Great Lakes Bioenergy Research Center, a joint venture between Michigan State University, the University of Wisconsin–Madison and others. The centre is one of three established by the US Department of Energy in 2007.

The key to navigating the field, says Ren, is to shift goals to fit new opportunities that might come up — such as her move into business development.

PLANNING FOR THE FUTURE

The future of the biofuels job market will depend on oil prices and how government policies evolve. "There is a lot of market uncertainty," says James McMillan, a biochemical engineer at the National Renewable Energy Laboratory (NREL) in Golden, Colorado.

Some government initiatives are sending positive signals. In January, for instance, the US Department of Agriculture announced new loan guarantees for commercial-scale plants — dependent on the recipients meeting pre-specified targets. Beneficiaries included Coskata, a start-up based in Warrenville, Illinois, which received $250 million to build a refinery that will make 200 million litres of ethanol a year out of wood chips; and Enerkem of Montreal, Canada, which got $80 million for a plant to make 38 million litres of ethanol a year out of municipal solid waste. The US military has ambitious targets for how much alternative fuel it will use in the future, and is a driving force behind research into algae-derived biofuels — an endeavour that has spawned dozens of companies. And in March, the Brazilian government announced US$600 million in loan guarantees and other funding for developers of biofuel technologies.

All this government support is creating new jobs. For example, Enerkem is currently hiring mechanical, environmental and electrical engineers, analytical chemists and a business-development director.

But government backing is not assured for the future. McMillan says that researchers at NREL wonder whether the US Congress will cut programme budgets in the next few years. And policies do not ensure a stable market price for the fuel that will emerge from the nascent industrial-scale facilities, says Jeff Passmore, a consultant and past chair of the Canadian Renewable Fuels Association in Ottawa.

Still, investors have been steadily financing smaller-scale efforts, creating jobs in research and development. Venture capital and private equity investment in advanced biofuels worldwide hit $525 million in 2010, according to

K. STEPNITZ/MSU COMMUNICATIONS

Bruce Dale advises biofuels hopefuls to get interdisciplinary experience.

Caroline Taylor, a bioenergy analyst at the EBI. That is up from $410 million during the recession in 2009. (Total investment spiked at $900 million in 2008, up from $300 million in 2007.) These numbers do not include the hundreds of millions of dollars spent each year by large companies such as DuPont, ExxonMobil and the Brazilian energy giant Petrobras.

NEW DIRECTIONS

Emerging business models could buffer biofuels companies financially, helping to lure investors and foster job opportunities. Last December, the trade journal *Biofuels Digest* ranked the leading bioenergy companies for 2010–11, and reported that 11 of the top 25 also had plans in place to produce high-value industrial chemicals. At the top of the list was Amyris of Emeryville, California, which opened its first industrial-scale facility in April, in Piracicaba, Brazil. The plant will turn sugar-cane syrup into farnesene, an industrial chemical that can be used to make cosmetics, perfumes and industrial lubricants. Farnesene is currently derived mainly from petroleum, and sells for around $265 per litre, according to a spokeswoman from Amyris, or seven times that for small quantities. The company is now hiring dozens of people, from computational scientists to fermentation specialists.

The ultimate vision for some of these companies is to build a 'biorefinery' to produce a range of products, from fuel to high-value, small-volume chemicals. Even if the technology does mature to the point at which such facilities can be built routinely, opportunities for scientists will continue to emerge, assuming that companies in the area maintain research and development units in the same way that the petroleum industry continues to innovate today, says McMillan.

Experts advise researchers entering the field to do their homework — many companies over-hype their technology or suffer from poor management, warns McMillan. Applicants who can network with established researchers in the field, and understand the science outside their own immediate areas of training, will be well placed to assess opportunities in a rapidly changing field. "If we are going to build up this industry, we are going to have to move a phenomenal amount of biomass around," says Dale. "In ten years, this field is going to look a lot different."

A desire to make a difference may be the defining feature of biofuels researchers as the field evolves. Many practitioners believe that they can help to wean the world off oil, and aid the environment and society in the process. "We are motivated by this big vision to get renewable energy on board," says Ren. "I dream about my work at night." ■

Charlotte Schubert *is a freelance journalist based in Seattle, Washington.*

COLUMN

Oh, Canada ...

A tax-law change has dealt a heavy blow to Canadian postdocs, argues **Lucie Low**.

As a newly appointed postdoc, I was excited, nervous and enthusiastic about moving to Canada last July. I had just completed a PhD, which is no small task in itself. I had applied for, and been awarded, an international scholarship to move from Britain to Canada and work in a world-renowned brain-imaging laboratory. The salary would apparently be tax-free. "Wow," I breathed to myself, "this is actually happening, this really is my dream come true!"

Fast-forward 11 months, and how things have changed. Postdoc salaries are no longer tax-free and I have learnt that the status of postdocs in Canada is generally 'undefined' (see *Nature* doi:10.1038/news.2010.429; 2010). All because last year's budget clarified the rules for tax credits, removing the 'student' loophole that gave some postdocs tax-free salaries.

Cue a hefty pay cut (I suppose I'll pay back my student loan when I win the lottery), mounting frustration over the paucity of resources and respect given to postdocs and a growing disillusionment with the whole situation. Oh, Canada! Country of maple syrup, lumberjacks and mounties! You could have got it so right. But you didn't.

Taxes are fine if they mean you get certain benefits — annual leave, for example. Then there's access to a pension scheme and maternity pay. Unfortunately, Canadian postdocs now find themselves paying full staff taxes, but still ineligible for these benefits. In many Canadian institutions, postdocs are classed as 'trainees', which seems to be a catch-all to describe being neither a student nor a staff member. This means that we are easily ignorable, often dealt with by departments more tailored to graduate-student issues — departments without the time, energy or resources to work out the complexities of postdoctoral status (be it 'student', 'trainee' or 'staff').

In an April 2010 letter to the Canadian Association of Postdoctoral Scholars (CAPS), the Canada Revenue Agency justified the use of 'trainee' as a term by arguing that postdocs are similar to "apprentices, articling students, and medical residents". (CAPS had requested clarification on the tax laws.) This would be fine if postdocs were getting the same salaries as, say, articling students (as newly qualified lawyers are called in Canada).

J. BARBER/CORBIS

But we're not. And, with the recently added tax burden, some postdocs are finding themselves in the bizarre situation of earning less than the graduate students in their labs. Taking into account that the average age of a postdoc in Canada is 33 (according to a CAPS poll) and that 48% of postdocs have dependents, this pay cut places a heavy burden on those with mortgages, children and other responsibilities. Now there's an incentive to complete your thesis.

I realize that few people become scientists for the money. Those keen on big salaries usually seek to become lawyers or surgeons. I'm also aware that postdocs aren't all equitable elsewhere. But if this inequitable treatment of Canadian postdocs continues, Canada will lose some of its brightest minds. My advice for now? Weigh up the pitfalls and limitations of your situation before considering a Canadian postdoc.

Still, all is not doom and gloom. There are countless opportunities to help improve the situation — by setting up institutional postdoc associations, working with CAPS or liaising with your institution to make it aware of the problems — which is exactly what I'm doing. *Vive la révolution*!

So I do see the benefits of my move, and of living and working in Canada. And for now, I am learning to ice skate and enjoying Canada's maple syrup. I'm also, I'm afraid, educating myself about the nuances of Canadian tax laws. ■

Lucie Low *is a postdoc in neuroscience at McGill University in Montreal, Quebec.*

Caroline Taylor, a bioenergy analyst at the EBI. That is up from $410 million during the recession in 2009. (Total investment spiked at $900 million in 2008, up from $300 million in 2007.) These numbers do not include the hundreds of millions of dollars spent each year by large companies such as DuPont, ExxonMobil and the Brazilian energy giant Petrobras.

NEW DIRECTIONS

Emerging business models could buffer biofuels companies financially, helping to lure investors and foster job opportunities. Last December, the trade journal *Biofuels Digest* ranked the leading bioenergy companies for 2010–11, and reported that 11 of the top 25 also had plans in place to produce high-value industrial chemicals. At the top of the list was Amyris of Emeryville, California, which opened its first industrial-scale facility in April, in Piracicaba, Brazil. The plant will turn sugar-cane syrup into farnesene, an industrial chemical that can be used to make cosmetics, perfumes and industrial lubricants. Farnesene is currently derived mainly from petroleum, and sells for around $265 per litre, according to a spokeswoman from Amyris, or seven times that for small quantities. The company is now hiring dozens of people, from computational scientists to fermentation specialists.

The ultimate vision for some of these companies is to build a 'biorefinery' to produce a range of products, from fuel to high-value, small-volume chemicals. Even if the technology does mature to the point at which such facilities can be built routinely, opportunities for scientists will continue to emerge, assuming that companies in the area maintain research and development units in the same way that the petroleum industry continues to innovate today, says McMillan.

Experts advise researchers entering the field to do their homework — many companies over-hype their technology or suffer from poor management, warns McMillan. Applicants who can network with established researchers in the field, and understand the science outside their own immediate areas of training, will be well placed to assess opportunities in a rapidly changing field. "If we are going to build up this industry, we are going to have to move a phenomenal amount of biomass around," says Dale. "In ten years, this field is going to look a lot different."

A desire to make a difference may be the defining feature of biofuels researchers as the field evolves. Many practitioners believe that they can help to wean the world off oil, and aid the environment and society in the process. "We are motivated by this big vision to get renewable energy on board," says Ren. "I dream about my work at night." ■

Charlotte Schubert *is a freelance journalist based in Seattle, Washington.*

COLUMN

Oh, Canada ...

A tax-law change has dealt a heavy blow to Canadian postdocs, argues **Lucie Low**.

As a newly appointed postdoc, I was excited, nervous and enthusiastic about moving to Canada last July. I had just completed a PhD, which is no small task in itself. I had applied for, and been awarded, an international scholarship to move from Britain to Canada and work in a world-renowned brain-imaging laboratory. The salary would apparently be tax-free. "Wow," I breathed to myself, "this is actually happening, this really is my dream come true!"

Fast-forward 11 months, and how things have changed. Postdoc salaries are no longer tax-free and I have learnt that the status of postdocs in Canada is generally 'undefined' (see *Nature* doi:10.1038/news.2010.429; 2010). All because last year's budget clarified the rules for tax credits, removing the 'student' loophole that gave some postdocs tax-free salaries.

Cue a hefty pay cut (I suppose I'll pay back my student loan when I win the lottery), mounting frustration over the paucity of resources and respect given to postdocs and a growing disillusionment with the whole situation. Oh, Canada! Country of maple syrup, lumberjacks and mounties! You could have got it so right. But you didn't.

Taxes are fine if they mean you get certain benefits — annual leave, for example. Then there's access to a pension scheme and maternity pay. Unfortunately, Canadian postdocs now find themselves paying full staff taxes, but still ineligible for these benefits. In many Canadian institutions, postdocs are classed as 'trainees', which seems to be a catch-all to describe being neither a student nor a staff member. This means that we are easily ignorable, often dealt with by departments more tailored to graduate-student issues — departments without the time, energy or resources to work out the complexities of postdoctoral status (be it 'student', 'trainee' or 'staff').

In an April 2010 letter to the Canadian Association of Postdoctoral Scholars (CAPS), the Canada Revenue Agency justified the use of 'trainee' as a term by arguing that postdocs are similar to "apprentices, articling students, and medical residents". (CAPS had requested clarification on the tax laws.) This would be fine if postdocs were getting the same salaries as, say, articling students (as newly qualified lawyers are called in Canada).

J. BARBER/CORBIS

But we're not. And, with the recently added tax burden, some postdocs are finding themselves in the bizarre situation of earning less than the graduate students in their labs. Taking into account that the average age of a postdoc in Canada is 33 (according to a CAPS poll) and that 48% of postdocs have dependents, this pay cut places a heavy burden on those with mortgages, children and other responsibilities. Now there's an incentive to complete your thesis.

I realize that few people become scientists for the money. Those keen on big salaries usually seek to become lawyers or surgeons. I'm also aware that postdocs aren't all equitable elsewhere. But if this inequitable treatment of Canadian postdocs continues, Canada will lose some of its brightest minds. My advice for now? Weigh up the pitfalls and limitations of your situation before considering a Canadian postdoc.

Still, all is not doom and gloom. There are countless opportunities to help improve the situation — by setting up institutional postdoc associations, working with CAPS or liaising with your institution to make it aware of the problems — which is exactly what I'm doing. *Vive la révolution*!

So I do see the benefits of my move, and of living and working in Canada. And for now, I am learning to ice skate and enjoying Canada's maple syrup. I'm also, I'm afraid, educating myself about the nuances of Canadian tax laws. ■

Lucie Low *is a postdoc in neuroscience at McGill University in Montreal, Quebec.*

VARIANTS

Evolution in action.

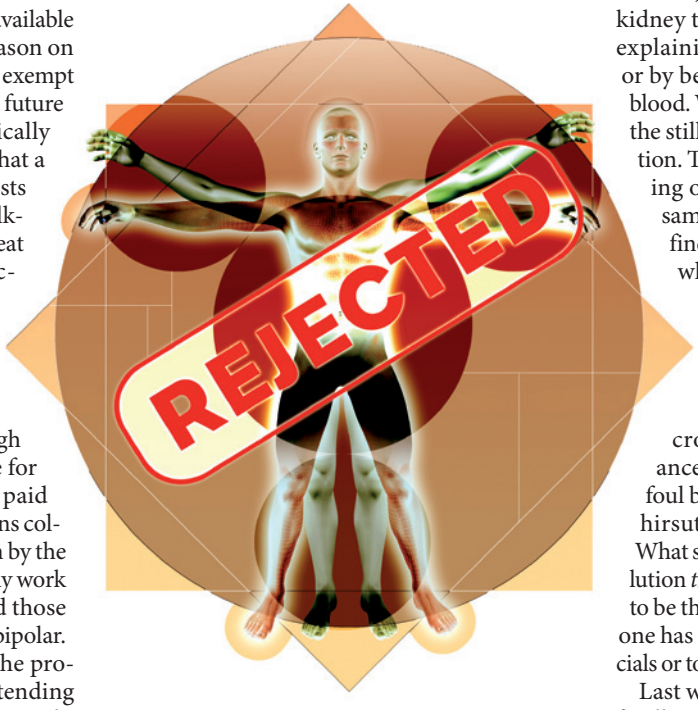
BY JOÃO RAMALHO-SANTOS

When the albinos started coughing up blood, no one much noticed. We were too busy getting the colony going. Building greenhouses; spreading MultiManure™; putting thawed goat and pig embryos into Accelerator™ incubators; assembling the drills. The Agency contract was very precise on the time window: self-sufficiency and profitable minerals available within six months, or one more season on BasicSupport™ rations, full non-tax exempt price, deductible with interest from future mining revenue. Which optimistically presumed there would be any — that a couple of outsourced XenoGeologists somewhere knew what they were talking about. Truthfully, it was the threat of another year on metallic BasicSupport™ aftertaste that spurred us on. Full price and uncertainty we could live with.

Three days later the matchstick couple started fading, and some were actually pleased. Although they worked for four, they also ate for seven and talked for ten. We only paid attention when the mauve-eyed twins collapsed, followed in quick succession by the bald cousins, the kids who could only work in the shade, all the asthmatics and those suspected of being some variety of bipolar.

The android doctors followed the program, tried to keep the peace, pretending all was as expected. But we could see right through them. Literally — they came in that translucent model designed to stand out in human company. The androids did look weird, but I guess we would too, if dozens of our charges started oozing blood out of every orifice; anaemic, wasting away, refractive to any known treatment; victims of yet another fascinating bug for someone to decipher after the fact. Although they were programmed to serve and protect, we (by now a handful of mostly grave-digging survivors) somehow presumed the androids were to blame. So we bashed their heads in as they stood grinning, bewilderment in their translucent faces. Liberating mayhem, but not a smart thing to do, the Agency would have to be reimbursed, and people had actually stopped dying, we knew the rest would remain alive.

The TruColony Program started out very scientifically, after the first missions comprising only athletic alpha-males and females failed at spectacular rates. The Agency had Xeno-everything in-house back then, XenoBiologists, XenoEpidemiologists. Recruitment took years of painstaking research, each new possible mining site thoroughly screened in terms of what the ideal genetic make-up of future colonists should be, who would adapt best, bring in the most profits. Avid oxyhaemoglobin variants for rarefied atmospheres, low metabolism for barren rocks, over-expression of mitochondrial uncoupling proteins for heat generation in frozen tundra. It was a great idea, the main problem being that, urban myth rumours of X-Men genes notwithstanding, no inherited trait can protect against raging volcanoes, assorted unknown plagues and indiscriminate death in general. For each site nailed perfectly, 7.46 others were a disaster.



JACEY

The Xenos were the first to be downsized, followed by the human medical staff. Then some MBA whizz-kid who had probably just switched from Cartoons to the Discovery Channel suggested that instead of finding the right colonization solution from the start, the Agency should just cheaply recruit crews so genetically diverse as to shame Babel. Dump them on the site with minimal preparation and let evolution do all the work. Even considering inevitable lawsuits and human-rights backlash, the desperation of the destitute always trumps careful plans, fuelling a healthy bottom line. Survival of the fittest all around.

According to the outsourced genotyping and 'omics analysis, our haematopoietic stem cells renew and differentiate at twice the normal rate, our detoxifying liver enzymes and kidney transporters are through the roof, explaining why we don't die of anaemia or by being poisoned by our own rotting blood. We are also missing some rare sugar the still-unknown parasite uses for infection. The diagnoses allowed for the flowing of new recruits carrying the exact same gene variants; it's much easier to find a pattern when you already know what it looks like. Except only some newbies survive, and, from the safe distance of their computer screens, the Agency staff is at a loss. What did haematopoiesis have to do with crossed eyes and crooked teeth, how is renal clearance related to dwarfism, adult acne, foul body odour, pathological anorexia, hirsutism or lack of facial symmetry? What sort of fitness is that? What was evolution *thinking*? Although this is rumoured to be the most successful site in history, no one has suggested us for Agency commercials or to dazzle embedded colony reporters.

Last week the Agency tacitly gave up by finally granting breeding rights, closing the gene pool. "Well," one onscreen representative couldn't quite help herself, "this is going to be one ugly colony."

I guess we could have explained that the aesthetics appreciation gene of theirs must not be universally shared. That presiding over broken androids and graves, we had pretty much felt just how the Agency had predicted we should feel. Winners, the last Noah's Ark gladiators left standing. We think of ourselves as being quite fit. ■

João Ramalho-Santos *is at the Center for Neuroscience and Cell Biology and the Department of Life Sciences, both at the University of Coimbra, Portugal. As a backup plan he owns 25% of the Comic Book Store Dr Kartoon and writes about Comics in* JL-Jornal de Letras. *He tips his hat to Tod Browning's* Freaks. *Some of his short stories can be found in English on LabLit.com.*

NATURE.COM
Follow Futures on Facebook at:
go.nature.com/mtoodm

PHOTOLIBRARY.COM

Lignocellulosic biomass, the woody parts of plants, can be enzymatically digested to yield chemical components that can be used to make biofuels.

LIGNOCELLULOSE

A chewy problem

The inedible parts of plants are feeding the next generation of biofuels. But extracting the energy-containing molecules is a challenging task.

BY KATHARINE SANDERSON

A potentially vast source of renewable energy sits just out of our grasp. Each year, more than 40 million tonnes of inedible plant material, including wheat stems, corn stover (the stalks and leaves) and wood shavings from logging, are produced — much of which is thrown away. Turning these discarded, woody bits of plants into 'second-generation' biofuels — which will not take food away from a hungry planet — has huge appeal.

The woody material that gives plants their rigidity and structure comprises three main types of carbon-based polymer — cellulose, hemicellulose and lignin — collectively called lignocellulosic biomass. When taken apart, these polymers yield chemical components that can be used to make biofuels. Cellulose, after all, is a polymer of glucose. And if this sugar can be extracted, it can be fermented to make ethanol or the longer-chain alcohol butanol. Hemicelluloses are polymers of various sizes that incorporate a range of different sugars, whereas lignin has a polymer backbone made from phenolic groups, which are ring-shaped, carbon-based structures. Other useful chemicals such as furans — molecules with a circular structure consisting of four carbon atoms and an oxygen atom — can be pulled out of lignocellulosic biomass and could serve as alternative, high-energy-density fuels. Most of the current efforts in second-generation biofuel production focus on ethanol, with furans and butanol at earlier stages of development.

The problem is that plants hold onto these chemicals with a tight grip. The glucose polymer chains in cellulose are largely insoluble and exist in crystalline microfibrils that make the sugars hard to reach (see 'From the field to the pump'). These cellulose microfibrils are attached to hemicellulose, which contains a variety of sugars, making it more complicated to convert to a single product such as ethanol. Surrounding all this is lignin, which protects the cellulose and hemicellulose. Lignin is a complex mess of polymers that are cross-linked to each other. The strong bonds that hold lignin's polymers together make it very difficult to break down. Added to that, the composition of lignin varies from plant to plant, and the true structure of this sturdy material remains unknown.

At present, the best way to break apart these lignocellulosic materials and extract their chemicals for fuel production involves heat and strong chemicals. This is a complex process: once it has been mechanically ground up, the woody biomass requires pretreatment using heat, acid or ammonia to rip apart the lignin and expose the cellulose and hemicellulose inside. Enzymes can then penetrate the biomass and liberate the sugars, which are

NATURE.COM
For latest research on using lignocellulose in biofuels
go.nature.com/4C9jkR

then fermented and distilled to produce ethanol. For now enzymatic biofuel remains at pilot scale. The enzymes so far discovered are not very efficient, to employ these techniques on a large scale is unsustainable — the existing process consumes more energy than is contained in the molecules it releases.

The good news is that there are alternative sources of enzymes. Some creatures happily survive on a diet of wood. So perhaps humans just need to find out how these enzymatic processes work and then apply them in an industrial setting.

THE TROUBLE WITH GRIBBLES

There are many potential sources of enzymes that digest wood. Some researchers are examining enzymes from the microbes found in termite guts or from wood-decomposing fungi that thrive on tree trunks. For example, biotechnology company Dyadic, based in Jupiter, Florida, has developed a designer fungus that it claims can inexpensively produce enzymes that rapidly digest lignocellulose. The enzyme technology developed by Dyadic is being used by Spanish energy company Abengoa in its Salamanca-based production facility, with a capacity of 5 million litres a year, to produce ethanol from corn stover and wheat straw.

At the University of York in the United Kingdom, plant cell-wall biologist Simon McQueen-Mason is investigating the marine wood borer *Limnoria quadripunctata*, also known as the gribble. These small crustaceans have long caused havoc in ports and harbours, eating away at moorings, piers and anything else made of wood. "This animal is very different from other animals that eat wood," says McQueen-Mason. "It has a gut devoid of microbial life." This means that the gribble — unlike the termite, which uses its intestinal bacteria to digest wood — must be able to secrete the enzymes needed to convert the wood into its constituent sugars.

McQueen-Mason and his colleagues are working to determine which genes allow gribbles to produce the correct mixture of cellulases — enzymes that can break down cellulose. So far, they have identified about 60 genes involved in production of glycosyl hydrolases, one type of cellulase. The gribble's cellulases seem to belong to three or four families, with each group working on a different part of cellulosic structure. It is not clear how gribbles deal with lignin, although a group of proteins called haemocyanins might be responsible. These enzymes are highly expressed in the gribble's digestive gland and have also been found in its gut, suggesting that they might be involved in digestion.

There is a lack of brave investors willing to take a gamble on a new technology.

It will take about three years for McQueen-Mason's team to fully investigate the gribble's digestive enzymes. In order to determine which enzymes are best at releasing sugars from unprocessed biomass, the researchers will work with enzyme firm Novozymes, headquartered in Bagsvaerd, Denmark. "It doesn't really matter where the enzymes come from, as long as we are able to produce them cheaply," says Novozyme's Claus Fuglsang.

Scaling up from the lab to industrial settings will not be simple, however. Digesting lignocellulosic material involves many enzymes for cleaving cellulose in different places. So, when a promising new enzyme is found, it must be tested in a range of enzyme mixtures to optimize the overall reaction. Improving the enzyme mixture is a priority says Fuglsang, who leads cellulosic ethanol research at Novozymes' laboratories in Davis, California: "With better enzymes comes the possibility of lowering chemical load and pretreatment."

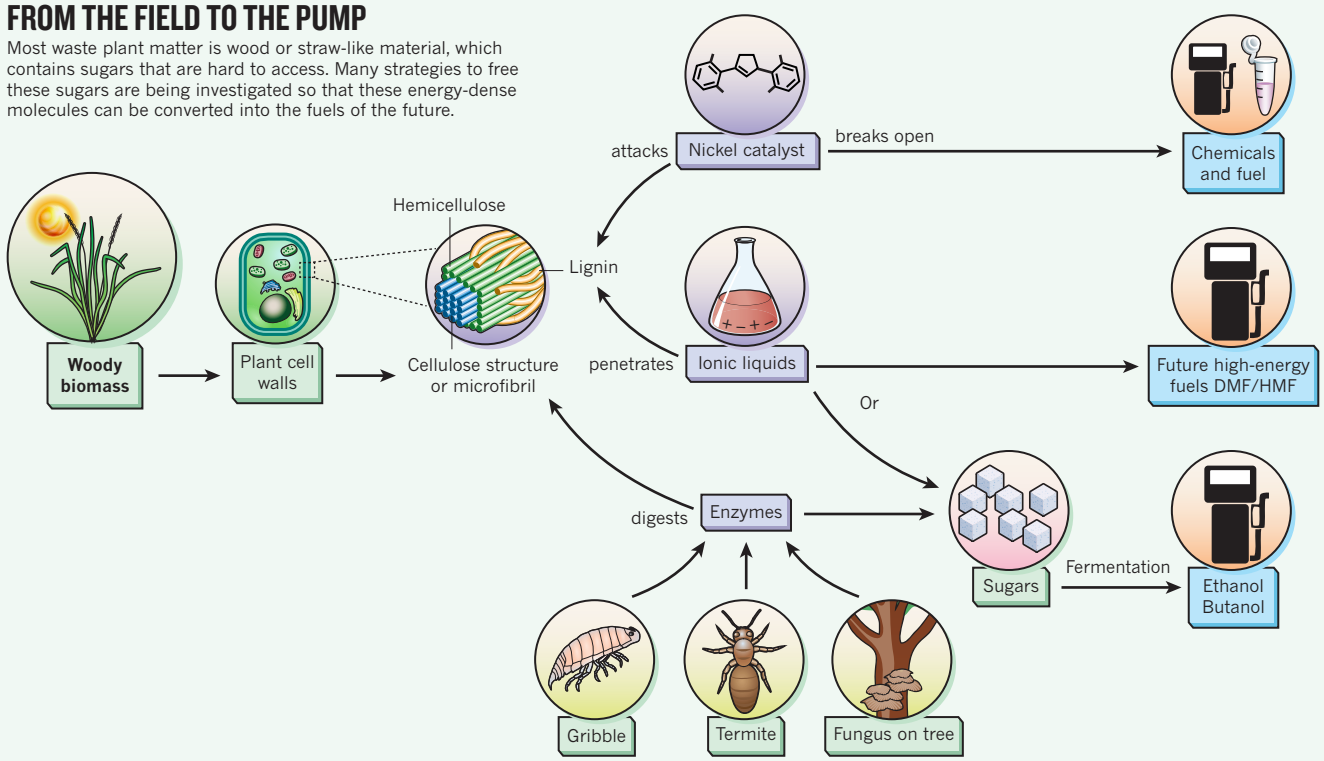
CHEMICAL SCALE

Chemical pretreatment processes are also in need of an overhaul. One weapon in the chemist's arsenal is a solvent called an ionic liquid. These are salts that become liquid at room temperature or just above, liberating their ions. They can penetrate lignin and liquify biomass, although exactly how is not yet fully understood. Chemist Brad Holmes and his colleagues at the Joint BioEnergy Institute in Emeryville, California, are using ionic liquids to dissolve crystalline cellulose and then reconstitute it as an easier-to-handle amorphous solid.

But the critical step in this process —

FROM THE FIELD TO THE PUMP

Most waste plant matter is wood or straw-like material, which contains sugars that are hard to access. Many strategies to free these sugars are being investigated so that these energy-dense molecules can be converted into the fuels of the future.



recovering the sugars from the ionic liquids — has proved challenging. Apart from the expense and chemical waste of losing some of the ionic liquid, there is another problem: if any ionic liquid is left in the sugar mixture, it prevents the enzymes from functioning in the subsequent fermentation step. Holmes' group is investigating boronic acid as a possible solution to getting the sugars out without destroying the ionic liquids, but the technology is still young. It probably will not be available on a large scale for ten years, he says.

Chemistry's role does not end with pretreatment. Processing plants for biofuels will need to be huge, requiring massive amounts of enzymes. Consequently, a purely chemical pathway to lignocellulosic biofuels may turn out to be simpler than an enzymatic route, says enzymologist-turned-chemist Ronald Raines, who works at the University of Wisconsin-Madison. "I have respect for enzymes, but they can be quite fragile," he says. "And they are certainly expensive." A chemical approach to the problem could take advantage of the accumulated wisdom of the chemical industry. "We know how to do things with chemistry on a very large scale," says Raines.

Raines' research group has developed a system that uses an ionic liquid consisting of N,N-dimethylacetamide and lithium chloride, which can dissolve cellulose without altering its chemical structure. This process can produce useful fuels from plant biomass in a single step at temperatures less than 140 °C (ref. 1).

The fuels that Raines' group are going after are from a different biofuel family from ethanol: the furans. With the help of a chromium catalyst, Raines can retrieve a chemical called 5-hydroxymethylfurfural, or HMF, which can, in turn, be used to make the promising petrol (gasoline) replacement fuel DMF, or 2,5-dimethylfuran. DMF can also be blended into fuel in much the same way as ethanol is now, and with around a 40% higher energy density than ethanol — a similar level to petrol — it has great potential as a biofuel. "We're very excited about DMF," he says.

Despite Raines' enthusiasm, DMF must undergo rigorous studies before it could be accepted as a biofuel. Safety is a huge concern, as DMF is known to be toxic to the nervous system, and the implications of DMF emissions aren't fully understood. Initial tests on DMF have been carried out and are promising: DMF has small-sized particulate matter emissions comparable to commercial petrol and a similar combustion performance[2].

In another approach, chemistry is also providing alternatives to the enzymes that serve as nature's catalysts. Chemists John Hartwig and Alexey Sergeev of the University of Illinois in Urbana-Champaign recently reported on a nickel catalyst that selectively removes oxygen atoms from lignin model compounds in precise places without destroying the entire molecule[3]. The less oxygen lignin has, the better it will work as a fuel.

The nickel catalyst works by breaking the carbon–oxygen bond that protrudes from related carbon-based alkyl ether, aryl ether and diaryl ether molecules, substituting a hydrogen atom for the oxygen but leaving the rest of the arene molecule undisturbed. Previous attempts to selectively break this carbon–oxygen bond have used high temperature and pressure, which resulted in the hydrogen atom attaching at other positions on the arene ring, creating a suboptimal and inefficient mixture of compounds. Using the nickel catalyst avoids this problem. However, the catalyst is currently in a soluble form, which makes it tricky to later separate the catalyst from the reaction products. Hartwig is working on developing a solid version of the catalyst so that liquids could flow over or through it, making separation much easier. The system also needs to be tested on naturally occurring lignin. In addition, to improve the sustainability credentials of this technology, Hartwig is looking for renewable or biomass sources for the hydrogen used in the reaction.

Until these technologies become a reality, practical applications focus on enzymes and ethanol. In April this year, the chemical company Mossi & Ghisolfi, headquartered in Tortona, Italy, announced that it had started work on the world's largest cellulosic ethanol plant — a commercial facility in Crescentino, in northwestern Italy, that is set to be operational by 2012. The plant will have the capacity to make about 50 million litres of cellulosic ethanol a year using Novozymes' enzymes. To put this in perspective, it is twice as much as the 2011 target set by the US Environmental Protection Agency for the production of cellulosic ethanol for blending into fuel.

For chemical plants such as this one, there is more to consider when using new technologies than the nitty gritty of chewing up lignin or cellulose. Such issues are at the forefront of Katherine Smart's mind as the programme leader of LACE (lignocellulosic conversion to ethanol), an industry–academia collaboration based at the University of Nottingham, UK. In addition to research into topics such as the digestion of wheat straw, LACE investigates the economics and social and ethical issues raised by the introduction of cellulosic ethanol. LACE works closely with several industrial partners, including oil company BP, British Sugar and the Dutch life and material sciences company DSM, and is optimizing processes for producing ethanol from various woody wastes.

For lignocellulosic biofuel production, there are obviously questions about sourcing the required materials. For Smart, one of the biggest challenges is logistics: transporting the huge quantities of enzymes and biomass feedstock, or starter material, that are needed to the production plant. It might be necessary, for example, to incorporate on-site enzyme production. Fuglsang argues that the entire infrastructure for biomass-derived ethanol will be different from that currently used for fossil fuel distribution. He suggests that another solution is to build smaller production facilities near to the biomass in question.

The biggest hurdle faced by these second-generation biofuel technologies is not a dearth of clever science. It is the lack of brave investors willing to take a gamble on a new technology and the absence of enforced legislation that would encourage the production of cellulosic biofuels over first generation crop-based fuels. Until these problems are solved, woody plants will be able to cling tight to their stash of fuel for a little longer. ■

SIMON CRAGG AND GRAHAM MALYON, UNIVERSITY OF PORTSMOUTH

The gribble, *Limnoria quadripunctata*, secretes a digestive enzyme that wrecks wooden harbours, but might be useful in making biofuels.

Katharine Sanderson *is a journalist based in Toulouse, France.*

1. Binder, J. B. & Raines, R. T. *J. Am. Chem. Soc.*, **131**(5), 1979–1985 (2009).
2. Zhong, S. *et al. Energy Fuels*, **24** (5), 2891–2899 (2010).
3. Sergeev, A. G. & Hartwig, J. F. *Science*, **332,** 439–443, (2011).

MARCUS ROHRER 2010

Algae farms require vast water surface areas to efficiently convert sunlight into an oil used as a biofuel.

ALGAE

The scum solution

The green slime that covers ponds is an efficient factory for turning sunlight into fuel, but growing it on an industrial scale will take ingenuity.

BY NEIL SAVAGE

When you imagine the crops that will provide biofuels, what is the first image that enters your mind? A field of corn or sugar cane? Maybe you should be picturing pond scum instead.

Algae, the organisms that cover ponds with a green film and turn tides red, are a promising source of biofuels. Researchers estimate that algae could yield 61,000 litres per hectare, compared with 200 litres to 450 litres from crops such as soya and canola. And, as the price of petroleum soars, that sort of yield is drawing interest from government and industry alike. Last year, the US Department of Energy gave US$44 million to create a research consortium to advance the technology for turning algae into fuel.

Industry is also gearing up. Sapphire Energy, a renewable energy company headquartered in San Diego, California, has received more than US$100 million in private investments to develop 'green crude', as well as another US$104 million from the US federal government's 2009 stimulus package. The oil company Exxon Mobil gave a US$300 million vote of confidence to algae by teaming up with the biotechnology company Synthetic Genomics in La Jolla, California. And aircraft maker Boeing helped establish the Algal Biomass Organization to promote the creation of algal jet fuel.

That alga draws so much research attention is a testament to its compelling potential. The organisms can be grown in artificial ponds on land that's unsuitable for agriculture, so they don't have to compete with food crops for space. They can be cultivated on the surfaces of lakes or coastal waterways, or in vats on wasteland. Algae reproduce rapidly, spreading over a body of water within hours. And they can thrive on what would otherwise be considered a clean-up problem — the water from waste-treatment plants and the carbon dioxide spewed from industrial chimneys.

Most algae being explored for biofuel production are single-celled organisms that turn carbon dioxide, hydrogen and nitrogen into carbohydrates, lipids and proteins. Depriving the organisms of nutrients causes the photosynthetic mechanism to switch from growing more algae to producing lipids. After a few days, a centrifuge is used to separate the algae from the water they grow in. Breaking open the cells then allows the extraction of an oil that can be turned into a hydrocarbon-based fuel. The alga's protein and carbohydrate remnants can be processed into pharmaceuticals or used as animal feed.

➲ NATURE.COM
to read the latest research on algae biofuels
go.nature.com/WSJkIr

But what's simple to describe can be difficult to accomplish efficiently. Just ask GreenFuel Technologies, a company founded in 2001 by researchers at the Massachusetts Institute of Technology (MIT) in Cambridge. GreenFuel built a series of ever-larger pilot plants that used waste gases from power plants as a food source for oil-producing algae, and it signed a US$92 million deal to build more plants in Spain. In 2009, the company shut down because of a lack of funds, having learned that harvesting algae was more expensive than it had anticipated. A recent study of algal biofuel production by the Energy Biosciences Institute at the University of California, Berkeley, which is funded by the oil company BP, found that much work remains before the struggle for economic viability can be won. According to Nigel Quinn, an agricultural engineer at Lawrence Berkeley National Laboratory who led the study, making fuel from algae using today's technology is a money-losing proposition, unless it's done in conjunction with another process, such as treating wastewater or producing valuable by-products.

To reach the big time, algal oil production must overcome several obstacles. For one thing, there's the question of space: for photosynthesis to work, light must reach the algae. If a layer of algae is more than a few centimetres thick, the organisms on the surface shade

those underneath, blocking the sunlight. One alternative is to spread horizontally — and wide. Algae would need to cover an area of 9.25 million hectares — about the size of Portugal — to derive enough biodiesel to cover Europe's annual transport requirement of 370 billion litres, according to René Wijffels and Maria Barbosa, environmental technologists at Wageningen University's Food and Biobased Research centre in the Netherlands.

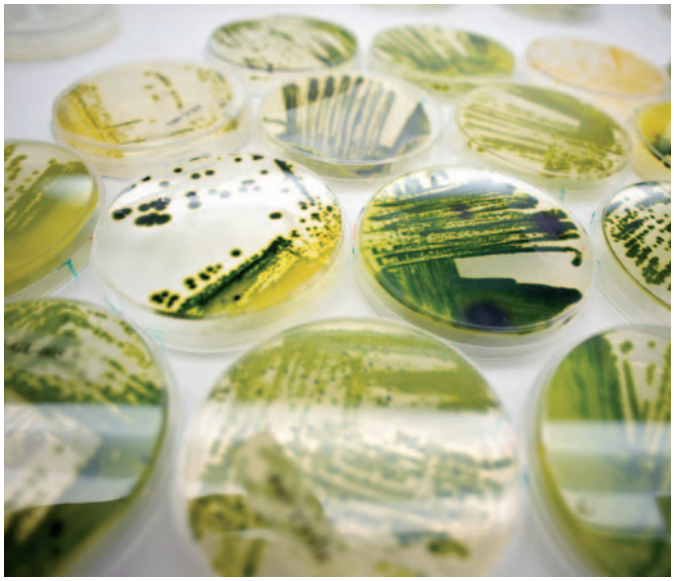
Realistically, only 5.5% of land in the United States is available to accommodate algae-growing ponds, estimates Mark Wigmosta, a hydrologist at the Pacific Northwest National Laboratory in Richland, Washington. With the current technology, that land could produce 220 billion litres of algal oil per year — equivalent to about half of the oil imported by the United States for transport each year. Furthermore, with current production processes, such a large-scale algae-growing enterprise would require roughly three times as much water as is devoted to all US agriculture, says Wigmosta. To assess whether water usage could be reduced, he looked at areas where the average levels of sunshine, precipitation and humidity would lead to more efficient algal growth: the Gulf Coast, the southeastern seaboard and the Great Lakes. He found enough land in these regions to replace about 17% of petroleum imports with biofuel, using only one-quarter of the water devoted to agriculture (roughly the same amount of water that bioethanol production requires). Wigmosta based this analysis on a system using open ponds of 30 centimetres depth and 4 hectares in area, assuming they are supplied by fresh water. Strains of algae that grow in salt water or waste water could make the equation more favourable.

It might be possible to use less water by switching from open ponds, which lose water through evaporation, to closed photobioreactors. In a typical reactor, an array of glass tubes circulates CO_2 through a mixture of algae and water; the goal is to expose all organisms to enough sunlight. But such systems — which Wigmosta says are popular in China — present their own difficulties. For example, because the reactors soak up sunlight, they need to be cooled, which often means spraying them with water, possibly cancelling out the savings made by avoiding evaporative loss.

The other input that algae need, besides water, is CO_2 — but algal cells can't efficiently tap into atmospheric CO_2 to support the rapid growth needed for a commercial operation. So algal farms might need to be situated near artificial CO_2 sources, such as coal-burning power plants. "If you have to pipe CO_2 four or five miles, the piping costs will eat you alive," Quinn says.

ADVANCED ALGAE

The renewable oils company Solazyme in South San Francisco, California, is trying to sidestep some of the problems with algal cultivation by exchanging photosynthesis for the same sort of fermentation used to produce ethanol. "The productivity is incredibly low when you grow algae in a direct photosynthesis process," says Solazyme president and chief technology officer Harrison Dillon. The company keeps its algae in the dark and feeds them sugar, which can be derived from any source. The organisms then convert the sugar into an oil. Dillon predicts that even without government subsidies, the company's biofuel will be priced competitively with petrol (gasoline). Solazyme is converting old factories to demonstrate its technology and has a contract to deliver 570,000 litres of alga-derived fuel to the US Department of Defense this year. The company hopes to be selling algal oil to commercial refineries to produce a hycrocarbon-based fuel by the end of 2013.



SOLAZYME, INC.

Algae at Solazyme are kept in the dark and fed sugar to produce an oil.

James Liao, a chemical and biomolecular engineer at the University of California, Los Angeles, also wants to move away from traditional ways of using algae. The main issue with restricting nutrients to force the organisms to make oil is that it trades off growth for oil production. Liao, instead, adds more nutrients. The result is an artificial algal bloom, which yields little oil but a lot of protein. The algae then become a feedstock for another organism, such as *Escherichia coli*, which digests the algae and produces alcohols such as ethanol and butanol. Those, in turn, can be built up into hydrocarbon-based fuels by using standard chemical processes. One advantage of Liao's proposal is its efficiency. "It's probably the fastest way to fix CO_2," Liao says. Another is that it avoids one potential problem of open ponds, the invasion of other organisms. Strains that have been genetically engineered to produce more oil may have trouble out-competing natural strains that enter the system. As a bonus, the conversion process produces ammonia as a by-product, and this nitrogen source can be used to fertilize the next round of growth.

Another possible fuel source is blue-green algae, which aren't strictly algae but bacteria of the genus *Cyanobacterium*. Whereas algal cells must be destroyed to extract their oil, cyanobacteria secrete their products. As a result, it's unnecessary to kill one batch and grow a new one, allowing continuous production. George Church, a geneticist at Harvard Medical School in Boston, Massachusetts, has engineered cyanobacteria to produce hydrocarbon molecules at the appropriate lengths for various fuels. "We're not making oils. We're making something much closer to petroleum," says Church, who cofounded Joule Unlimited in Cambridge, Massachusetts, to commercialize the technology. Moreover, Church says, tweaking the bacterium's genes will eventually make it possible for the organisms to soak up atmospheric CO_2 efficiently — an advance that would liberate fuel production from the need for an artificial source of CO_2. The company is testing its technology in a pilot plant near Austin, Texas, and expects to begin commercial production in 2012. The Joule team contends that its process will produce 140,000 litres per hectare per year.

There is plenty of room for progress. Current estimates are that biological photosynthesis can convert at most about 10% of the sunlight reaching the Earth's surface to chemical energy; Wigmosta says today's algae convert about 1.1%. Genetic engineering could create algae that produce more oil and are more efficient at converting solar energy to biomass. Engineers are working on improved designs for growth systems, such as structures that stack algae in layers for better sunlight exposure, and harvesting systems that could use microwaves or sound waves to extract the oil.

Quinn says it could easily take ten years of such research to make algal biofuels economically viable, but it's certainly possible to replace some proportion of the petroleum we use now. "We don't necessarily know what the path's going to be," he says. But "we're optimistic". ■

Neil Savage *is a science writer based in Lowell, Massachusetts.*


PHOTOLIBRARY.COM

PERSPECTIVE

Don't foul the water

Shifting from corn to perennial crops in making biofuels is essential to save clean water, argues **Jeremy Martin**.

The most important environmental impacts of biofuels occur not at a car's tailpipe or biorefinery's smokestack, but on the farms that grow the crops to be rendered into liquid transportation fuels. Biofuel production is already at a scale that is reshaping agriculture around the world. In the United States, the production of corn ethanol is polluting surface water and groundwater, while undermining existing efforts to ensure long-term agricultural productivity. Switching to next-generation biofuels made from perennial crops offers the chance to expand biofuel production while protecting water resources and building a more diverse and resilient agricultural system. Yet realizing this potential requires important changes in both biofuel and agricultural policy.

The use of corn grain for ethanol in the United States more than tripled from 2005 to 2010, and more than a third of the US corn crop now goes to ethanol facilities[1]. That trend will continue: the US government mandate for annual biofuel consumption — almost entirely corn ethanol — grows from 45 billion litres in 2010 to around 57 billion litres in 2015. This rate of growth far exceeds projected yield increases for corn. And given that demand for food is hardly standing still, the result will be consistently high prices for corn, which will in turn lead to expanded and more intensive corn production.

That could spell trouble for water resources. Although it occupies just 10% of US farmland (including cropland and pasture), corn consumes 40% of the fertilizer used in the country. Expansion of corn planting therefore will lead to an even larger leap in the amount of fertilizer applied to the soil, resulting in substantial increases in the amounts of nitrogen and phosphorus infiltrating the ground- and surface waters. Elevated levels of nitrate and pesticides have been found in shallow groundwater in more than half of the United States' agricultural watersheds. Water in more than 20% of these watersheds exceeds safe drinking water standards for nitrate[2], which is a potential risk factor for cancer, reproductive problems and methaemoglobinaemia, or 'blue-baby syndrome', which interferes with oxygen transport in the blood of infants.

Most US corn is grown in the Midwestern states. The farm runoff travels down the Mississippi River, ultimately into the Gulf of Mexico, where the influx of excess fertilizer leads to annual algae blooms that deplete the oxygen in the water, killing fish and other marine life. The dead zone that results has over the past five years extended over an area averaging more than 15,000 square kilometres.

Many of the existing programmes to reduce agricultural water pollution were developed in an era of low crop prices, before the expansion of corn ethanol, and rely on incentive payments for voluntary conservation measures. But high crop prices make the incentive payments uncompetitive. High prices also induce farmers to plant corn on erodible and marginally productive land that had previously been set aside for conservation — an agricultural expansion that further degrades water quality, since this land requires more fertilizer and is more susceptible to the loss of fertilizer through leaching or erosion. With corn prices expected to remain high, conservation programmes will continue to lose the bidding war with production.

The shift to second-generation cellulosic biofuels will help to spare water. Perennial crops such as switchgrass (*Panicum virgatum*), *Miscanthus* or fast-growing trees require less fertilizer, pesticide and herbicide than corn. Moreover, their deep year-round root systems scavenge fertilizer, so less of it makes its way into the rivers. These roots also sequester carbon (helping to mitigate global climate change) while enhancing long-term soil productivity. Research on *Miscanthus* and switchgrass at the University of Illinois found that perennial bioenergy crops reduced inorganic nitrogen leaching by more than 90% compared with a corn–soybean rotation. The study also found that these perennial crops could reduce drainage flows that cause erosion and carry pesticides to surface waters.

The most immediate obstacle to realizing the water quality benefits of perennial bioenergy crops is the absence of a cellulosic biofuel industry. The key challenge facing pioneers in this field is raising capital to build the first commercial-scale biorefineries. Supporting the startup of the first commercial facilities would cost a small fraction of the US$5 billion a year the US government now spends on tax credits for corn ethanol.

But investments in energy technology alone will not ensure that biofuels play a constructive role in agriculture. Expanded biofuel production must be accompanied by robust funding and administration of existing agricultural conservation programmes, and smart new policies that target the highest-impact opportunities to protect water resources and the environment even as production expands. Although our ambitions for biofuels may be measured against our oil consumption — and this is as true for biodiesel in the European market as for ethanol in the United States — ultimately the scale at which biofuels can be produced must be constrained by the scarce natural resources upon which we depend for food, fuel and essential environmental services. ■

Jeremy Martin *is a senior scientist at the Cambridge, Massachusetts-based Union of Concerned Scientists.*
email: JMartin@ucsusa.org

1. US Department of Agriculture. Feedgrains Yearbook. Updated September. (US Department of Agriculture, 2010).
2. US Environmental Protection Agency. 2008 Report on the Environment. EPA/600/R-07/045F (NTIS PB2008-112484) (US EPA, 2008).

FACT FOUNDATION

Workers in Garalo, Mali, press jatropha seeds that contain an oil that can be burned to produce electricity.

LOCAL BENEFITS

The seeds of an economy

Biofuels could help poor nations modernize, but scaling up aid supported projects to commercial operations is far from easy.

BY NATASHA GILBERT

After nightfall, the streets around Garalo, a small town in southern Mali, used to be pitch black. The town's 10,000 or so inhabitants had no electricity and so used paraffin (kerosene) to light their homes and places of work, and battery-powered torches to find their way around.

For years the locals had been requesting extension of the national grid to the area, but the job had proved to be too expensive. In 2006, Garalo's farmers took matters into their own hands. Realizing their cotton had ceased to earn them a decent wage, they planted those fields instead with jatropha (*Jatropha curcas*), whose seeds produce oil that can be burned to produce electricity. To achieve their ambition, Garalo's villagers teamed up with a non-governmental organization called Fuels from Agriculture in Communal Technology (FACT) Foundation (www.fact-foundation.com). This Netherlands-based foundation assists developing countries to produce and use biofuels.

Today, local farmers cultivate 600 hectares of land with jatropha, which provides electricity to 350 homes — roughly half of Garalo's population — and businesses, and also powers streets lights. Another 200 households have expressed an interest in becoming connected, says Winfried Rijssenbeek, director of the FACT Foundation.

In addition to providing light, heat and fuel for transportation, biofuels have given Garalo's businesses and tradespeople the tools and confidence to modernize and expand. One tailor, for example, bought an electric sewing machine; a furniture maker bought powered tools. Both artisans can now do their jobs faster and better and thus earn more income, says Rijssenbeek. "Electricity has a big impact on the villagers. It creates a positive attitude in the people living there," he adds.

The Garalo project is a testament to how biofuel production can greatly improve the lives of poor people in developing countries, so much so that the FACT Foundation is rolling out the model to another 10 nearby villages in order to be able to support more farmers and add another 900 hectares of jatropha. Rijssenbeek says his long-term goal is to provide biofuel-generated electricity to 100 villages.

GROUP POWER

What makes the Garalo project so successful, says Rijssenbeek, is that locals took charge of and responsibility for the initiative themselves. Garalo's farmers organized themselves into a cooperative that collects and presses the jatropha seeds. The farmers make decisions and deal with problems together, collectively guaranteeing fair prices for their crop and the electricity they buy.

To kick-start the project, several organizations, including Stichting Het Groene Woudt (a Dutch private foundation) and a Malian government department funded by the World Bank, provided funding to build a facility to extract oil from jatropha seeds and a power station to convert the oil to electricity. The power station is run by a local non-governmental organization, the Mali Folkecenter (www.malifolkecenter.org). In addition, the FACT foundation helped local businesses gain access to financial credit to take advantage of the new electricity supply.

Could projects like Garalo's be expanded across Africa to make a significant impact on alleviating poverty and improving livelihoods? It is a tempting prospect (see 'A new hope for Africa', page S20). Many land-locked African nations with poor infrastructure have little or no access to the fossil fuels on which the developed world depends. Biofuels, made from a choice of crops suitable for the region, could provide an affordable, locally grown source of fuel, and could also be traded internationally, says Tom Richard, a bioengineer at Pennsylvania State University. Large-scale production will generate an export market for biofuels from developing countries enabling farmers to "diversify their income generation," says Richard. In this way, the burgeoning biofuel industry could help poor nations to strengthen their economies, and create jobs and new streams of income for farmers.

Despite its many successes, the Garalo biofuels initiative falls well short of such a grand ambition. Its benefits are still limited to the homes and businesses now on the grid. In addition, the project is still dependent on outside funding. "It is very expensive to run small-scale distilleries," says Frank Rosillo-Calle, an honorary research fellow in energy policy at Imperial College London.

Charitable projects such as this are good first steps but are unlikely to help Africa achieve wider economic growth and energy independence. Rijssenbeek's aim that the project will run as a self-sustaining business is still a distant hope. The income generated from the sale of electricity covers operating expenses, but barely makes a dent in the costs incurred in establishing the project. "At this stage it is not

possible to run the enterprise as a commercial project," Rijssenbeek concedes.

The road to large-scale biofuel production for developing countries is bumpy, and what works as a small-scale aid project rarely translates into commercial success. The lion's share of investment in biofuel production in developing countries comes from large foreign-owned companies looking to establish industrial-scale plantations, says Laura German, a specialist in the governance of forests at the Indonesia-based Centre for International Forestry Research (CIFOR). A study[1] by German and Jan Willem van Gelder, of economic research firm Profundo, found that between 2000 and 2009, US$ 5.7–6.7 billion was invested by 10 large companies in Asia, Africa and Latin America (with Brazil accounting for just over half). However, the foreign-owned commercial biofuel projects established so far have largely failed to deliver the anticipated benefits for local populations.

German has studied commercial biofuel production initiatives in Ghana and Zambia, where foreign companies have purchased more than 1.1 million hectares and 600,000 hectares, respectively, mostly to grow jatropha[2]. Local communities and farmers will often sell their land to foreign companies in the hope of gaining employment, or they might sign binding contracts to grow jatropha for a company. But companies often reneged on contracts, German says, and the promised jobs never materialized. In one case in Zambia, farmers were contracted to grow jatropha for a fixed price only for a particular company. But as the crops began to bear fruit, "the company was nowhere to be seen," she says.

The foreign companies hold most of the cards in these situations. Locals often lack the experience and legal knowledge to negotiate favourable terms, and have little power to hold the companies to account. Consequently, many communities end up tied to long-term lease agreements of up to 50 years in which companies dictate the prices paid to farmers.

"People are desperate for investment but they are starry eyed about the benefits they could get," German warns. "Expectations are high, but there is little understanding of what is required to get those benefits."

It isn't all bad news: German and her colleagues came across a handful of companies in Ghana that were operating by the book. The companies had registered with appropriate government agencies, obtained environmental permits, and offered good pay and fair deals to locals. However, all biofuels ventures are "tarnished" in the eyes of civil society groups and human rights campaigners, who protect the rights of local people. Ironically, it is the responsible ones, who talk to local groups and keep authorities informed, that are the most heavily investigated, says George Schoneveld, a researcher in forest governance also at CIFOR. "The more visible and engaging a company is, the more they tend to come under scrutiny," he says. "The more secretive and shady ones have received almost zero exposure."

"People are desperate for investment but they are starry eyed about the benefits they could get."

ETHICAL EXPANSION

The efficiencies and economies of scale that come from commercializing and centralizing biofuels production have limited benefits for locals. In a report published in April, the Nuffield Council on Bioethics, a London-based think-tank, said that European policies to source 10% of Europe's transport energy needs from renewable fuels by 2020 are driving an "unsustainable" and "unethical" expansion of global biofuel production[3].

Nevertheless, given the potential to help tackle climate change and reduce poverty, the Nuffield report concludes that governments have a "duty" to develop biofuels. If countries can ensure that strict ethical and environmental standards are met — that is, the fuels do not infringe people's health and work rights nor contribute to net emission of greenhouse gas — then they should encourage foreign investment in biofuel schemes, says Richard Templer, director of the Porter Institute for Sustainable Bioenergy Research at Imperial College London. "If they can do biofuels properly, they should," he says.

The Nuffield authors suggest that the European Commission should establish a monitoring system to check that any biofuel consumed in Europe meets human-rights standards and to enable swift sanctions against those found to be abusing these rights by, for example, not paying a fair wage, or employing children in the production process. They also call on the United Nations Environment Programme to develop an international standard to promote "a market for environmentally sustainable and human-rights friendly biofuels," says Joyce Tait, a science-policy researcher and lead author of the Nuffield study.

Governments can also play a greater role in regulating the expansion of the biofuels industry to ensure its benefits are realized for local communities, says German. This could include requesting concrete commitments from companies and holding them accountable when these promises are broken, to ensure that people get legal compensation. Also, says German, governments could provide more financial support and incentives for the establishment of smallholder biofuel schemes and processing, such as those set up by the FACT Foundation, helping to bridge the gap between small aid projects and large-scale commercial ventures. Such international standards and regulations would help communities to realize the benefits of biofuels, encourage better practice among investing companies — and improve the reputation of biofuels.

There is time to learn lessons and make changes for the better. More projects demonstrating the positive impacts of biofuels, such as that in Garalo, will help to overcome the scepticism of those groups concerned about the negative impact of biofuels on social justice and food security, says Tom Richard.

"It's easy to think that we would do large-scale biofuels badly in the future because we do it badly now," Richard says, adding, "We must get out of this mindset." Biofuels, developed on both industrial and local scales, he says, need to be part of the options to help put the world on a more sustainable and equitable footing. ■

JOERG BOETHLING/PHOTOLIBRARY.COM

Farmers in Garalo, Mali, grow 600 hectares of jatropha to supply half the town with electricity.

Natasha Gilbert *is a reporter for* Nature *based in London.*

1. van Gelder, J. W. & German, L. www.cifor.cgiar.org (2011).
2. Schoneveld, G. *et al*. www.cifor.cgiar.org (2010).
3. Nuffield Council on Bioethics. www.nuffieldbioethics.org/biofuels-0 (2011).

nature OUTLOOK

BIOFUELS

23 June 2011 / Vol 474 / Issue No. 7352

COVER ART: NIK SPENCER

It was too good to be true: a world powered by an energy-packed liquid that we can conveniently tap from the ground. As more countries took advantage of these liquid riches it became obvious that there isn't enough crude oil to carry on indefinitely, especially as the biggest stores are in some of the least stable countries. What's more, we have begun to appreciate the huge environmental cost of releasing vast amounts of prehistoric carbon.

Since kerosene began to replace whale oil in lighting more than 150 years ago, petroleum products have come to power everything from leaf blowers to jet fighters — and we have become utterly dependent on them. This is not sustainable. We need an alternative that functions like crude oil, but without the same drawbacks; that can break our addiction to fossil fuels and reduce greenhouse-gas emissions, without risky deep-sea drilling or fear of oil spills.

Biofuels might be the answer. These fuels can be made from many different starting materials, from waste wood to algae or even genetically engineered bacteria. Such variety means a range of biofuels could be produced in disparate locations to fulfil different roles.

Biofuels have big boots to fill; expectations are high, yet initial attempts have had chequered success. The oil- and sugar-rich crops that need premium land to grow shift the problem, turning fuel shortages into food scarcity. There are other issues: production of these first-generation biofuels is water and energy intensive; few actually yield a net reduction in greenhouse-gas emissions. But these are surmountable challenges. Indeed, they are teething problems for a young technology. Next-generation biofuels are already learning to walk. It might take 20–30 years before biofuels are free to run, but it's an enterprise worth fostering. We are pleased to acknowledge the financial support of Biotechnology and Biological Sciences Research Council, BP, Ceres, São Paulo Research Foundation and US Department of Energy's BioEnergy Science Center. As always, *Nature* retains sole responsibility for all editorial content in this *Outlook*.

Michelle Grayson
Associate Supplements Editor, Nature Outlook.

CONTENTS

Nature Outlooks are sponsored supplements that aim to stimulate interest and debate around a subject of interest to the sponsor, while satisfying the editorial values of *Nature* and our readers' expectations. The boundaries of sponsor involvement are clearly delineated in the *Nature Outlook* editorial guidelines available at http://www.nature.com/advertising/resources/pdf/outlook_guidelines.pdf.

CITING THE OUTLOOK
Cite as a supplement to *Nature*, for example, *Nature* **Vol XXX**, No. XXXX Suppl, Sxx–Sxx (2010).

VISIT THE OUTLOOK ONLINE
The *Nature Outlook Cancer Prevention* supplement can be found at http://www.nature.com/nature/outlook/biofuels. It features all newly commissioned content as well as a selection of relevant previously published material.

All featured articles will be freely available for 6 months.

SUBSCRIPTIONS AND CUSTOMER SERVICES
For UK/Europe (excluding Japan): Nature Publishing Group, Subscriptions, Brunel Road, Basingstoke, Hants, RG21 6XS, UK. Tel: +44 (0) 1256 329242.Subscriptions and customer services for Americas – including Canada, Latin America and the Caribbean: Nature Publishing Group, 75 Varick St, 9th floor, New York, NY 10013-1917, USA. Tel: +1 866 363 7860 (US/Canada) or +1 212 726 9223 (outside US/Canada). Japan/China/Korea:Nature Publishing Group — Asia-Pacific, Chiyoda Building 5-6th Floor, 2-37 Ichigaya Tamachi, Shinjuku-ku, Tokyo, 162-0843, Japan. Tel: +81 3 3267 8751.

CUSTOMER SERVICES
Feedback@nature.com

SIMON MAINA/AFP/GETTY IMAGES

Despite the bounty in this Kenyan market, Africa suffers from poor infrastructure and low productivity.

PERSPECTIVE

A new hope for Africa

Bioenergy could help bring food security to the world's poorest continent, say **Lee R. Lynd** and **Jeremy Woods**.

It has widely been assumed that increased production of energy from biomass requires a sacrifice in food security, particularly for the world's poor. Yet closer scrutiny suggests that modern bioenergy — in the form of fuel, electricity or heat — could be developed in ways that actually enhance food security.

In Africa, where the incidence of food insecurity is highest, hunger persists because of multiple compounding factors: poverty, poorly developed agricultural infrastructure and support, degraded land and armed conflict. Underlying these factors is a legacy of three decades of neglect for agricultural development in the poorest countries of the world.

After the success of the Green Revolution in conquering famine in India, Pakistan and elsewhere in Asia in the 1960s, the world entered a prolonged period of plentiful and cheap food. As documented by Roger Thurow and Scott Kilman in their book *Enough: Why the World's Poorest Starve in an Age of Plenty*, this abundance shifted attention away from further agricultural development, particularly for Africa. The prevailing thinking in the West was: our food is cheap and abundant; developing countries can buy food from us; if people are hungry, we'll feed them.

Global support for agricultural development in Africahas declined from US$8 billion a year in the 1980s to less than US$3 billion, and aid from the United States has plummeted more steeply. This period also saw the introduction of the 'structural adjustment' philosophy of fiscal austerity at leading development agencies such as the World Bank. Poor countries were urged to shrink the size of government, disband inefficient public sector enterprises and let the private sector take over. But the private sector was inexperienced and undercapitalized, and both suppliers and lenders considered smallholder farmers to be too poor to be a worthy market.

The result, observes Thurow, was that agricultural schools and research institutes in Africa were woefully underfunded. Extension services, which had distributed the latest technology and advice to farmers, were neglected. Rural infrastructure collapsed. Development of storage facilities, farm-to-market roads, rural finance and commodity markets was ignored. One-third to one-half of agriculture production routinely went to waste before it could get to a market. African agricultural productivity fell further behind the rest of the world (see 'The Africa lag').

Without government support or financial mechanisms such as futures pricing and crop insurance, Africa's farmers were left bearing all the risk of an inherently risky business. Add to this inferior infrastructure, outdated knowledge and technology, and unimproved seeds, and it is clear why many found it impossible to compete and so abandoned production of surplus crops.

There are some encouraging recent signs. Prompted in part by the food price spike of 2007–2008, the world has refocused attention on global food security. Sensibly, agricultural development is replacing emergency aid as the lead strategy in fighting hunger. In addition, Africa's farmers are now being seen as indispensable if the world is to meet the challenge of roughly doubling its food production to meet predicted increases in population by 2050. Nowhere else is the potential to increase agricultural yields so great.

Consideration of the impact of bio energy on African food security has tended to focus on land competition and to overlook bioenergy's marked potential to promote rural development. However, potentially productive land is rather plentiful in much of Africa whereas lack of development is the most important underlying cause of hunger. August Temu of the World Agroforestry Centre observes: "Africa has 12 times the land area of India, similar land quality, and 30% fewer people. And yet India produces enough food to feed itself and Africa does not. The green revolution bypassed Africa primarily owing to serious organizational and institutional weaknesses, not geographically limited capacity."

The impact of bioenergy on food insecurity — and its frequent companions poverty, and underdeveloped rural areas - depends on the crop grown, the land used, the technology employed, and how the bioenergy supply chain is integrated into agricultural, social and economic systems. A range of options need to be considered.

Bioenergy could in principle be produced from inedible plants that grow on land that is not well suited for growing food. Grass has notable potential for regeneration of degraded soils, and plants such as *Agave* can grow on particularly dry land. Sugar cane requires higher quality land than grass or *Agave*, whereas annual row crops require the highest quality land.

Technology is established for electricity production from biomass, and for biofuel production from sugar cane, corn and oil seeds (e.g. soy, sunflower, oil palm, jatropha). This is not, however, the case for biofuel production from grass or *Agave* fibre. Although oil seeds are more expensive to produce and most yield less fuel per unit land than other crops, they require the least elaborate processing technology and are thus more practical to produce on a small scale. Configurations involving integrated production of food, animal feed and bioenergy are potentially attractive.

Some bioenergy supply chains would make food security worse, whereas others could make it better. Consider the multiple food-security benefits of producing bioenergy from inedible crops grown on marginal land:

- Employment, and development of marketable skills, for rural Africans who have few opportunities for either.
- Introduction of sorely needed agricultural infrastructure and knowhow.
- Improved balance of payments and currency valuation. As economic development proceeds in Africa, demand for electricity and fuel will continue to increase sharply; the cost of importing oil imposes an ever-growing burden on Africa's economies and farmers.
- Energy democratization, self-sufficiency and availability for agricultural processing. Restricted access to clean, affordable energy impedes development and food production, amplifies losses in the food supply chain and exacerbates hunger.
- An economically rewarding way to regenerate Africa's vast areas of degraded land.
- A route to advancing agriculture in Africa, largely independent of factors that have made this difficult in the case of food production. North America and Europe export large amounts of food at prices difficult for African farmers to compete with. But these regions do not export biofuels and are unlikely to do so in the future, and exporting heat and electricity is not feasible.
- Lessened conflict, which is widely recognized as both a cause and an effect of hunger and poverty in Africa. Using bioenergy to improved both food security and economic security could help replace a vicious cycle with a virtuous one.

Crops grown on higher quality land would in general realize the benefits listed above, but the situation is more complex than for marginal land because of stronger interactions with food production. These interactions are a potential asset if managed well, but represent a risk to food security if managed poorly.

Assessments of biomass production potential consistently identify Africa and Brazil as the two regions with the greatest capacity. Although each situation is unique, Africa might hope to replicate Brazil's recent success in several respects.

The past decade has seen 10% of the Brazilian population lifted out of poverty. Ethanol production nearly doubled during this period, as Brazil has emerged as an international food production and export powerhouse. About half of Brazil's primary energy supply is renewable, and sugar cane provides a third of this from 0.6% of the country's total land area in a region far from the environmentally sensitive Amazon rainforest. Residues from cane ethanol production have become an important source of electricity, and expanding their use for cattle feed is under consideration. Brazil has analysed the impacts of its ethanol industry, documenting macroeconomic benefits as well as increases in the income of labourers and in the level of education and nutrition available to their children.

Like Brazil, Africa has the geographical capacity to produce much more food and bioenergy than it requires, while using water sustainably and not compromising biodiversity. We believe that bioenergy can, increasingly over time, improve the food security and well-being of poor Africans, thus turning an assumed liability into an asset.

Three outcomes are possible. First, Africa's bioenergy potential could go undeveloped. Given the continent's pressing and persistent problems, this would be unfortunate. Second, bioenergy could be developed in ways that increase food security and prosperity for poor Africans. And third — the big concern —bioenergy could be developed in a way that benefits few people and further marginalizes many more.

The question, then, is: what needs to be done to ensure that we cut with the beneficial edge of bioenergy's double-edged sword?

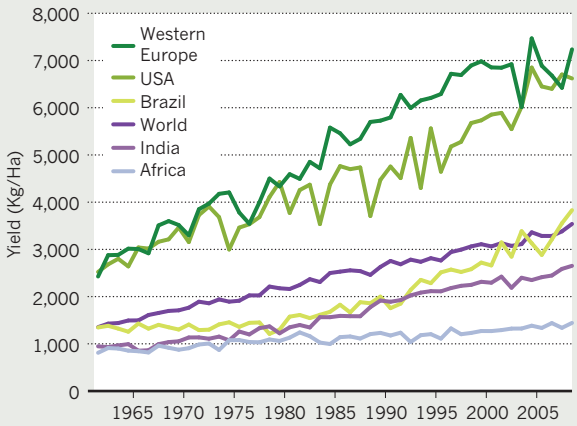
To start with, we suggest that proposed bioenergy projects in Africa be expected to demonstrably improve food security at a local level. This would move the discussion from the abstract to the concrete while honouring legitimate concern over the possibility of exploitation. To achieve this goal will require planning and monitoring. Emerging frameworks and standards for evaluating bioenergy, such as those of the Roundtable on Sustainable Biofuels and the Global Bioenergy Partnership, will need to be adapted to the African context. Private–public partnerships are in principle attractive ways to harness the economic engine of private enterprise in order to realize social benefits. Farmer cooperatives could be a way to aggregate the output and represent the interests of smallholder farmers supplying biomass processing facilities. Research is needed on biomass production, conversion technologies and societal integration, emphasizing Africa-specific aspects.

Important strategic issues include the different attributes of village-scale and industrial-scale processing, the need to implement near-term options in ways that foster rather than impede deployment of advanced technologies as they become available, and how to access export markets in ways that serve the needs and aspirations of Africans.

As humanity faces the multiple challenges of our time, our need to avoid the many approaches that do not foster sustainable outcomes is exceeded only by our need to identify and implement the (probably fewer) approaches that do. Bioenergy is a case in point. The first step towards reaching 'win–win' outcomes with respect to bioenergy, food security and poverty reduction is to recognize that such outcomes are possible. ■

THE AFRICA LAG

The green revolution largely bypassed Africa, where cereal crop yields have barely improved in 50 years.



FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, (FAOSTAT), 2011.

➲ NATURE.COM
to read the latest news on biofuels in Africa
go.nature.com/EQwUi2

Lee R. Lynd *is a professor at Dartmouth College's Thayer School of Engineering in Hanover, New Hampshire, and coordinator of the executive committee of the Global Sustainable Bioenergy project.*
Jeremy Woods *is a lecturer at Imperial College London, and codirector of the Porter Alliance.*
e-mails: lee.r.lynd@Dartmouth.edu; jeremy.woods@imperial.ac.uk

We thank R. Thurow for discussions about the causes of food insecurity in Africa, and C.H. Brito Cruz of the Sao Paulo Research Foundation for information about bioenergy in Brazil.

DANIEL DEITSCHEL/GETTY IMAGES

POLICY

Fuelling politics

Biofuels have been hailed as key to reducing our fossil-fuel dependence, yet their environmental and social impacts remain uncertain. A complex task lies ahead for policy makers.

BY MARTIN ROBBINS

In recent years a consensus of sorts has developed around energy, signalled most strongly in 2006 when George W. Bush admitted that the United States had an "addiction" to oil. Petroleum has come under attack from almost every point on the political spectrum, from people concerned about the effects of fossil fuels on the environment to right-wing patriots alarmed at their nation's dependency on a resource imported from some of the world's most troubled nations.

The effects of this addiction are especially apparent in the developing world, where petroleum often makes up a significant proportion of total imports. Conflict in Libya sent petrol (gasoline) prices soaring in Nairobi, with taxi drivers increasing their fares daily to match the increases in pump prices. In Kampala recently drivers were forced to tour local fuel stations in an attempt to find the few that hadn't run dry. Coupled with anger at rising food prices, their frustrations rapidly boiled over into riots. Oil isn't just an economic problem; increasingly it's a major security issue.

Biofuel is seen by many as the answer: a renewable source of energy that can be grown locally, exploited with existing technology and pumped straight into cars. A recent Eurobarometer report showed 83% public approval for encouraging sustainable biofuels across the European Union (see 'Incentives and targets').

Yet nobody has come up with a set of policies that can make biofuels a viable near-term substitute for petrol and diesel. Regulations, incentives and penalties are scattered across countries and often serve to shift problems from a highly regulated country to a less regulated one. Moreover, a growing collection of non-governmental organizations, scientists, politicians and even some biofuel producers have expressed concerns about the rapid and largely unchecked growth of biofuels, focusing especially on the need for adequate regulation in the developing world, impacts of land-use and food security.

BIOFUELS TARGETS

Globally, governments are providing billions of dollars in subsidies for bioenergy, but the amounts pale into insignificance next to the spend on fossil fuels. Bloomberg New Energy Finance estimates that in 2009 governments provided subsidies worth at least US$43 billion to the renewable energy and biofuel industries combined, but the International Energy Agency's figures for 2008 show global fossil-fuel subsidies of $557 billion. This imbalance places biofuels at a significant disadvantage. The "higher pricing of biofuels versus fossil fuels" is a considerable obstacle to development, says Matti Lehmus, executive vice president of biodiesel producer Neste Oil.

The most generous government support for biofuels development comes from the United States. The International Food and Agricultural Policy Council (IPC) estimates US subsidies of biofuels to be worth nearly US$7 billion a year (see 'Incentives and subsidies'), compared to nearly US$5 billion from the European Union. Financial aid includes a direct subsidy of 12 US cents per litre and tax credits for blenders worth 26 US cents per litre. Meanwhile, the US Food, Conservation and Energy Act 2008 has stimulated public investment in the form of grants and loan guarantees through the US Department of Agriculture.

Quantifying the exact level of government spending in any country is difficult, however, and estimates are often politically spun. True figures are obfuscated by a complex mixture of direct and indirect support largely inseparable from wider agricultural policy. In the United States and European Union this picture is complicated further by the variety of policies pursued by different states. Lack of transparency is also an issue, although the IPC notes that fossil-fuel policies suffer from a similar opacity.

Globally, the policy with the largest impact on biofuel production is the blending quota (see

➲ NATURE.COM
For more on how biofuels are a political matter
go.nature.com/la78Je

'Incentives and targets'). Governments in at least 17 countries have provided a huge stimulus to biofuel demand by introducing targets requiring the blending of 5–10% bioethanol with petrol, or 2–5% biodiesel with diesel fuel — mixtures that most cars and vans can run on with ease.

The US Environmental Protection Agency's Renewable Fuel Standard sets clear, long-term targets for quadrupling biofuel use in transport to 36 billion gallons by 2022 while also curbing greenhouse gas emissions. Europe is on target for biofuels to make up 10% of transport fuel by 2020, but progress varies wildly between member states (see 'Incentives and targets').

Meeting the increased demand that biofuels targets stimulate will require imports from major producers like Brazil. Both the United States and the European Union, however, have erected protectionist barriers; high tariffs on imports hamper competition and disconnect sustainability from cost. The United States, for example, imposes a 14 US cents per litre tariff on Brazilian bioethanol, even though this type of fuel is widely considered more sustainable than domestic, corn-based fuels. And Lehmus argues that if biofuels are to become a global reality, policy makers in Europe need to understand that "a significant share of its biofuels and biofuel feedstock will come from outside the EU".

The dominance of corn in the US biofuel scene has frustrated those advocating more advanced solutions. "The reason we founded Advanced Biofuels USA is that so many people we spoke to defined biofuels as corn-based ethanol, if they knew what they were at all," says Joanne Ivancic, executive director of the lobbying outfit based in Frederick, Maryland. "Very few people have heard that biofuels might be made from a wide variety of feedstock using a wide variety of technologies."

Although the United States is the world's largest producer of bioethanol, the country's biofuel policies might be hindering progress. For one thing, the United States is over-reliant on inefficient first-generation corn technology. In addition, Environmental Protection Agency policy dictates that fuel can only be 10% ethanol (recently increased to 15% for post-2007 vehicles) — a 'blend wall' that places a limit on future demand and will remain in place until the effects of more aggressive fuel blending on vehicles and the environment are clearer.

Europe has made its own policy missteps. In 2009, the European Union became the producer of the majority of the world's biodiesel (60%), yet EU lawmakers are wrestling with a growing sense that targets have been rushed through without full understanding of the wider impacts on land use and food prices. Although there is still political support for biofuels, Lehmus voices a common complaint when he says that "legislation within the EU is becoming increasingly fragmented" and bogged down in indecision. Britain, for example, requires electricity suppliers to meet renewable energy targets (which means more consumption of biomass) and favours biofuels by setting a lower road-fuel tax than is imposed on fossil fuels. But what the UK government gives with one hand, it takes away with the other: subsidies for biofuel producers were scrapped in April 2010.

With pressure from producers and advocates to scrap blend targets, Europe's politicians are due to decide what action to take on the issue of land use in July 2011 (see 'Next generation biofuels', page S2). Two serious options are on the table — adding penalties to biofuels based on their land-use impact, or raising the environmental standards that fuels must meet to qualify for support. Confounding the difficulty, European policy and existing transport infrastructure are heavily geared not for bioethanol, as is the case in Brazil, but for rapeseed-based biodiesels that require more intensive use of fertilizers. With many farmers heavily invested in rapeseed cultivation, significant change in policy will enrage many in the agricultural industry, creating a headache for European leaders.

Producers need consistent government support that levels the playing field with fossil fuels.

Developing nations have different concerns, given their lesser reliance on fossil fuels. India and China, with their huge and growing populations, face the most intense pressures on energy supply over the coming decades. China is the world's third largest producer of bioethanol, and provides around US$2 billion in direct subsidies for renewable energy alongside a programme of low-interest loans to support R&D and capital investment by producers, and an ethanol blending target of 10% by 2020. Although India's biofuel industry currently lags behind, the country has set an ambitious target to meet 20% of its diesel demand with plant-derived fuel as early as 2017.

The key problem for both countries is the scale needed to support their vast populations. China is experimenting with a variety of crops, but most of its ethanol is produced from fairly inefficient corn — a reliance that could trip political sensitivities at a time when crop prices are rising. India, on the other hand, is staking its biofuels future on jatropha (*Jatropha curcas*) — an oil-rich, perennial shrub seen by many as a perfect biofuel crop as it can grow in semi-arid or saline areas not suitable for food production. India plans to set aside 140,000 square kilometres for jatropha — more than three times the area of all UK food crops. The success of the country's jatropha policy will depend on where that land is, how productive it turns out to be and what varieties of the crop are used: it is still unclear whether the yield or oil quality will be sufficient for large-scale biofuel production.

Brazil is blessed with vast tracts of arable land in a subtropical climate, and its sugarcane has an energy density several times greater than corn. The military government of the 1970s pushed hard to make the nation energy-independent in the wake of the oil shocks of the period. As a result, Brazil today is a global leader in bioethanol production, second only to the United States. But things have not always run smoothly. Demand for sugar for food consumption competes with sugar for biofuels, a tension that has forced the Brazilian government to take heavy-handed measures: the partially state-owned producer Petrobras is forced to adjust production to prevent upsetting ethanol prices. The Brazilian model has been successful but, with its reliance on government intervention and abundant natural resources, it may not be suitable elsewhere (see 'Lessons from Brazil', page S25).

The International Energy Agency recently claimed that biofuels could meet 27% of global transportation fuel demand by 2050, but for that to happen key economic, political and technological conditions need to be met. Producers need consistent government support that levels the playing field with fossil fuels, with longer-term policies obviously providing greater certainty for investors. Ivancic points out that many of the benefits of moving away from fossil fuels are predominantly public-sector gains: for example energy security, economic development, military flexibility and climate change mitigation. Without economic incentives, she contends, "private industry can't get a return on investment equal to the value of biofuels".

THE HUMAN ELEMENT

Biofuels are ultimately a way to harness energy from the Sun. Unfortunately, many of the sunniest parts of the world are among the poorest, and lack governments able to introduce effective policies or deal with endemic corruption. Without proper regulation, biofuels could perpetuate human rights abuses.

Among producers, opinion is divided on whether the best regulatory approach is global or local. Annegrethe Jakobsen, communications manager for Danish enzyme producer Novozymes, suggests that "the best way to ensure the global sustainability of biofuels is by an ISO standard", that is, a guideline established by the International Organization for Standardization, comprising private and public institutes in over 150 countries. Similarly, the UK-based Nuffield Council on Bioethics advocates a 'Fairtrade' type of scheme, with certification for sustainable producers. Others prefer local solutions. "In an ideal world, countries would develop indigenous laws and regulations that best fit their own economic, environmental and social goals," says Ivancic.

Land-use change for biofuel production is another focal point for debate, both for increased carbon dioxide emissions and for its

potential impact on food prices (see 'Beyond food versus fuel', page S6). Some biofuels will be more sustainable than others, but the current flat-rate subsidies and blend targets tend to promote all biofuels equally, regardless of how environmentally friendly — or otherwise — they might be.

Lehmus advocates legislation that is "technology neutral and feedstock neutral, and that emphasizes the importance of emission reductions and sustainability". Similar concerns are top of the environmental advocates' agenda: Kenneth Richter, biofuels campaigner at Friends of the Earth, calls for policies "that guarantee sustainability and significant carbon savings from biofuels". Many fuels currently fall short; the Nuffield Council on Bioethics claims that only a third of biofuel used in Britain meets any kind of environmental standard, and that current EU policies for transport fuel are "unsuitable and unethical" as they neither protect the environment nor avoid human-rights abuses.

The trouble is, no-one is sure how to measure biofuel impact. The UN Environment Programme called for further research on sustainability in a 2009 report that lists many of the key issues surrounding biofuels and essentially scrawls "citation needed" next to them. Within the industry there are frustrations with existing research: recent US and EU studies on Brazilian ethanol derived emissions values varying from 3.8 g of carbon dioxide per megajoule of fuel energy to 17 g or even 46 g. "It is the same crop and the same country," Jakobsen observes. "Such different results are a clear indicator that the science is not mature."

Policy makers are all too aware of the knowledge gap, and Norman Baker, UK undersecretary of state for transport, recently conceded that "there have been shifts in biofuels policy in the past" and that future policy decisions need to be "robust and stable" to give businesses enough confidence to invest. Hampering these decisions, though, is "scientific uncertainty about the sustainability of biofuels and their wider socioeconomic impacts", said Baker. Until this is resolved, it will be a huge challenge to create policy consistent enough for investors, yet flexible enough to deal with our improving know-ledge.

For consumers, a complex network of subsidies and tariffs has obscured the connection between a fuel's sustainability and its cost. Meanwhile policy makers need better-quality research to draw on so that they can identify suitable, sustainable crops and production methods to support over the longer term and encourage investors. For everyone, better information about the consequences of biofuels use is the key to weaning us off our fossil fuel addiction. ■

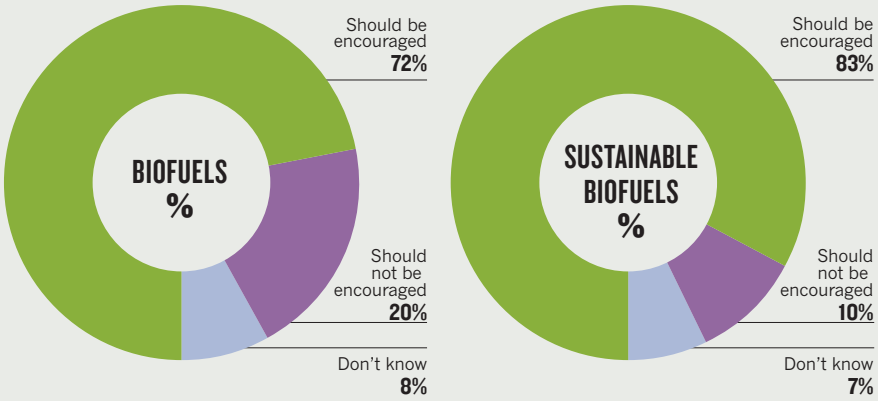
Martin Robbins *is a writer in Maidenhead, UK.*

INCENTIVES AND TARGETS

Sustainable biofuels have strong public support, but governments worldwide differ in the financial support they offer and their progress towards meeting targets.

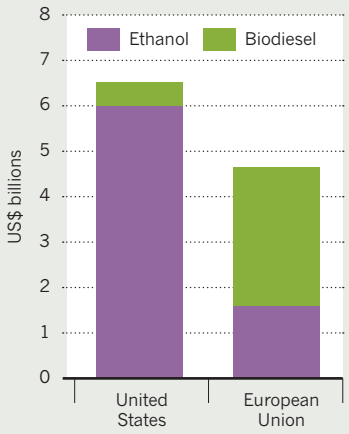
PUBLIC SUPPORT

Biofuels enjoy considerable public support within European Union nations, and sustainable biofuels even more so.



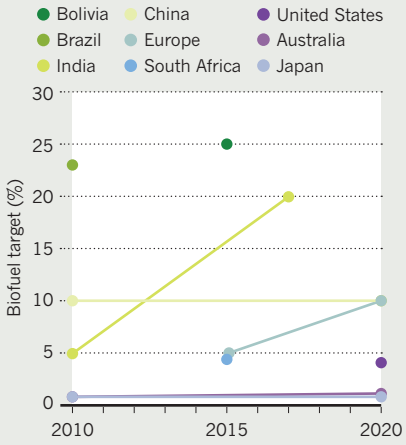
EUROPEAN AND AMERICAN SUBSIDIES

Estimated support for bioethanol and biodiesel in 2006.



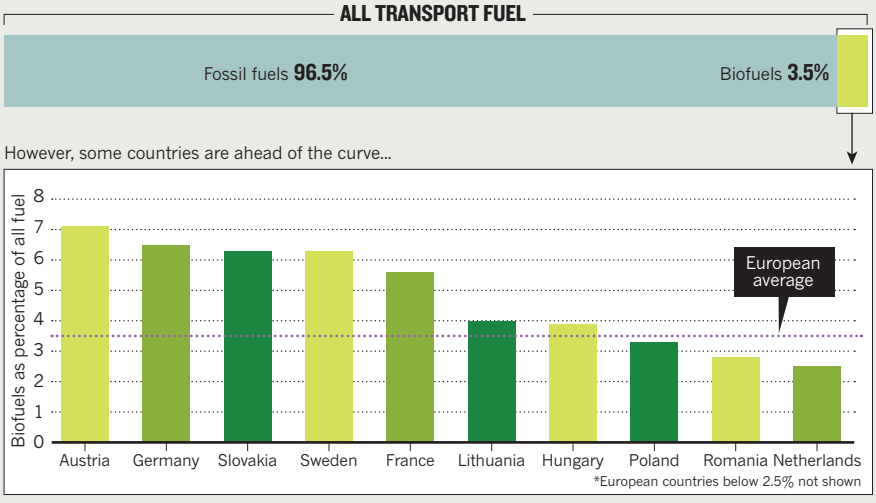
BIOFUEL TARGETS BY NATION

Many of the world's most developed countries have less-ambitious targets for biofuel blending.



EUROPEAN BREAKDOWN

Biofuels represent a small fraction of total transport fuel in Europe.



SOURCES: PIE CHARTS: EUROBAROMETER SURVEY, 2010.SUBSIDIES: GLOBAL SUBSIDIES INITIATIVE TARGETS: VARIOUS, COMPILED BY BIOENERGYWIKI BREAKDOWN: EUROSTAT.

NEWSCOM/PHOTOSHOT

Moema sugar mill in Orindiuva helps supply one of 25,000 biofuel filling stations across the country.

PERSPECTIVE

Lessons from Brazil

Thirty five years of experience has taught one of the world's leading biofuels producers several essential lessons, which other countries should heed, says **Marcia Moraes**.

There was nothing inevitable about the path Brazil took to where it is today — a country with a vibrant biofuel industry that annually turns half of its sugarcane harvest into 24 billion litres of ethanol to power 12.5 million vehicles. In 2009, nearly 1.2 million jobs in Brazil stemmed from the sugarcane, sugar and ethanol sectors. About half were held by poorly educated workers on sugarcane plantations. Sugarcane workers are among the best-paid in Brazilian agriculture, and their children have better socioeconomic outcomes than the children of other agricultural workers. Here are some of the lessons that Brazil has learnt as it has built its biofuels industry:

LESSON 1: PROVIDE STABLE GOVERNMENT OVERSIGHT

In 1975, there was no market for sugarcane ethanol. The Brazilian government had to intervene to create this market, and for decades tightly controlled it. Clear and stable rules allowed companies to make the long-term investments needed to achieve their goals.

The late 1990s saw a sharp reduction of state interference in the biofuel industry. Production quotas were abolished, as was the ethanol subsidy. Sugar and ethanol prices were left to the market. Sugarcane producers are now remunerated by processors based on sugar and ethanol prices; and ethanol producers and distributors are free to negotiate prices and quantities. The free market brought considerable efficiency gains. But planning to meet the market demand is tricky. Unlike corn and soy, sugarcane is a perennial crop, going through a six-year cycle (planting in the first year and harvesting in the next five years). Planning the supply of sugarcane means considering the estimated demand for the next five years, for both ethanol and sugar. A market economy is not always up to this task.

LESSON 2: PLAN FOR MAXIMUM FLEXIBILITY

From the outset, Brazil chose to produce two types of ethanol: anhydrous ethanol, which is blended with petrol (gasoline), and hydrated ethanol, used in vehicles that run only on ethanol and also in dual-fuel (flexible fuel) cars. State intervention was essential for the creation of the necessary infrastructure for both hydrated and anhydrous ethanol, including the installation of hydrous ethanol pumps at more than 25,000 filling stations. Today, government intervention is basically limited to determining the proportion of the anhydrous ethanol blend, setting the tax rate on sugar exports and enforcing the existing environmental and labour rules. The focus on anhydrous ethanol certainly reduces the necessary investments, aids the planning of the supply and allows the co-existence of fossil and renewable fuel — simply blending anhydrous ethanol with petrol (gasoline) in different proportions yields whatever grade fuel is desired.

LESSON 3: ESTABLISH CLEAR RULES ABOUT ETHANOL STOCKS

Ethanol is produced from an agricultural product, subject to weather conditions and the length of the growing season. So although sugarcane harvesting and ethanol production last for 6–8 months every year, ethanol is sold year round. There must be a policy for buffering stocks to avoid shortages or sharp price oscillations throughout the year.

LESSON 4: GET THE SCALE RIGHT

When Brazil launched its biofuel initiative in 1975, the Programa Nacional do Álcool encouraged the establishment of small-scale distilleries. These distilleries lacked economies of scale and most closed through economic inefficiency.

LESSON 5: PROTECT THE ENVIRONMENT

Minimizing environmental impact is imperative. Brazil regulates issues such as where sugarcane may be grown[1] — for example prohibiting production in the sensitive biomes of Amazonia and Pantanal. The government regulates use of water, use of the by-product vinasse as a fertilizer, preservation of forests and burning of sugarcane stalks as a method of straw removal. Even without mandates, biofuel producers have seen the virtue of minimizing waste. Biorefineries burn the fibrous sugarcane bagasse to generate electricity for the ethanol plant and for sale.

LESSON 6: WELCOME PRIVATE INVESTMENT

Brazil has greatly expanded its sugar and ethanol production in recent years, relying on both domestic and foreign private investment. External capital has enabled the industry to introduce new technologies and management standards. Investment from other countries in particular has helped give Brazilian biofuel producers access to foreign markets that may be shielded by protectionist policies.

No two countries will face exactly the same circumstances. But the strategies that have guided this nation of almost 200 million people should show the rest of the world that the future of energy need not depend on oil. ■

Marcia Moraes *is a professor of economics at the University of São Paulo, Brazil.*
e-mail: mafdmora@esalq.usp.br

1. Leopold, A. & Aguilar, S. *Case Studies on Bioenergy Policy and Law: Options for Sustainability* (eds Morgera, E., Kulovesi, K. & Gobena, A.) FAO Legislative Study, **102** (FAO, 2009).

A high-tech look at the fine structures of pre-treated biomass at the US National Renewable Energy Laboratory in Colorado.

DENNIS SCHROEDER/NREL

INTRODUCTION

Next generation biofuels

Proponents of biomass-based fuels push for sustainability against a steady tide of conflicting analysis, but can advanced biofuels cut the mustard?

BY PETER FAIRLEY

Sustainable, supersonic fighter jets. President Barack Obama conjured up that improbable image in the spring of 2011 to help students at Georgetown University, Washington DC, grasp why biofuels deserve a starring role in his energy strategy. Just the week before, a US Air Force pilot had throttled a fighter jet up to one and a half times the speed of sound on a 50/50 blend of petroleum-based fuel and biofuel from *Camelina sativa*, an inedible cousin of the mustard plant — powerful evidence, claimed Obama, that biofuels could help the country kick its oil habit. "If an F-22 Raptor can exceed the speed of sound using biomass," Obama half-joked, "your old beater can do the same."

Such unqualified advocacy is bold following criticisms heaped on biofuels in recent years. The rapid scale-up of biofuels, fermented or refined from foodstuffs such as corn (maize), sugarcane and soybeans, has contributed to higher food prices and deforestation. And as production of biofuels relies mainly on fossil fuels, these early deployments have done little to cut greenhouse-gas emissions. In April 2011, summing up the findings of a 14-month inquiry into the ethics of biofuels, Nuffield Council on Bioethics, a think-tank in London, said policies and targets to encourage biofuels had "backfired badly".

Yet, as Obama's speech shows, the unpopularity of biofuels is giving way to a recognition that for all their present flaws, biofuels need to be part of our energy future. Other renewable sources of energy, such as the Sun and wind, will contribute to electricity capacity, but the reality is that the vast majority of motor vehicles require liquid fuel and will do so for the foreseeable future. And any attempts to sustainably power the world's millions of motor vehicles must include biofuels.

In early 2011, organizations as diverse as the London-based financial giant HSBC, and environmental advocacy group Oceana, based in Washington DC, have endorsed biofuels use as a critical means of squaring transport, economic growth and environmental protection. If appropriately implemented, biofuels could maximize use of the waste materials of farming, forestry and city life. That approach could help save land that provides carbon sequestration and other critical environmental services.

In May, the International Energy Agency (IEA), based in Paris, laid out a 'roadmap' to ramp-up the use of biofuels from around 2% of global transport fuel today to 27% by the year 2050. Biofuels, according to the IEA, could displace enough petroleum to avoid the equivalent of 2.1 gigatonnes of carbon dioxide emission each year if produced sustainably — about as much as net carbon dioxide absorbed by the oceans.

Such estimates have convinced the IEA, Obama and others that biofuels are not only preferable to petrol (gasoline) but that, with technological advances, there is also hope for biofuels' environmental and social redemption. Biofuel advocates are counting on a new generation of advanced biofuels to maximize petroleum (crude oil) displacement and minimize side effects. "Continuing the pathway we

➲ NATURE.COM
to read the latest news and research on biofuels
go.nature.com/xxxxxx

started in the last ten years with conventional biofuels really isn't possible," says biofuels researcher Anselm Eisentraut, lead author of the IEA's biofuels roadmap. "We need more land- and resource-efficient technologies."

The biofuels that inspire enthusiasm among advocates are to be bred and brewed from a variety of non-food materials — be they ethanol from corn stalks or other 'cellulosic plants' or even from municipal garbage, or jet fuel from dedicated energy crops such as the fast-growing *Camelina sativa*. The order of the day is to demonstrate that the required agronomy and technologies deliver at industrial scale, and that the net reduction in petroleum use and greenhouse-gas emissions are real, lasting and ethical.

FUELING A TURNABOUT

Biofuels derived from crops other than corn or sugarcane appear to be on the cusp of commercialization — again. In 2007, advanced biofuels developers were gearing up to build their first processing facilities. They expected explosive growth based in part on a surge in governmental support. For example, the US Congress had created mandates to spur cellulosic ethanol production up to about 950 million litres by 2011. But the US Environmental Protection Agency — recognizing the dearth of supply available — mandated blending of just 30 million litres of cellulosic ethanol and 57.7 million litres of corn ethanol into petrol this year. The United States "had an aggressive plan and we've done essentially nothing," says Jeremy Martin, a policy analyst at the Union of Concerned Scientists, a Massachusetts-based advocacy group pushing advanced biofuels.

A series of blows knocked the backers of advanced biofuels to the floor. Researchers studying the impact of biofuels over their entire life cycle — from crop to car — predicted surprisingly large greenhouse-gas emissions. A report by researchers at Princeton University published in *Science* in 2008 estimated that greenhouse gases caused by increased cultivation of corn for ethanol would exceed the greenhouse-gas emissions of petrol. Despite the study's explicit focus on first-generation biofuels, the unexpected magnitude of the estimates cast a pall over the entire biofuels enterprise.

Biofuels of all flavours took another hit when spiking food prices sparked food riots in Mexico in December 2007, which subsequently spread to impoverished communities throughout the developing world. Annual output of bioethanol and biodiesel had exploded from 16 million litres worldwide in 2000 to more than 100 million litres in 2010, outpacing growth in supply of corn, sugarcane and vegetable oil. Although many economists agree that biofuels contributed less to the price hikes than commodities speculators, oil prices and weather, protestors and food rioters claimed that the conversion of staple foods into biofuels distorted global agricultural markets.

As the future for biofuels dimmed, electrical vehicle technology pushed forward to fill the void. Major automakers began mass-marketing their first electric vehicles, including Michigan-based General Motors' Chevy Volt and Japanese carmaker Nissan's Leaf. Worldwide, government, industrial and private investments in developing electric vehicles and battery technology — including RMB 100 billion (US$15 billion) budgeted by Beijing — eclipsed investments in biofuels.

Advanced biofuels hold promise of an escape from their predecessors' food-versus-fuel conundrum (see 'Beyond food versus fuels, page S6). The problem is their presently higher production cost compared with first-generation biofuels and petroleum fuels. Prices for bioethanol account for its relatively low energy density; bioethanol packs only 70% the energy of petrol. And according to the IEA, the cost of producing and retailing diesel and ethanol from cellulose is about US$1.10 per litre of petrol-equivalent, or about US$ 4 a gallon. By way of comparison, corn and cane ethanol are 62-75 US cents per litre-equivalent, whereas petrol itself is about 54 US cents. Consequently, at the moment, advanced biofuels are far from competitive — but that should change. The IEA estimates that by 2050, cellulose-based biofuels will be produced for as little as 75 US cents per litre-equivalent. But bridging the gap will require continued investment, research and development.

Ground is shifting back towards biofuels, partly owing to a series of natural disasters and societal shifts.

But the ground is shifting back towards biofuels, partly owing to a series of natural disasters and societal shifts that have heightened the urgency of avoiding petroleum. Obama set his energy strategy in the context of growing instability in the oil-rich Arab world and shrinking petroleum reserves. HSBC's analysis asserts that the supply squeeze created by steeply rising demand from developing countries should be of concern to "anyone who drives a car, heats a home, or runs a factory." The BP Deepwater Horizon oil-spill disaster in 2010 prompted Oceana's call for the elimination of oil drilling in the Gulf of Mexico by 2020.

Then in 2011 came Japan's nuclear crisis, which shone a harsh light on an important source of low-carbon electricity. Major economies including China, Germany and the United States had been counting on nuclear power to help charge millions of electric vehicles. Without more nuclear and massive new investments in renewable energy, electric vehicles will fuel-up on coal-fired power, doing the environment little good.

Can entrepreneurs revive biofuels? With government financial support from the United States and Europe, several biofuels developers are likely to begin building new industrial-scale plants over the next year or so. The IEA's roadmap says that biofuels plants capable of displacing nearly 175 million litres of petrol annually have already been built, and it estimates that capacity to replace another 1.9 billion litres per year is in construction or advanced planning. The Obama administration promises federal funds to help at least four advanced biofuels plants break ground by 2013.

Many of the early movers in advanced biofuels have demonstrated pilot-scale production of cellulosic ethanol through some combination of physical, enzymatic and fermentation steps (see 'A chewy problem', page S12). Take the 50 million litre per year ethanol factory that Italian chemicals firm Gruppo Mossi & Ghisolfi began building north of Turin in April 2011, which is designed to displace about 34 million litres of petrol. The plant will mash-up straw feedstocks, then use enzymes supplied by Denmark's Novozymes to break the long-chain carbohydrates into smaller sugars for fermentation.

Mascoma, based in Lebanon, New Hampshire, plans to begin construction on a 150-million litre per year cellulosic ethanol plant in Michigan in late 2011. The company genetically engineered thermophylic microbes to secrete cellulase enzymes. "The organism produces its own enzymes, so you cut out one of the most costly components," explains Jonathan Mielenz, who heads the Bioconversion Science & Technology Group at Oak Ridge National Laboratory in Tennessee and has collaborated with Mascoma.

Another innovator is Illinois-based Coskata, which has a US$250 million federal loan guarantee to construct a cellulosic ethanol facility in rural Alabama. The plant will produce 208 million litres of ethanol per year by gasifying woody biomass. In Coskata's novel process, the carbon monoxide and hydrogen gases released by the biomass are fermented to ethanol by anaerobic bacteria. One advantage of gasifying biomass is that the process accommodates a wider variety of feedstocks than enzymatic approaches. "You can use any carbonaceous material — even rubber tires," says Mielenz. "It's much more forgiving."

These processes all consume a range of biomass feedstocks and, like oil refineries that split and upgrade petroleum into an assortment of fuels and chemicals, churn out their target fuels plus other commodities. Co-products from these biomass 'biorefineries' include solid residues that can be burned to generate steam and power or sold as animal feed; carbon dioxide that can be purified for carbonation; and glycerine for pharmaceuticals and cosmetics. Co-products boost revenues and, in some cases, provide a free and renewable source of energy to power the biorefinery. Such integration may also cut carbon emissions — a factor that first-generation biofuels producers are

also now trying to exploit.

And there are novel processes and fuels on the way. Exxon, for example, has dedicated US$600 million to finance an R&D collaboration with Synthetic Genomics, based in La Jolla, California, founded by human genome entrepreneur Craig Venter, to optimize photosynthetic algae to produce refinery-ready oils (see 'A scum solution', page S15). This spring, UCLA biomolecular engineer James Liao reported three novel bioprocessing ways to make butanol — a promising alcohol-based fuel that packs about 25% more energy per litre than ethanol. In one elegant demonstration of metabolic engineering, Liao made a biofuel by fermenting proteins instead of sugars. The source of the protein is the same algae from which Exxon and others are trying to harvest oils. Liao's approach to using algae makes sense because algae are roughly two-thirds protein, and at best one-quarter oil.

Liao's discovery of unimagined routes to butanol shows that this is a "pretty exciting time" for advanced biofuels R&D, says Mielenz. "The combination of thermochemical breakthroughs and the ability to genetically modify organisms means there are many more pathways and routes that we haven't yet discovered."

DOES ADVANCED MEAN GREEN?

Will biofuels advances translate into more sustainable transportation? They could, but only if producers and policymakers sweat the details and investigate the impact of biofuels on overall energy and environmental sustainability.

Studies suggest that it should be feasible to responsibly harvest sufficient biomass to meet ambitious biofuels targets. In May 2011, the Intergovernmental Panel on Climate Change (IPCC) issued a report on renewable energy projects concluding that biomass can by mid-century sustainably provide up to 300 exajoules — more than four times the biomass needed to meet the IEA target of 27% of the worlds' transportation fuel coming from biofuels. (One exajoule is one billion gigajoules, or one billion gigawatt-seconds.)

The question is how to ensure that the biomass is harvested as the IEA and IPCC recommend — maximizing the use of farm, forestry and municipal wastes, and encouraging cultivation of dedicated energy crops away from lands that provide carbon sequestration and other critical environmental services. One way to channel the development of biofuels along the greenest path is to develop a set of standards and practices that biofuels producers comply with, either voluntarily or by mandate. For example, in March 2011 the Roundtable on Sustainable Biofuels, run by the Energy Center at EPFL in Lausanne, Switzerland, released details of a certification system for biofuels to qualify between sustainable biofuels and those that are environmentally destructive.

That proposal has plenty of company. The IEA roadmap cites no fewer than 67 initiatives underway worldwide to develop just such criteria for biofuels sustainability — a proliferation that could muddy the waters further. When it comes to determining sustainability, international methods are needed, according to the IEA's Eisentraut, to head-off confusion that might discourage producers from trying to operate sustainably. "There's a risk of creating market distortions," he warns.

The problem is a lack of consensus on how to measure environmental impacts of biofuels. The issue that most bedevils the biofuels debate is that of indirect land use changes — the main problem cited in the 2008 *Science* paper. In the classic case, cultivation of energy crops on existing farmland displaces other crops to freshly-cleared forest lands, thereby causing a release of carbon from both the forest soils and trees. Factoring such indirect effects into life-cycle assessment of biofuels' environmental impact is a controversial exercise; a variety of methodologies and data input results in a wide range of results. The verdict on the greenhouse-gas emissions from producing a volume of a given biofuel from a given source of biomass from a given land can swing from sustainable to lamentable as various knock-on effects and feedback loops are layered into the equations.

How to handle biorefinery co-products such as excess power generation is another vexing area for life-cycle modelers. Geoffrey Hammond, director of the Institute for Sustainable Energy and the Environment at the University of Bath, United Kingdom, who heads-up a Biotechnology and Biological Sciences Research Council-funded programme on cellulosic ethanol, has studied methodological impacts on life-cycle studies for ethanol derived from wheat and barley straw. "Depending on which accounting methodologies you use the total amount of greenhouse-gas emissions can vary by at least a third," says Hammond.

Hammond's group examined three protocols used in the United Kingdom for calculating emissions reductions and for attributing those to biofuels and biorefinery co-products. In all cases the environmental life-cycle assessment model found the straw-fed biorefineries and their ethanol to be net carbon-cutters. But the magnitude of the benefit is hard to pin down — the carbon footprint reduction ranges from a slim 9% to 38% depending on protocols.

Recent research has delivered some good news on the environmental impact of bio-fuels. Satellite analysis from Stanford University and the Carnegie Institution, Washington DC, reported in April 2011 in *Nature Climate Change,* identified a local cooling effect from planting sugarcane as an ethanol feedstock. Analysis showed that the sugarcane planted in central Brazil's Cerrado region reflects light better (higher albedo) and transpires more than the crops and pasture that the cane displaced. The net result of the conversion: a local cooling of 0.93°C. Such regional cooling could trim land-use impacts significantly by driving up agricultural productivity in arid regions such as the US southwest, says David Lobell, who studies food security at Stanford University and one of the report's authors. Extrapolating from modeling of high-albedo bioenergy crops such as switchgrass (*Panicum virgatum*) and *Miscanthus*, Lobell says that extensive plantings in the southwestern United States. could boost farm yields and thereby drive down prices for agricultural commodities. These lower prices would in turn dampen the incentive to convert forests and wilderness into farmland. "Local cooling could suppress indirect land-use change," says Lobell. "It's the kind of thing that we hadn't even thought about."

Progress in biofuels requires not just technical advances, but a level of predictability in the economic and policy terrain. Uncertainty can paralyze. The European Commission (EC) has yet to incorporate indirect land-use effects into its 2009 *Renewable Energy Directive*, which tasks European states with boosting

JOHN JAMES/ALAMY

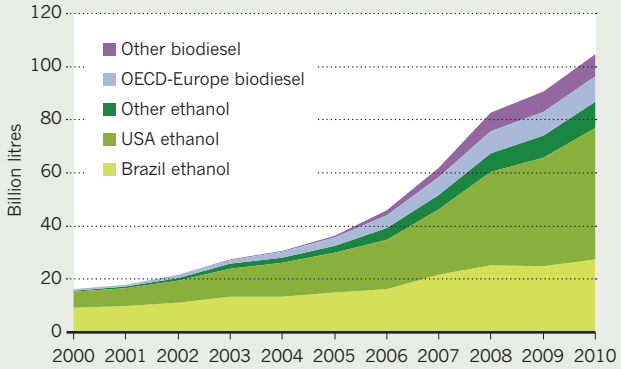
A field of *Miscanthus,* one of several non-edible crops being used in next-generation biofuels.

THE RISE OF BIOFUELS

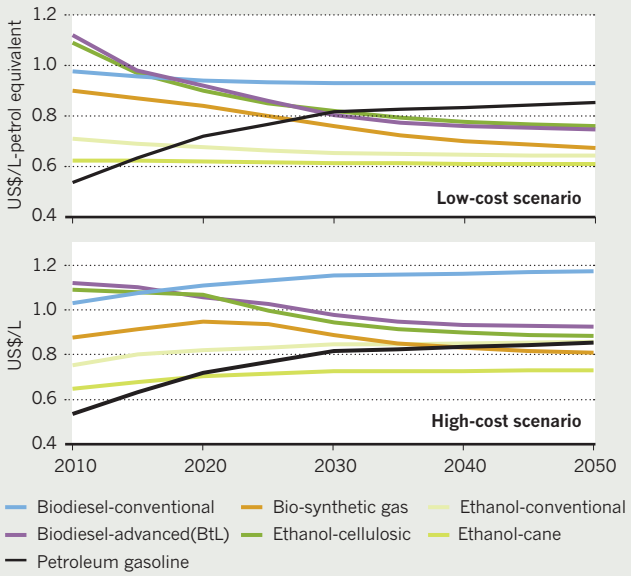
Biofuel production now tops 100 billion litres per year. Different fuel types vary in their costs, carbon emissions and impact on land use.

SOURCE: INTERNATIONAL ENERGY AGENCY'S TECHNOLOGY ROADMAP: BIOFUELS FOR TRANSPORT, 2011.

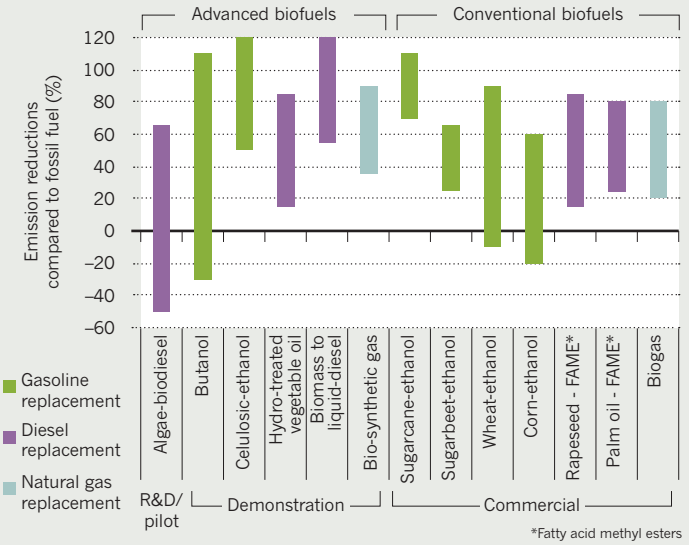
GLOBAL BIOFUEL PRODUCTION



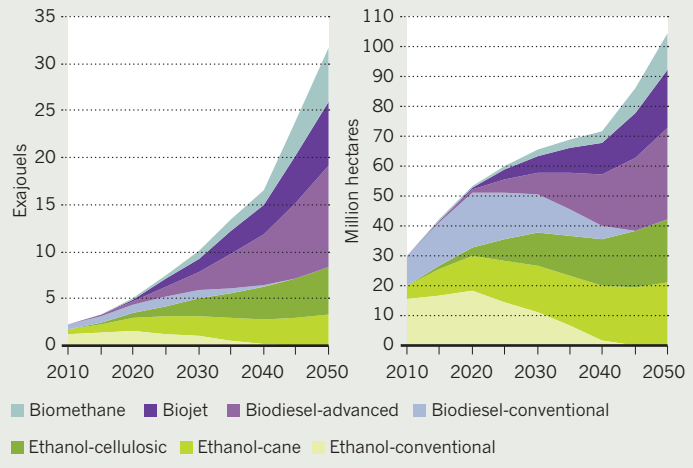
COSTS OF VARIOUS BIOFUELS COMPARED TO PETROL (GASOLINE)



LIFE-CYCLE GREENHOUSE-GAS EMISSIONS OF CONVENTIONAL AND ADVANCED BIOFUELS



DEMAND FOR BIOFUELS (LEFT) AND RESULTING LAND DEMAND (RIGHT)



renewables to 10% of transport energy by 2020 (see 'Fuelling politics', page S22). The directive recognizes only those biofuels whose carbon footprint is at least 35% lower than petrol (a threshold that rises to 60% by 2018). But guidance promised in December 2010 on how to handle indirect land use in the carbon footprint calculations was deferred to July, and possibly longer: "taking no action for the time being, while continuing to monitor impacts," is one option the EC is weighing up.

Regulatory paralysis freezes further incorporation of advanced biofuels into the energy supply. That worries both biofuels advocates and policy analysts, who say the status quo is untenable. "If we continue to just use fossil carbon the planet is doomed and indirect land-use changes could become fairly unimportant," says Jonathan Mielenz at Oak Ridge. "It's easy just to say biofuels are bad but the alternative is to just continue to use petroleum."

Aviation has an especially big stake in the biofuels. There is little prospect for battery-powered jetliners, so if air travel is going to become environmentally sustainable as the sector accelerates from an estimated 2.5 billion passengers worldwide in 2011 to a projected 16 billion passengers in 2050, the industry will need liquid fuels from renewable sources. The blend of refined camelina oil and petroleum jet fuel that powered Obama's F-22 Raptor may soon complete the multiyear certification process for new fuels for air travel. But commercial-scale biorefineries still remain but a figment of financiers' imaginations.

This stasis exasperates the Union of Concerned Scientists' Martin. The best way to determine how big a role advanced biofuels can play in the energy picture is to build twenty or so commercial-scale biorefineries to test the efficiency and impacts of growing energy crops, gathering and storing tonnes of biomass at a facility, and novel conversion processes, such as Mascoma's and Coskata's, to produce current and next-generation fuels. "Let's get to the first billion gallons. That will clarify a lot," says Martin.

One fact is beyond doubt: the roughly US$4 billion total pricetag in government incentives that Martin estimates is needed is a pittance compared to the subsidies now lavished on food-based fuels, which garner more than US$5 billion each year from the US government. If there is any consensus emerging around biofuels, it is that this set of funding priorities will need to change. Only then can vehicles powered by contemporary biomass, rather than plants long since dead and buried, have a chance to deliver the world to a cleaner and safer future. ■

Peter Fairley *is a science journalist in Victoria, British Columbia.*

JONATHAN BURTON

AGRICULTURE

Beyond food versus fuel

The most controversial aspect of biofuels is the perceived competition for farmland. Will advances in biofuels and agriculture send this trade-off speeding towards the history books?

BY DUNCAN GRAHAM-ROWE

When Ferrari's Formula 1 drivers nabbed first and second place in last year's Bahrain Grand Prix they had a little something extra in their fuel tanks — lignocellulosic, or 'second-generation' biofuel. Even the high-octane world of F1 racing has woken up to the reality of dwindling oil reserves and the threat of climate change. Since 2008, the rules have stated that at least 5.75% of all Formula 1 fuel must be derived from plants. But the fuel in Ferrari's tank is different. Made from the non-edible parts of plants, Ferrari's fuel additive demonstrates just one of the ways in which biofuels can meet our energy needs without reducing food production.

Recent rises in food prices are some of the sharpest since records began, according to the Food and Agriculture Organization (FAO) of the United Nations. These increases have been attributed to the growing use of 'first-generation' biofuels, derived from the edible parts of food crops such as sugar cane and corn (maize), for example, which are often blended with today's gasoline (petrol). With food riots becoming increasingly common in developing countries, many are concerned about how prudent — or ethical — it is for Western governments to be establishing ambitious targets for the uptake of biofuels. Is there a way forwards that does not rob the world's population of the food it needs?

In Ferrari's case, the answer is yes. The team's choice of biofuel is ethanol derived from straw, a waste product of agriculture. Developed by the Ottawa-based biotechnology company Iogen, the race car fuel is one example of how scientific advances are helping researchers avoid some of the problems of first-generation biofuels. If the new generation of fuels could be used in combination with improved land-use strategies and increased agricultural yields, they might vault us beyond the point where we are forced to choose between food and fuel.

ENERGY ISSUES

Using crops to make liquid fuel is not a recent idea. As far back as the nineteenth century, Rudolf Diesel designed his eponymous engine to run on vegetable oil or peanut oil. And Brazil has mandated the addition of sugarcane-derived ethanol to its fuel since 1929. However, it is only in the past decade, with mounting fears about energy security and climate change and with the relentless escalation in energy consumption, that biofuels have begun to take a more central role in global energy policy.

The haste to develop such fuels has led to tensions over land use. In Malaysia, for example, policies encouraging the use and production of palm-oil-based biodiesel have resulted in large swathes of Borneo's jungle being replaced by palm plantations. And, where edible crops such as corn are being used to make fuel, significant chunks of the global harvest are now diverted from food production into biorefineries. In 2007, the United States planted a record 92.9 million acres (about 375,000 square kilometres) of corn, but one-third of it was used to produce ethanol. This diversion, it has been argued, has led to unprecedented rises in

➲ NATURE.COM
Can science feed the world?
go.nature.com/vt3Rqo

the price of corn, which spiked 73% towards the end of 2010. Early this year, the FAO reported that its index of food prices was the highest in the index's 20-year history.

Pick apart the data, however, and it becomes clear that biofuels are only part of the equation. According to the 2008 World Bank report *A Note on rising food prices*, although biofuels are the most important factor, there is good evidence to suggest that other factors, such as stock market speculation and extreme climatic events, have helped to drive up the price of food since 2002. The rising cost of petroleum should also not be overlooked, because this has a direct impact on food prices, says Ottoline Leyser, a plant geneticist at the University of Cambridge, UK. Leyser, who was also a co-author of a recent report on biofuels by the London-based Nuffield Council on Bioethics, points out that it takes a lot of energy — mainly in the form of fossil fuels — to harvest, process and transport food.

To give biofuel critics their due, food price is not their only concern. According to Angela Karp, scientific director of the Centre for Bioenergy and Climate Change at the Harpenden-based Rothampsted Research, the largest agricultural research centre in the UK, first-generation biofuels also tend to require intensive inputs, such as water and nitrogen, and this has an impact on their greenhouse gas emissions. Although most biofuels weren't developed specifically to address rising carbon dioxide levels, it is nonetheless expected they will help alleviate the problem. However, biofuels often show only minimal reductions in greenhouse gases compared with their fossil fuel equivalents, according to a life-cycle analysis by Nigel Mortimer, former chair of sustainable energy development at Sheffield Hallam University, UK, who now works at Stocksfield-based North Energy, a renewable energy and sustainability consultancy. And research by Kenneth Stone, an agricultural engineer at the US Agricultural Research Service's branch in Florence, South Carolina, suggests that if the US Department of Energy's biofuel target for 2030 is met using corn-derived ethanol only, agricultural water use could increase six-fold.

One of the trickiest problems arising from the competition between biofuels and food production is land use. Ideally, biofuel crops would be grown on marginal land that has a low carbon stock, leaving the higher quality soil for food production, says Ian Crute, chief scientist at the Kenilworth-based Agriculture and Horticulture Development Board. But, in reality, the profitability of energy crops has caused farmers to shift production from food crops to fuel crops, creating a ripple effect, says Olivier Dubois, coordinator of the FAO's Bioenergy Group in Rome. "Food crops still have to be produced somewhere," he says. And that means finding additional land and converting it for agricultural use.

"Large-scale, second-generation biofuels will take five to ten years."

Land-use issues are not just restricted to crops: livestock has a big impact. According to research published by Bruce Dale, a chemical engineer at Michigan State University's Biomass Conversion Research Laboratory in East Lansing, more than 80% of US agricultural production goes into animal feed. Global increase in meat consumption is one of the main driving forces behind high food prices. As Leyser says, "It's spectacularly wasteful."

And with the inexorable growth of the global population, additional land is set to become increasingly scarce. By 2050, the world's population is expected to exceed 9 billion people. According to the FAO's 2009 report *How to feed the world in 2050*, this means we'll need to feed 2.3 billion more people. Moreover, because of the steady increase in meat consumption and average calorie intake, the amount of food needed to meet this demand will be disproportionately larger than the 34% increase in population (see 'Growing challenges'). "We need 70% additional food by 2050," says Dubois.

Even before this point, the US Energy Information Administration predicts that there will be additional demand for land coming from the energy sector. By 2035, the energy consumption in developing countries is set to rise by 84%, and nearly one-third of this additional fuel is expected to come from biofuels.

Land for both food and fuel production will become scarce unless farming practices change. The global amount of arable land could, in theory, double over the next four decades. But the FAO predicts that the net increase is more likely to be just 5%, because developed countries more commonly convert arable land to accommodate urban expansion. So to raise agricultural productivity, crop yields will have to increase, says Dubois.

Trends over the past half century are not encouraging, however: although average global productivity is increasing, the annual rate of growth has declined from 3.2% in 1960 to 1.5% in 2000. Crute sees a silver lining: productivity is still increasing — and, he says, there is plenty of room for improvement. One way is to close the yield gap between countries. For example, large parts of Africa are underutilized, mainly because they lack infrastructure. If farmers can't get their product to market, then they will grow only what they need to feed their family, says Crute. Growing biofuels in these regions has a big chance of making a positive impact on both infra structure and agricultural output (see 'A new hope for Afica', page S20).

Genetic modification might also help improve agricultural productivity, and not just by generating crops that are more pest resistant —a modification in use today. "On the horizon is drought tolerant maize," which can survive harsher conditions and requires less water, says Crute. Salt tolerance can be incorporated too. With such crops, more land would be viable, and plants could flourish in regions where rising sea levels have increased the salinity of the water table. Furthermore, certain crops, such as corn, have much more efficient photosynthesis mechanisms than wheat or rice, says Crute. These highly productive mechanisms could be 'borrowed' from one plant species and incorporated into others.

GETTY IMAGES

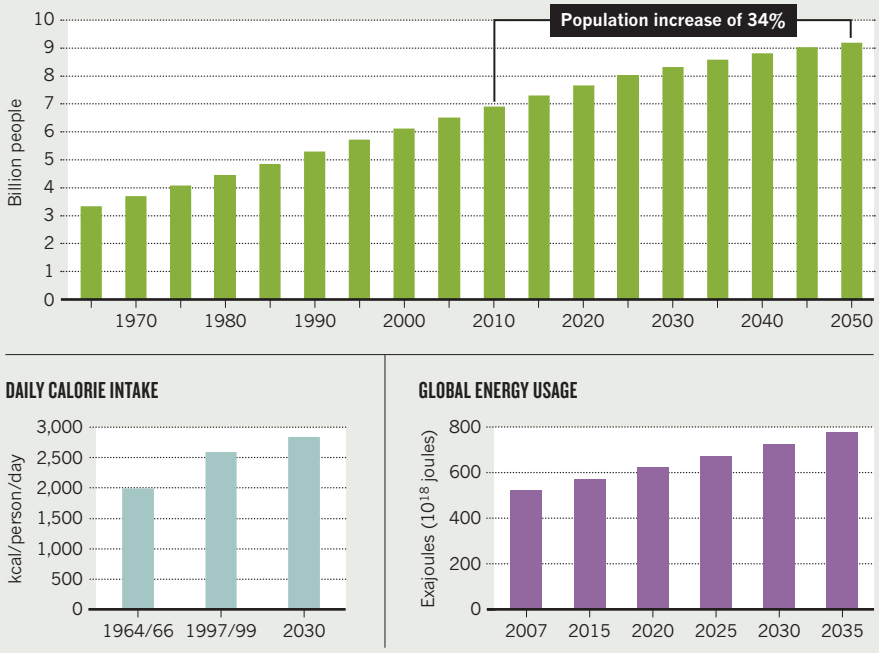
Willow trees planted to supply a biofuel power station in Lockerbie, Scotland, United Kingdom.

THE NEXT GENERATION

On the other side of the equation from improved farming practices, advanced biofuels could overcome many of the issues that have brought bioethanol and biodiesel into

GROWING CHALLENGES

Between now and 2050 the world's population is predicted to increase to more than 9 billion people. Each person will also be consuming more calories per day and using more energy to power their lives. Food and fuel supplies will need to massively increase to meet these needs.



SOURCE: POPULATION DIVISION OF THE DEPARTMENT OF ECONOMIC AND SOCIAL AFFAIRS OF THE UNITED NATIONS SECRETARIAT, 2007.

contention with food production. By using the entire plant and not just the edible parts, vastly different types of biomass become available as starting materials, or feedstock. There is a lot more energy stored in the rest of the plant than the grain, but it's harder to get at, says geneticist Chris Somerville, who directs the Energy Biosciences Institute at the University of California, Berkeley. Consequently, this process is currently much more expensive than the production of corn ethanol (see 'A chewy problem', page S12).

To create Ferrari's straw-based fuel, Iogen uses an enzyme from the 'jungle rot' fungus *Trichoderma reesei*. The fungus deploys this enzyme to extract nutrients from trees, by digesting lignin, one of the main components of lignocellulose, the woody part of plants. But the search is on to find more efficient enzymes. Enzymes for corn-ethanol processing are highly efficient and cost only about 2 US cents per US gallon, says Somerville. Enzymes that break down lignocellulose cost about 13–25 US cents per litre and that's where the challenge lies. At the Energy Biosciences Institute, "we would like to reduce that cost by about half," he says.

It might be possible to genetically engineer plants with lignin that is easier to break down, but this is a long way off, says Somerville. A more promising idea is to use selective breeding or genetic analysis to identify plants that are high in energy. Charles Wyman, a chemical and environmental engineer at the University of California, Riverside, studied 47 variants of the poplar tree to determine which were most willing to give up sugar from their lignin — and what made that possible. In March 2011, his team reported that the sugar yield could range from as low as 28% up to 92% of the theoretical maximum, depending not only on the lignin content but also on its structure.

So, if these technologies can be perfected and the costs of processing the waste parts of plants (or residues) can be reduced, this should sidestep the food versus fuel issue. For example, grain from corn could be harvested for food, while the rest of the plant could be rendered into fuel.

Although elegant in principle, this dual-use approach is dauntingly complicated to put into practice, says Dubois. "It's seen as a win–win solution, but competing use of residues is a big issue. They're needed as a cheap form of organic fertilizer."

In 2009, research by Guy Lafond, an agronomist with Agriculture and Agri-Food Canada in Indian Head, showed that if more than 40% of straw was removed from land and not ploughed back, the quality of the soil declined. Iogen's demonstration facility produced an average of 256,000 litres of ethanol per year since it opened in 2004, despite having an annual capacity of more than 1.9 million litres. Iogen attributes this shortfall to the fact it is a test facility and so does not operate continuously. But if Shell, which part owns the company, were to start producing ethanol from straw on a commercial scale, this would require 20–30 tonnes of straw per day as a feedstock, and that might put strains on supply.

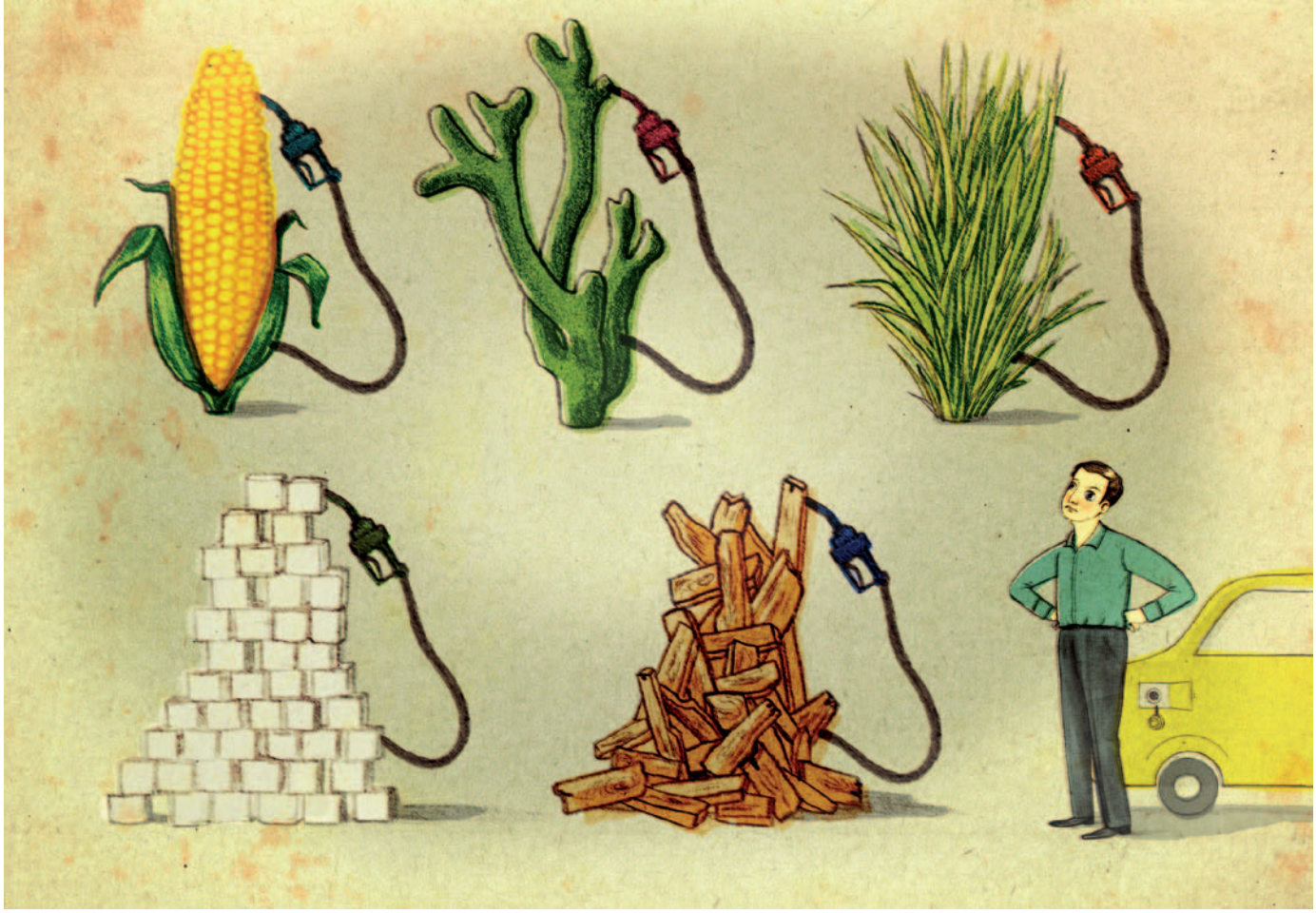
"The main emphasis will come not from dual use but dedicated crops," Leyser predicts. By 'dedicated' she means using fast-growing plant species that are harvested entirely for biofuel production. To avoid the energy crops vying with food crops for land use, she says, will require careful selection of species. Fast-growing species of willow or poplar trees, for example, can be grown on polluted soil, which also helps to reduce soil contamination. These trees convert carbon dioxide into biomass more rapidly than most other plants — a trait that should translate into high biomass yields. Perennial grasses such as *Miscanthus* and switchgrass (*Panicum virgatum*) are perhaps even better candidates. These plants use little water and store nutrients in their roots, which remain in the soil for the next year's crop. "What you harvest is very nutrient poor, which means almost no fertilizer is needed," says Leyser. Perennial grasses can also be grown on land that is unsuitable for arable crops. Indeed, says Somerville, waste lands could be transformed into bioethanol energy fields using drought-tolerant plants such as agave, one species of which is used to make tequila.

Dubois says that second-generation biofuels "have their drawbacks". They still require land, and even plants that need minimal input will compete with food crops for some resources. And first-generation bioethanol has a head start of more than half a century. "Large-scale, second-generation biofuel production will take five to ten years," he says.

That process has begun. The oil company BP is building a US$400 million cellulosic ethanol plant in Highlands County. The facility will use a high-energy sugar cane, known as 'energy cane', as well as switchgrass or *Miscanthus* from the area, to generate about 136 million litres of fuel a year, which by the US Department of Energy's definition would make it a commercial-scale biorefinery. This may sound like a lot, but it is trivial compared with the 1.3 million litres of gasoline per day that Somerville says his local oil refinery produces.

Although they may take some time, advanced generations of biofuels are on their way. Regardless of whether it is a second-generation biofuel that ends up in our fuel tanks — and the tanks of Formula 1 racing cars — or a third- or fourth-generation that comes in its wake, biofuels will have an important role in meeting the world's energy needs. And through technology, careful land management and considered use of resources, they should allow us to have our fuel and eat it. ■

Duncan Graham-Rowe *is a science writer based in Brighton, UK.*



JONATHAN BURTON

FUEL OPTIONS

The ideal biofuel

A biomass-based fuel needs to be cheap and energy dense. Gasoline sets a high standard.

BY NEIL SAVAGE

A single word can sum up the biofuel of the future as envisioned by the people trying to create it — petrol.

"Our goal really is to come up with methods to make all the same molecules found in gasoline, jet fuel and diesel," says George Huber, a chemical engineer at the University of Massachusetts-Amherst, who is working on ways to turn plant organic matter, or biomass, into transport fuels. Like petrol (gasoline), an ideal biofuel should drop into today's infrastructure and carry enough juice to get any vehicle where it's going.

Ethanol, long the focus of the biofuel industry, doesn't meet those requirements. Compared with petroleum-based fuels, it's much less dense in energy: a litre of ethanol takes a car only about 70% as far as a litre of petrol, and ethanol cannot provide enough power for heavy trucks or aircraft. What's more, ethanol mixes with water from the environment, resulting in a more dilute fuel. It's also corrosive and so cannot easily be used in today's engines or be shipped cheaply through existing pipelines.

To overcome those limitations, researchers are trying to turn biomass into more complex alcohols than ethanol, as well as into hydrocarbons that more closely resemble those in petroleum, which is a mixture of different lengths of hydrocarbons. These scientists are developing processes — both biological and chemical— to produce substances that can be either placed directly in the fuel tank or slotted into the processing chain in existing refineries. And they want to use as much of the available biomass as possible, not just the simple carbohydrates that can be derived from sugarcane or kernels of corn, but also the harder to break down cellulose and lignin of corn stalks, wood and switchgrass (*Panicum virgatum*).

BIGGER ALCOHOLS

A 'higher alcohol' — a molecule with more atoms of carbon and hydrogen than ethanol — comes closer to the ideal. Ethanol has two carbon atoms linked to five hydrogen atoms, plus a hydroxyl group. The carbon–hydrogen bonds are where the useful energy is stored; breaking these bonds through combustion releases energy. Adding two more carbon atoms and four hydrogen atoms creates butanol, which is less corrosive and packs much more punch. With its four carbons, butanol has about 90% of the energy of petrol, which is a mixture of molecules with five to eight carbon atoms. Diesel and jet fuel consist of molecules with 9 to 16 carbon atoms.

Today, ethanol is often added to petrol to prevent pre-ignition, or "knocking." A move to higher alcohols could make alcohol-rich fuel blends more feasible. Today's petrol blends generally don't exceed 10% ethanol, but it's easy to imagine a fuel blend containing 50–70% of the higher energy, less corrosive alcohol butanol. Some proponents even say that unmodified car engines could work on butanol alone. "You can reach much higher levels of renewables in fuels if you go to these longer-chain molecules," says Michelle Chang, a chemist at the University of California, Berkeley who has induced microbes to produce butanol. Ultimately, butanol might also be cheaper to make than ethanol. When a batch of yeast produces ethanol from the sugar derived from, say, corn, the alcohol comes out mixed

➲NATURE.COM
research into manipulating microbes to make alcohol
go.nature.com/f8yNIX

with water. The water has to be boiled off, which means using more energy to produce the fuel. But butanol and water don't mix, so they can be separated by less energy-intensive processes.

The problem lies in producing enough alcohol in the first place. Generating large volumes of ethanol is relatively easy. Yeast strains were domesticated for brewing alcoholic drinks thousands of years ago and are natural fermenters, turning plant-derived glucose into ethanol. Genetically manipulating them to make more ethanol is straightforward. Although there are organisms that naturally produce butanol and higher alcohols, most of them make those alcohols in tiny quantities, Chang says. "The question to ask is, how do you get the same yields that you get with ethanol. Nobody has matched those yields yet."

Chang recently found a way to boost butanol production by tenfold, at least in the laboratory. She plucked a combination of genes from different organisms and expressed them in *Escherichia coli*. Some of the genes are from a strain of the bacterium *Clostridium* that naturally produces butanol. The challenge, Chang says, is that *Clostridium* has its own agenda, geared more towards its survival than towards making large amounts of the alcohol. When the cell determines that there's too much butanol, the same enzymes that created the butanol start to break it down. Similar attempts by other researchers have therefore suffered from low productivity, yielding about half a gram of butanol per litre of glucose solution. So, instead of importing the entire butanol-making pathway of *Clostridium*, Chang mixed in genes of two other bacteria, *Treponema denticola* and *Ralstonia eutropha*. Those genes encode slightly different versions of the enzymes that control the fermentation — enzymes that are less likely to break down butanol.

This sort of manipulation is more difficult with longer carbon chains. The bigger the molecule, the more steps required to make it. And each step gives the cell the opportunity to divert production in a different, more natural direction. "To get good production, cells need to be healthy," says Shota Atsumi, a chemist at the University of California, Davis. "If we remove some pathways, many times, cells become sick." Atsumi has also inserted genes from other organisms into *E. coli* to induce the bacteria to produce various forms of butanol, as well as the five-carbon alcohol pentanol.

Atsumi's collaborator on the project — James Liao, a chemical and biomolecular engineer at the University of California, Los Angeles — recently produced a higher alcohol in a process that combines two sought-after advantages over ethanol production from corn. He started with a species of *Clostridium* that, unlike many butanol-producing strains, can digest cellulose, thereby enabling more of the available biomass to be used. Instead of pressing *E. coli* into service, Liao altered a pathway in *Clostridium* so that the bacterium produced the branched version of butanol — isobutanol. Isobutanol has the same chemical formula as butanol but a different structure that improves its engine performance and makes it easier to synthesize into other chemicals.

Several companies are trying to commercialize butanol or isobutanol production from biomass. In December 2010, the biotech company Green Biologics, of Abingdon, United Kingdom, announced a deal to provide its fermentation technology, based on a *Clostridium* strain, to two Chinese biochemical companies. Butamax Advanced Biofuels, a joint venture between the oil company BP and the chemical giant DuPont, has opened a demonstration plant in Hull, United Kingdom, and expects to have a commercial plant operating by 2013. And Gevo, an advanced biofuels company in Englewood, Colorado, is converting an ethanol production facility in Minnesota to produce about 68 million litres of isobutanol per year from 2012. Like ethanol, butanol would probably enter the market as a petrol blend.

Atsumi says that isobutanol yields need to increase by at least a couple of orders of magnitude to be economically viable. He doesn't see a need to coax microbes into making alcohols with longer carbon chains, because it's fairly easy to use conventional processes to convert isobutanol into other useful molecules. "Isobutanol is already a great biofuel," he says, adding that four- and five-carbon alcohols are "good enough for the fuel industry". Isobutanol can, for example, be dehydrated to form the hydrocarbon isobutene, which can in turn be used to make anything from petrol to jet fuel.

HYDROCARBON HEAVEN

No alcohol is likely to be the end point, however, if the aim is to fit into today's fuel storage and distribution system. "With alcohol fuel, and with ethanol in particular, you sacrifice a great deal of your gas mileage," says chemical engineer John Regalbuto of the University of Illinois at Chicago. Alcohol is "just not as energy dense as hydrocarbons," says Regalbuto, who is also a former director of the US National Science Foundation's catalysis and biocatalysis programme, which sponsors projects to convert biomass into fuel and other useful chemicals.

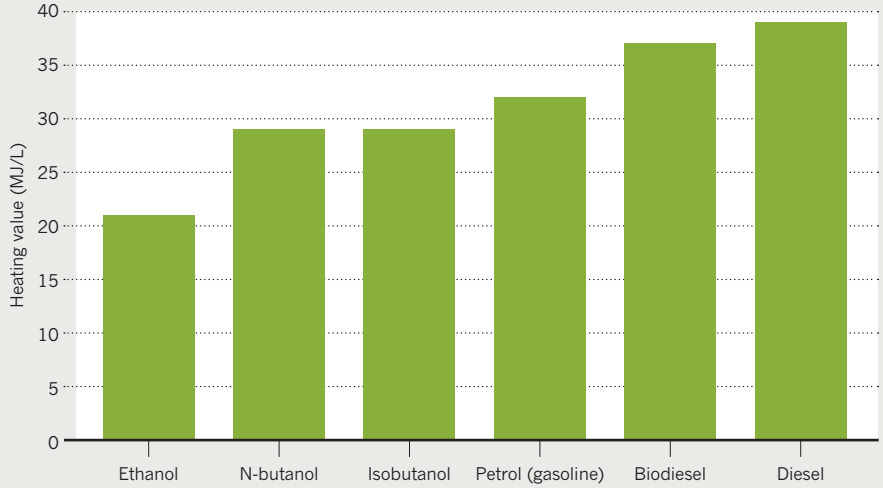
The closer the molecules that the microbes produce are to the molecules being burned in today's engines, Regalbuto says, the easier they'll fit into existing infrastructure. They'll be what biofuel experts call 'drop-in fuels'. Regalbuto likes to quote the motto of LS9, an industrial biotechnology company in South San Francisco, California, that is engineering microbes to produce chains of hydrocarbons: "The best replacement for petroleum is petroleum."

LS9 is one of several ventures using synthetic biology to redirect the fermentation process to produce hydrocarbons instead of alcohols. LS9's researchers identified genes that enable one kind of microbe, cyanobacteria, to naturally produce alkanes — one family of molecules consisting of only carbon and hydrogen atoms. The smallest alkanes (methane, propane and butane) are flammable gases, whereas petrol and diesel consist mainly of longer-chain alkanes. The company has been running a pilot-scale facility for more than two years and, in December 2010, announced it had raised US$30 million in investor funding to move towards commercial production.

Another way to get hydrocarbons from plant matter is to work with fatty oils, such as those produced by palm trees or soya. A chemical process called transesterification converts these oils to biodiesel. But because

PACKING HEAT

Biofuels vary in their energy density, but they all fall short of their petroleum-derived models.



HAN, L. ET AL. ANNU. REV. CHEM. BIOMOL. ENG. 1,19–36 (2009).

Pyrolysis: cooking wood chips (left) above 400 °C for a few seconds with a catalyst (middle) produces a 'bio-oil' (right).

DR. JOUNGMO CHO

of the requirement for land to grow the crops, Regalbuto says biodiesel is unlikely to be produced on a large enough scale to meet fuel demand. Much work is being done to use algae to produce oil for converting to biodiesel, but progress has fallen short of the enthusiastic projections (see 'A scum solution', page S15).

Not every approach to producing hydrocarbons relies on microbes to digest biomass. James Dumesic, a chemical and biological engineer at the University of Wisconsin–Madison, derives fuel from plant matter through a multi-step chemical process. He turns biomass into a clear liquid called gamma-valerolactone, or GVL. Like ethanol, GVL can be blended into petrol. But GVL has an important advantage: it can be processed further, to become a hydrocarbon.

To produce GVL, Dumesic applies sulphuric acid to the cellulose in corn stover (the stalks and leaves left over after harvesting), sawgrass or wood. By contrast, during fermentation, enzymes are often added to biomass to break down the cellulose into the simpler sugar glucose, which the microbe can handle. "Here, we're going right past the sugar," says Dumesic, adding that sulphuric acid is far less expensive than enzymes.

This step produces equal amounts of formic acid and levulinic acid. Mixing a catalyst made of ruthenium and carbon into the levulinic acid transforms it into GVL, which contains 97% of the energy from the original biomass. Whereas fermentation requires several days to convert biomass, catalysis takes just "tens of minutes", Dumesic says. GVL can be shipped through existing pipelines or tanker trucks to a refinery for further processing. There, heating at high pressure in the presence of zeolite (an alumino-silicate catalyst commonly used in petroleum cracking) converts GVL to butene plus carbon dioxide. The butene molecules can be combined (with the help of another common catalyst) to yield longer hydrocarbon chains for diesel or jet fuel.

One problem with this process is that the sulphur in the acid tends to deactivate the carbon–ruthenium catalyst, a problem Dumesic has solved by using a ruthenium–rhenium catalyst. Now he's working on developing catalysts that use a cheaper metal than ruthenium.

Some researchers are trying to improve chemical methods, such as gasification and pyrolysis, that have long been in use for converting biomass to hydrocarbon-based fuels. Gasification, which dates back more than a century, involves heating carbon-containing materials to high temperatures in the presence of oxygen. The resulting syngas can be burned as fuel or converted to liquid fuel using the Fischer–Tropsch synthesis, a process developed in the 1920s. Pyrolysis works in a similar way: biomass is heated to between 400 °C and 600 °C for a few seconds, in the absence of oxygen, and then cooled rapidly to produce a liquid known as bio-oil. Bio-oil, analogous to crude oil, is a mixture of compounds that can be 'upgraded' to hydrocarbon-based fuels.

This upgrading involves adding a large amount of hydrogen to the carbon in the oil, which can cost more than the bio-oil itself, says Huber, a former student of Dumesic's. Huber has developed a pyrolysis process that, through the addition of a catalyst, doesn't stop at the bio-oil stage. The biomass is ground up and rapidly heated, and the resultant vapours flow through zeolites, which convert the vapours to benzene, toluene and xylene. These aromatic hydrocarbons can then be blended to yield a fuel that can be used in high-performance cars, for instance, as these require a high percentage of toluene. The whole process takes only minutes. "We think it's going to be significantly cheaper than gasification or fermentation," says Huber. The university has licensed his technology to the New York-based start-up company Anellotech, which Huber co-founded. "As long as we have a cheap feedstock, we can make our products at under $3 a gallon," says Huber. This April, gasoline prices were about US$4 per gallon in the United States (about US$1 per litre) and about US$2 per litre in the United Kingdom.

THE ONCE AND FUTURE KING

Huber doesn't foresee a single technology emerging as the king of biofuel processing. Instead, he says, there will be a mix that makes the best use of available resources and fits in with the various demands for fuels. "The future biorefinery is going to be like the petroleum refinery today," Huber predicts. "You're going to have a series of different units that all make different products."

But fuel will continue to be made of the same compounds that it is now. There's no reason to try to invent some new liquid, says George Church, a geneticist at Harvard Medical School in Boston, Massachusetts, because "alkanes are still a pretty good fuel". There's no better way to store energy for transport; petrol is "like a battery that's 50 or 100 times higher in energy density", says Church, whose synthetic biology research has contributed to LS9's technology and that of other biofuel companies.

Regalbuto is optimistic that biomass-derived, hydrocarbon-based fuel will soon slip seamlessly into everyday use. "I wouldn't be surprised if we're putting 'green gasoline' in our gas tank in five to seven years," he says. "And we won't even know it, because it will be a drop-in replacement." Longer term, he expects conventional cars, with their tanks of liquid fuel, will give way to battery-powered vehicles that depend on electricity generated from a mix of nuclear and renewable energy sources. Heavier vehicles — boats, aircraft, tanks and trucks — will rely on biofuel. Such a strategy, he says, could enable oil-dependent economies to end their reliance on imported petroleum. "Electricity for the light vehicles, biomass for the heavies, and we're energy independent in two decades," he says.

"Electricity for the light vehicles, biomass for the heavies, and we're energy independent in two decades."

Liao, who thinks the most promising feedstock will be algae, says a biofuel will be successful only if it can be made affordably and in large volume. "It has to be something that can be produced at the rate that we currently dig out oil from underground," he says. "Then we can talk about replacing petroleum." ■

Neil Savage *is a science writer based in Lowell, Massachusetts.*

LETTER

doi:10.1038/nature10082

Tracking apex marine predator movements in a dynamic ocean

B. A. Block[1], I. D. Jonsen[2], S. J. Jorgensen[1], A. J. Winship[2], S. A. Shaffer[3], S. J. Bograd[4], E. L. Hazen[4], D. G. Foley[4], G. A. Breed[2,5], A.-L. Harrison[5], J. E. Ganong[1], A. Swithenbank[1], M. Castleton[1], H. Dewar[6], B. R. Mate[7], G. L. Shillinger[1], K. M. Schaefer[8], S. R. Benson[9], M. J. Weise[5], R. W. Henry[5] & D. P. Costa[5]

Pelagic marine predators face unprecedented challenges and uncertain futures. Overexploitation and climate variability impact the abundance and distribution of top predators in ocean ecosystems[1–4]. Improved understanding of ecological patterns, evolutionary constraints and ecosystem function is critical for preventing extinctions, loss of biodiversity and disruption of ecosystem services. Recent advances in electronic tagging techniques have provided the capacity to observe the movements and long-distance migrations of animals in relation to ocean processes across a range of ecological scales[5,6]. Tagging of Pacific Predators, a field programme of the Census of Marine Life, deployed 4,306 tags on 23 species in the North Pacific Ocean, resulting in a tracking data set of unprecedented scale and species diversity that covers 265,386 tracking days from 2000 to 2009. Here we report migration pathways, link ocean features to multispecies hotspots and illustrate niche partitioning within and among congener guilds. Our results indicate that the California Current large marine ecosystem and the North Pacific transition zone attract and retain a diverse assemblage of marine vertebrates. Within the California Current large marine ecosystem, several predator guilds seasonally undertake north–south migrations that may be driven by oceanic processes, species-specific thermal tolerances and shifts in prey distributions. We identify critical habitats across multinational boundaries and show that top predators exploit their environment in predictable ways, providing the foundation for spatial management of large marine ecosystems.

Large marine predators are important top consumers in marine ecosystems, and their depletion can have cascading effects on lower trophic levels in both coastal and open-ocean waters[3,7,8]. Although some taxa are protected internationally from direct harvest, intense overexploitation by high-seas fisheries and anthropogenic environmental change are threatening global marine biodiversity. The management and conservation of highly migratory marine predators depends on understanding how movements relate to ocean processes; however, the dynamic relationships among behaviour, physiology and oceanography over a range of temporal and spatial scales remains unclear[4,9,10].

Tagging of Pacific Predators (TOPP) examined the distribution of species that inhabit and migrate into the coastal upwelling region off the west coast of North America. Using simultaneous tag deployments, we tracked seven top-predator guilds (sharks, tunas, albatrosses, shearwaters, turtles, pinnipeds and rorqual whales) concurrently for nearly a decade. Habitat use, residency patterns and migration pathways to North American waters were previously unknown or poorly documented for most of these animals[5,6,11–13].

Deployments of 4,306 electronic tags yielded 1,791 individual animal tracks from 23 species, totalling 265,386 animal tracking days (Fig. 1 and Supplementary Tables 1–3). The tracking technologies used had different location precisions, and include ARGOS satellite tags ($n = 1{,}183$), archival and satellite transmitting tags ($n = 1{,}008$) and archival geolocation tags ($n = 2{,}115$); the latter two provide estimates of position based on sunrise and sunset times and sea surface temperature (SST). We used a state-space modelling approach[14] to estimate daily locations, and their uncertainty, from each track while accounting for the differing precisions of the tracking technologies.

The tracking data show that the California Current large marine ecosystem (CCLME; Supplementary Fig. 1) is an important habitat (Figs 2–4) for tunas (Pacific bluefin, *Thunnus orientalis*; yellowfin, *Thunnus albacares*; albacore, *Thunnus alalunga*), sharks (shortfin mako, *Isurus oxyrinchus*; white, *Carcharodon carcharias*; salmon, *Lamna ditropis*; blue, *Prionace glauca*; common thresher, *Alopias vulpinus*), cetaceans (blue whale, *Balaenoptera musculus*), pinnipeds (northern elephant seal, *Mirounga angustirostris*; California sea lion, *Zalophus californianus*), seabirds (Laysan albatross, *Phoebastria immutabilis*; black-footed albatross, *Phoebastria nigripes*; sooty shearwater, *Puffinus griseus*) and sea turtles (leatherback, *Dermochelys coriacea*; loggerhead, *Caretta caretta*). Annual migratory periodicity was evident in the movements of many tagged animals that showed fidelity to the cool, nutrient-rich waters of the CCLME (Figs 2 and 3). Extended residency within the CCLME was revealed by examining tracks that spanned multiple seasons, using a behavioural switching state-space model[14]. Numerous species (leatherback sea turtles, black-footed albatrosses, sooty shearwaters, bluefin tunas and salmon sharks; Fig. 2b) had a strong attraction to the CCLME and undertook long migrations (>2,000 km) from the western, central or south Pacific basin. Some species (bluefin and yellowfin tunas; mako, white and salmon sharks; blue whales; male elephant seals; and leatherback sea turtles; Fig. 3a and Supplementary Fig. 2) undertook a seasonally recurring north–south migration in the North Pacific and within the CCLME. Other taxa undertook movements between near-shore and offshore waters, with a residency period within the CCLME or the Gulf of Alaska, followed by an offshore migration that ranged into the North Pacific transition zone (NPTZ) (female elephant seals, salmon sharks and Laysan albatrosses), the subtropical gyre and north equatorial current (blue and mako sharks and leatherback sea turtles), or the 'Café' region of the eastern Pacific and the Hawaiian Islands (white sharks; albacore tunas, *Thunnus alalunga*; and black-footed albatrosses). The mechanisms and cues underlying fidelity to seasonally modulated migration pathways are not entirely known, but may represent a capacity to discriminate among areas of seasonal significance for foraging or reproduction.

The quarterly predator density in relation to the median satellite-derived SST measurements within the CCLME (Fig. 3b and Supplementary Fig. 3) suggests that seasonal warming (quarter three) along the Baja California peninsula may trigger northward movements of predator populations. Increased metabolic expenditures in ectotherms

[1]Stanford University, Biology Department, Hopkins Marine Station, Pacific Grove, California 93950, USA. [2]Dalhousie University, Department of Biology, Halifax, Nova Scotia B3H 4J1, Canada. [3]San Jose State University, Department of Biological Sciences, San Jose, California 95192, USA. [4]NOAA Southwest Fisheries Science Center, Environmental Research Division, Pacific Grove, California 93950, USA. [5]University of California, Santa Cruz, Department of Ecology & Evolutionary Biology, Long Marine Laboratory, Santa Cruz, California 95060, USA. [6]NOAA Southwest Fisheries Science Center, Fisheries Research Division, La Jolla, California 92037, USA. [7]Oregon State University, Hatfield Marine Science Center, Newport, Oregon 97365, USA. [8]Inter-American Tropical Tuna Commission, La Jolla, California 92037, USA. [9]NOAA Southwest Fisheries Science Center, Protected Resources Division, La Jolla, California 92037, USA.

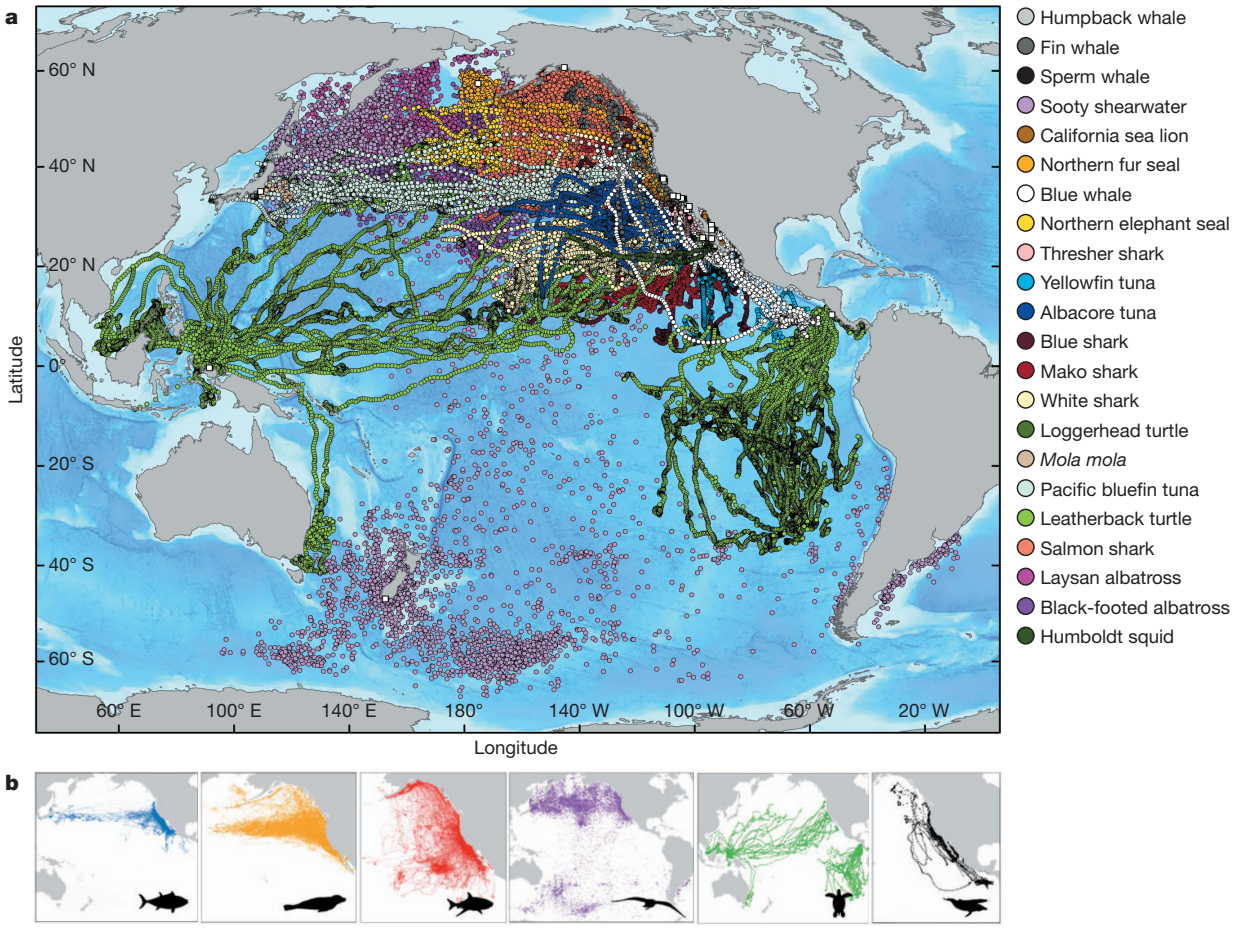


Figure 1 | All TOPP species state space position estimates and distribution from electronic tagging. **a**, Daily mean position estimates (circles) and annual median deployment locations (white squares) of all tagged species. **b**, Daily mean position estimates of the major TOPP guilds (from left): tunas (yellowfin, bluefin and albacore), pinnipeds (northern elephant seals, California sea lions and northern fur seals), sharks (salmon, white, blue, common thresher and mako), seabirds (Laysan and black-footed albatrosses and sooty shearwaters), sea turtles (leatherback and loggerhead) and cetaceans (blue, fin, sperm and humpback whales).

as the ocean warms, subsurface hypoxia[15] and declines in primary productivity (Fig. 3b and Supplementary Fig. 3g) all potentially decrease the suitability of this habitat. In the northern waters of the CCLME (Fig. 3b and Supplementary Fig. 3a–d), cooler temperatures (9–11 °C), particularly during the spring upwelling season, limit cardiac function[13,16] in many fishes and sharks, reducing their capacity to exploit these environs. Predators were observed moving south into warmer water despite regions of persistent productivity in the north (Fig. 3b). A similar relationship may hold for endotherms, such as seabirds, pinnipeds and whales, but the linkage is indirect as the physiological constraints are largely on their prey.

Tracking data typically were acquired for less than one year; however, for some taxa (tunas, turtles and sharks) multiyear tracks were obtained (maximum of 1,380 days, for salmon sharks; Supplementary Table 1 and Supplementary Fig. 4). To account for abbreviated tracks and for variation in sample size among taxa, we applied time weighting and species normalization schemes before examining multispecies density patterns (Supplementary Methods). The relative spatial density of large marine predators indicates that high-use areas in the northeastern Pacific occur both on the high seas (NPTZ; Fig. 4a and Supplementary Fig. 1) and within the exclusive economic zones of North America (CCLME; Fig. 4b). These observed patterns differ markedly from those expected

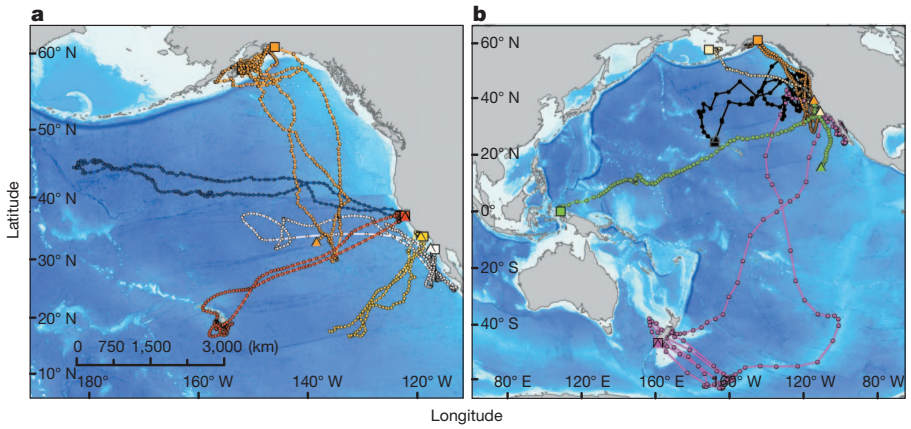


Figure 2 | Fidelity and attraction to the CCLME. **a**, Examples of pelagic predators released and electronically tracked in the CCLME that show fidelity to deployment locations and the CCLME. We show the release locations (square), pop-up satellite end point locations (triangle) and daily mean positions (circles) of the following species: yellowfin tuna (yellow), bluefin tuna (white), white shark (red), elephant seal (blue) and salmon shark (orange). **b**, Individual tracks of pelagic animals released >2,000 km away from the CCLME that are indicative of cross-basin or ecosystem attraction to, and temporary residency within, the eastern North Pacific. Symbols are as in **a**, for leatherback sea turtles (green), sooty shearwaters (pink), fur seals (pale yellow), black-footed albatrosses (black) and salmon sharks (orange).

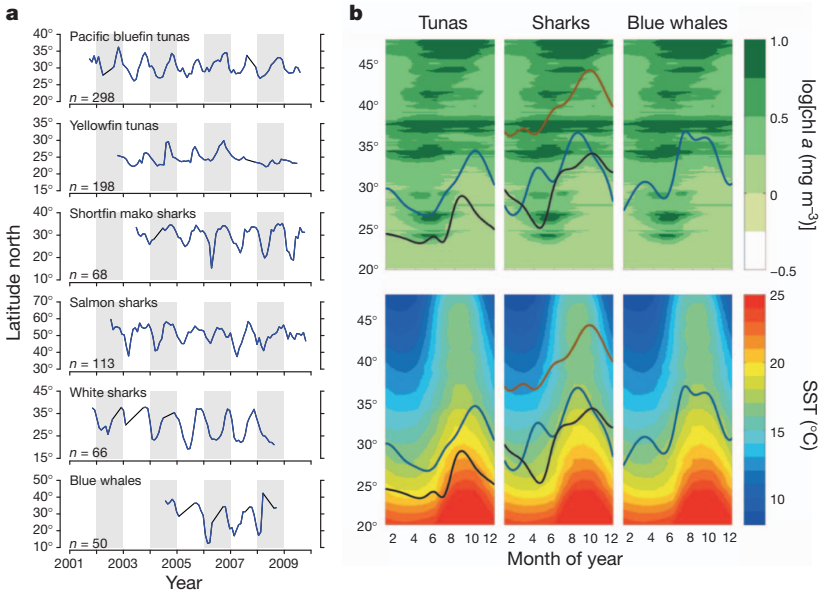


Figure 3 | Latitudinal migration cycles and seasonal climatologies within the CCLME. **a**, Monthly mean latitudes of predators residing within or migrating to the CCLME. Black line segments denote gaps where no data were available. Sample sizes indicate the numbers of individual tracks contributing to the time series. **b**, Seasonal climatologies in the California Current for tunas (Pacific bluefin, blue; yellowfin, black), sharks (salmon shark, brown; shortfin mako, black; white shark, blue) and blue whales relative to median chlorophyll *a* densities and SST values between 2000 and 2009.

under a null model of random diffusion from tag deployment locations (Supplementary Fig. 5 and Supplementary Methods), confirming that the observed density patterns are not driven by the tag deployment locations.

The CCLME is a highly retentive area for many species tagged there, and is an attractive area for animals undergoing long migrations from the western and central North Pacific and the Gulf of Alaska (Figs 2 and 4a, b). Pacific bluefin and yellowfin tunas spent significantly more time in the CCLME than expected on the basis of null model simulations (Supplementary Fig. 6). Several species (tunas, white, mako and salmon sharks, leatherback sea turtles and blue whales) had substantial residency periods within, or were return migrants to, the CCLME, according to behavioural inferences from a switching state-space model[14] (Fig. 4c). Additionally, all species tagged outside the CCLME spent significantly more time on average in the CCLME than expected on the basis of null model simulations (Supplementary Fig. 6). Representatives from several guilds exhibited cross-basin migrations (>2,000 km) into the CCLME from the western (leatherback sea turtles and bluefin tunas), central (black-footed albatrosses and salmon sharks) and south Pacific basins (sooty shearwaters; Fig. 2b). The retention within and attraction to the CCLME is consistent with the high productivity of this region, which supports large biomasses of krill, sardines, anchovies, salmon, groundfish and squid that provide a predictable forage base for top predators[2,17]. The NPTZ is another important region, serving as an east–west migration corridor (Fig. 4a) and foraging region for elephant seals, salmon and blue sharks (Fig. 4c), Laysan and black-footed albatrosses, and bluefin tunas (Fig. 1). This is a complex region encompassing an abrupt north-to-south transition between sub-Arctic and subtropical water masses with dynamic frontal regions[18].

To investigate which aspects of the biophysical environment putatively attract these predators, we explored both presence/absence and relative habitat use with generalized additive mixed models[19]. We examined the collective response of 16 marine predator species to environmental covariates (Supplementary Fig. 7 and Supplementary Methods). In the binary presence/absence model, predator incidence showed a strong positive relationship with SST across a broad temperature range that peaked near 15 °C (Supplementary Fig. 7a). Tagged animals occupied a small portion of cool, nutrient-rich water in coastal regions and northern latitudes in preference to the broadly available warm oligotrophic waters in lower latitudes (Supplementary Fig. 3). Relative density models showed a strong positive relationship between density and chlorophyll *a* suggesting that the suite of tagged species preferentially occupy regions of high productivity (Supplementary Fig. 7c). The observed patterns of predator distribution in this study may be indicative of trade-offs between access to areas of higher productivity and physiological thermal tolerances, either directly by the predators or indirectly by their lower-trophic-level prey.

To examine how closely related taxa partition marine resources, we compared thermal preferences from *in situ*, tag-based SST measurements for sympatrically occurring species within three guilds (albatrosses, tunas and sharks). Differences in habitat use evident among congeneric species illustrate how more-recently divergent species partition the oceanic environment (Fig. 5). During the June–November post-breeding phase, black-footed albatrosses were associated with a broader range of higher SSTs primarily in the eastern Pacific, whereas Laysan albatrosses were associated with a narrower range of lower SSTs in the western and central North Pacific (Fig. 5a and Supplementary Table 9). Bluefin tunas ranged farther north in the colder waters of the CCLME, whereas yellowfin tunas occupied warmer waters in the southern CCLME (Fig. 5b and Supplementary Table 10). These differences are consistent with physiological specializations in bluefin tuna cardiac performance[16]. The lamnid sharks had a more complex separation of habitats. Salmon sharks, with their cold-tolerant cardiac physiology[13], occupied the cooler, sub-Arctic waters in the North Pacific. Most salmon sharks, but not all, migrated seasonally into the warmer NPTZ and CCLME waters (Fig. 5c and Supplementary Table 11). White sharks overlapped with salmon sharks in the near-shore CCLME but also migrated into warmer, offshore waters of the subtropical gyre[12] and the Hawaiian Islands (Fig. 5c and Supplementary Table 11). Shortfin mako sharks were distributed throughout the CCLME and into the subtropical gyre but occupied a thermal range intermediate between the two modes of the white shark range (Fig. 5c and Supplementary Table 11).

Species such as black-footed albatrosses, sooty shearwaters, leatherback sea turtles and salmon sharks have evolved migration or foraging strategies that entail long-distance journeys to productive foraging habitats within the CCLME and the NPTZ. Shearwaters[6] and leatherback sea turtles tagged in the south and western Pacific arrived in the central and northern portions of the CCLME each year, when SSTs were seasonally highest. These patterns indicate that resources within the CCLME and the NPTZ are valuable to species throughout the Pacific basin, highlighting the importance of maintaining ecological connectivity across hemispheres. It is likely that other areas of the Pacific, not extensively covered by our tagging efforts, are similarly important. For example, estimates of global marine species richness peak around Southeast Asia, and richness of primarily oceanic taxa peaks at latitude 30° (north or south) along boundary currents such as the Kuroshio in the western Pacific[20].

The concentration and residency of multiple predators at predictable locations in the Pacific reveals the importance of these regions as biological hotspots. Given the high density of tracked animals in these regions, predators may have an important top-down role in structuring

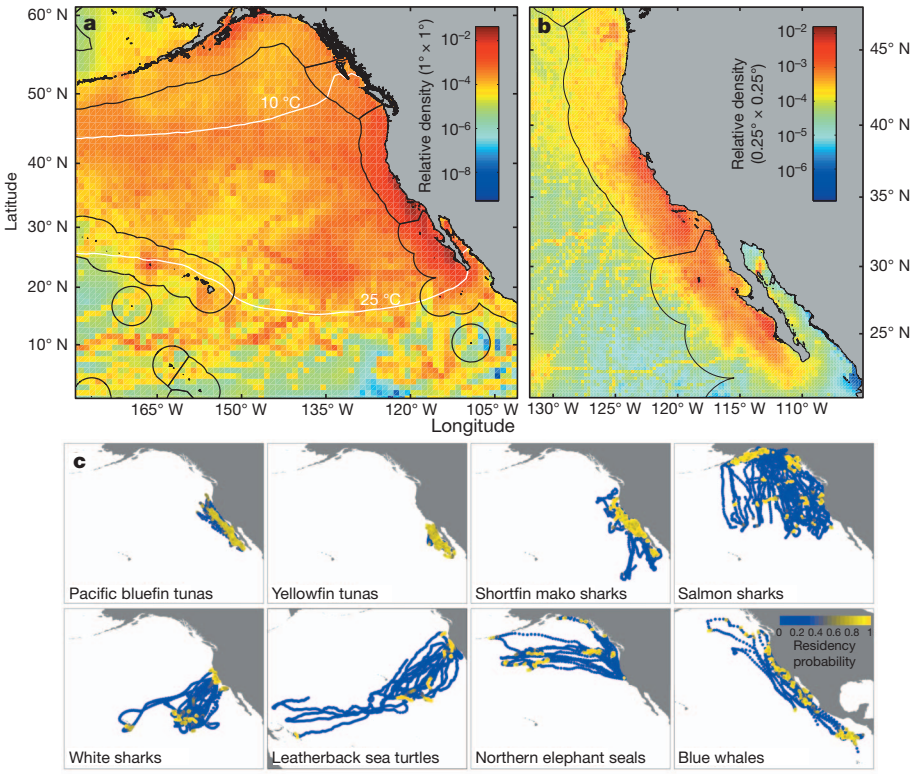


Figure 4 | Predator density maps and residency patterns. **a**, Density of large marine predators within the eastern North Pacific. Densities of the time-weighted and species-normalized position estimates of all tagged individuals were summed within $1^\circ \times 1^\circ$ grid cells. **b**, Density of large marine predators within the CCLME at a $0.25^\circ \times 0.25^\circ$ resolution. **c**, Patterns of resident (slow, area-restricted movements) versus transient (fast, directed movements) behaviours of predators that primarily occupied or migrated to the CCLME, estimated using a switching state-space model. The coloured points grading from blue to yellow display the posterior mean probability of the resident behaviour associated with each daily mean position estimate. Each panel displays residency patterns for ten individuals. Uncertainty in position estimates in **a** and **b** is included by calculating densities using all 2000 Markov chain Monte Carlo samples from the joint posterior distributions of the daily positions, rather than using only the posterior means. SST contours in **a** are denoted by solid white lines. Exclusive economic zones are delineated by solid black lines.

these ecosystems. The impact that the removal of top predators has had on the trophic dynamics of coastal and oceanic ecosystems is contentious[3,21]. For example, a failure to manage marine resources effectively in the Atlantic has led to a decline of pelagic species such as the Atlantic bluefin tuna[3] and the porbeagle shark[22], and has brought demersal fish, such as cod, to the brink of ecological extinction[23]. In the Pacific, commercial fisheries removed over 50 million tonnes of tunas and sharks from pelagic ecosystems between 1950 and 2004[21], resulting in significant declines in populations of large fish. Furthermore, unsustainable levels of fishery bycatch mortality has decimated populations of Pacific sea turtle (leatherback and loggerhead) and threatened some populations of albatross[24,25]. The impact of these top-predator biodiversity losses on ecosystem stability is unknown. Establishing the population baselines for top-predator populations and conserving remaining populations in regions such as the CCLME, and quantifying the ecosystem services they provide, remains an important management priority[26]. Our study provides a clearer understanding of where predators are concentrated in the North Pacific and which environmental parameters can be used to predict their occurrence.

Tracking data can be used to examine behaviours and distributions, inform population assessments and improve estimates of natural and fishery-induced mortality[27–29]. Despite their high monetary and conservation value, population assessments of many predators in this study are rare or non-existent. Long-term observations of predator movements provide information on the spatial extent of highly migratory populations and potential rates of exchange among them. High archival tag returns (Supplementary Table 1) for bluefin (56%) and yellowfin (39%) tunas indicate that high fishery-induced mortality currently occurs within the CCLME. The recovery of cetacean populations, such as blue whales, which are endangered, can be enhanced through the use of electronic tracking data to identify high-use areas and coordinate policy actions that mitigate anthropogenic risks[30] (for example ship strikes). Finally, an improved understanding of the horizontal and vertical spatiotemporal distribution of leatherback sea turtles and North Pacific albatrosses in relation to pelagic fisheries could better inform management protocols (for example time–area closures and gear mitigation) to reduce bycatch of these endangered and threatened species.

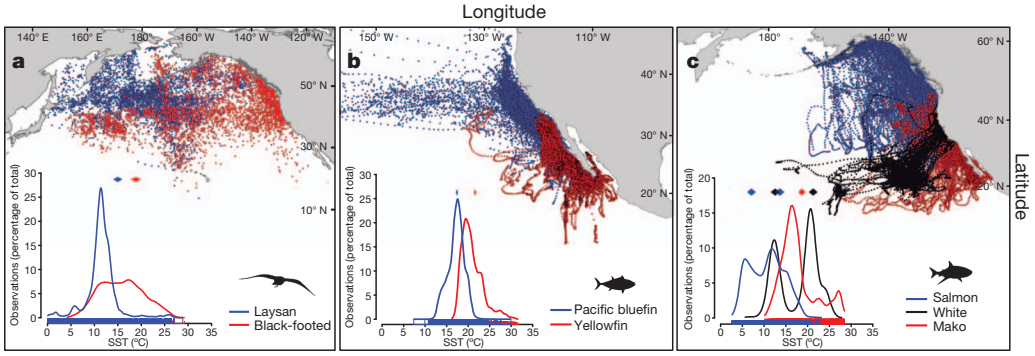


Figure 5 | Niche separation within three predator guilds. Spatial distribution and thermal habitat use (insets) across three guilds of sympatric species: Laysan and black-footed albatrosses ($n = 55$ individuals, 8,743 daily SSTs; **a**), Pacific bluefin and yellowfin tunas ($n = 376$ individuals, 75,177 daily SSTs; **b**), lamnid (salmon, mako and white) sharks ($n = 137$ individuals, 12,971 daily SSTs; **c**). SST profiles are daily means of tag-derived SSTs. Linear mixed-effects model estimates of mean (±95% confidence interval) SST for each species are displayed as diamonds at the top of each inset graph. The distribution of daily mean SSTs for each species is displayed along the *x* axes.

The decadal record of multitaxa electronic tagging generated by TOPP provides information on cross-border movements between US, Mexican and Canadian waters that has important implications for conservation and management efforts. This extensive data set serves as a baseline for monitoring and forecasting these seasonal patterns and assessing shifts in abundance among countries. Linking movements to ocean temperatures can help predict the potential ecosystem-level effects of climate variability. This baseline information could provide the foundation for launching a management scheme on a scale never before achieved for a pelagic ecosystem. International policy vehicles such as a UNESCO Marine World Heritage designation could provide an effective approach to generating global interest and, in turn, funding to support ecosystem-based management efforts within the CCLME. Furthermore, if a conservation corridor was extended across the NPTZ, key ecological foraging hotspots and migratory corridors that link the eastern and western Pacific basins for transoceanic migrants could be sustainably managed. Without an aggressive effort to zone and effectively manage these resources, the predator populations they support will decline and the biodiversity of this open-ocean wilderness will be irreplaceably lost.

METHODS SUMMARY

We used a Bayesian state-space model[14] to account for observation error among tag types and to obtain regular daily location and uncertainty estimates. Argos satellite errors were estimated previously[14]. Errors for archival geolocation and pop-up satellite archival tags were estimated from taxon-specific double-tagging experiments (Supplementary Methods) and captive sea pen studies. We used a behavioural switching state-space model[14] to infer patterns of residency versus transience from a subset of species' tracks that at least partly inhabited the CCLME.

Unnaturally abbreviated tracks were normalized by weighting each location estimate by the inverse of the number of individuals of that species that had location estimates for the same relative day of their track. We imposed a threshold relative day of track (85th percentile) above which locations received the same weight as on the threshold day. We normalized the data set by species to account for the over-representation of species with large sample sizes. The normalized and weighted tracks were summed within $1^\circ \times 1^\circ$ or $0.25^\circ \times 0.25^\circ$ grid cells to obtain relative densities of 16 predator species across the North Pacific and within the CCLME, respectively. The influence of the tag deployment locations on the spatial density pattern was assessed through null model simulations. The relative density of predator species in the $1^\circ \times 1^\circ$ grid cells was related to oceanographic variables using generalized additive mixed models[19], with season and year as random covariates. Satellite-derived chlorophyll *a*, SST and sea surface height anomalies and variances, and wind stress curl, were summarized quarterly in each grid cell between June 2002 and June 2009. A two-model approach dealt with over-dispersion, non-normality and non-constant variance of the relative density data. We used a binary presence/absence habitat model with a logit link function to model the occurrence of predators within grid cells. All cells with non-zero density were log-transformed and fitted to environmental correlates to create a relative-use habitat model.

Further details of the methods are included in Supplementary Information.

Received 31 December 2010; accepted 1 April 2011.
Published online 22 June 2011.

1. Robinson, R. A. *et al.* Travelling through a warming world: climate change and migratory species. *Endanger. Species Res.* **7,** 87–99 (2009).
2. Chavez, F. P. *et al.* From anchovies to sardines and back: multidecadal change in the Pacific ocean. *Science* **299,** 217–221 (2003).
3. Myers, R. A. & Worm, B. Rapid worldwide depletion of predatory fish communities. *Nature* **423,** 280–283 (2003).
4. Costa, D. P. *et al.* Approaches to studying climate change and habitat selection of Antarctic pinnipeds. *Integr. Comp. Biol.* **50,** 1018–1030 (2010).
5. Shillinger, G. *et al.* Persistent leatherback turtle migration corridor presents opportunities for conservation. *PLoS Biol.* **6,** e171 (2008).
6. Shaffer, S. A. *et al.* Migratory shearwaters integrate oceanic resources across the Pacific Ocean in an endless summer. *Proc. Natl Acad. Sci. USA* **103,** 12799–12802 (2006).
7. Scheffer, M., Carpenter, S. & de Young, B. Cascading effects of overfishing marine systems. *Trends Ecol. Evol.* **20,** 579–581 (2005).
8. Springer, A. M. *et al.* Sequential megafaunal collapse in the North Pacific Ocean: an ongoing legacy of industrial whaling? *Proc. Natl Acad. Sci. USA* **100,** 12223–12228 (2003).
9. Steele, J. H. & Henderson, E. W. Coupling between physical and biological scales. *Phil. Trans. R. Soc. Lond. B* **343,** 5–9 (1994).
10. Whitehead, H., McGill, B. & Worm, B. Diversity of deep-water cetaceans in relation to temperature: implications for ocean warming. *Ecol. Lett.* **11,** 1198–1207 (2008).
11. Bailey, H. *et al.* Behavioural estimation of blue whale movements in the Northeast Pacific from state-space model analysis of satellite tracks. *Endanger. Species Res.* **10,** 93–106 (2009).
12. Jorgensen, S. J. *et al.* Philopatry and migration of Pacific white sharks. *Proc. R. Soc. Lond. B* **277,** 679–688 (2010).
13. Weng, K. C. *et al.* Satellite tagging and cardiac physiology reveal niche expansion in salmon sharks. *Science* **310,** 104–106 (2005).
14. Jonsen, I. D., Flemming, J. M. & Myers, R. A. Robust state-space modeling of animal movement data. *Ecology* **86,** 2874–2880 (2005).
15. Prince, E. & Goodyear, C. Hypoxia-based habitat compression of tropical pelagic fishes. *Fish. Oceanogr.* **15,** 451–464 (2006).
16. Shiels, H. A., Di Maio, A., Thompson, S. & Block, B. A. Warm fish with cold hearts: thermal plasticity of excitation-contraction coupling of bluefin tuna. *Proc. R. Soc. Lond. B* **278,** 18–27 (2011).
17. PICES. *Marine Ecosystems of the North Pacific.* PICES Spec. Publ. 1 (North Pacific Marine Science Organization, 2004).
18. Polovina, J. J., Howell, E., Kobayashi, D. R. & Seki, M. P. The transition zone chlorophyll front, a dynamic global feature defining migration and forage habitat for marine resources. *Prog. Oceanogr.* **49,** 469–483 (2001).
19. Wood, S. *Generalized Additive Models: An Introduction with R* (Chapman & Hall/CRC, 2006).
20. Tittensor, D. P. *et al.* Global patterns and predictors of marine biodiversity across taxa. *Nature* **466,** 1098–1101 (2010).
21. Sibert, J., Hampton, J., Kleiber, P. & Maunder, M. Biomass, size and trophic status of top predators in the Pacific ocean. *Science* **314,** 1773–1776 (2006).
22. Campana, S. E. *et al.* Population dynamics of the porbeagle in the northwest Atlantic Ocean. *N. Am. J. Fish. Manag.* **22,** 106–121 (2002).
23. Swain, D. P. & Chouinard, G. A. Predicted extirpation of the dominant demersal fish in a large marine ecosystem: Atlantic cod (*Gadus morhua*) in the southern Gulf of St. Lawrence. *Can. J. Fish. Aquat. Sci.* **65,** 2315–2319 (2008).
24. Peckham, S. H. *et al.* Small-scale fisheries bycatch jeopardizes endangered Pacific loggerhead turtles. *PLoS ONE* **2,** e1041 (2007).
25. Lewison, R. L. & Crowder, L. B. Estimating fishery bycatch and effects on a vulnerable seabird population. *Ecol. Appl.* **13,** 743–753 (2003).
26. Halpern, B. S. *et al.* Mapping cumulative human impacts to California Current marine ecosystems. *Conserv. Lett.* **2,** 138–148 (2009).
27. Hays, G. C. *et al.* Satellite telemetry suggests high levels of fishing-induced mortality in marine turtles. *Mar. Ecol. Prog. Ser.* **262,** 305–309 (2003).
28. Heupel, M. & Simpfendorfer, C. Estimation of mortality of juvenile blacktip sharks, *Carcharhinus limbatus*, within a nursery area using telemetry data. *Can. J. Fish. Aquat. Sci.* **59,** 624–632 (2002).
29. Kurota, H. *et al.* A sequential Bayesian methodology to estimate movement and exploitation rates using electronic and conventional tag data: application to Atlantic bluefin tuna (*Thunnus thynnus). Can. J. Fish. Aquat. Sci.* **66,** 321–342 (2009).
30. Berman-Kowalewski, M. *et al.* Association between blue whale (*Balaenoptera musculus*) mortality and ship strikes along the California coast. *Aquat. Mamm.* **36,** 59–66 (2010).

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

Acknowledgements This manuscript is the culmination of a Census of Marine Life cross-project synthesis between TOPP and Future of Marine Animal Populations (FMAP). Funding for this work was provided by the Sloan Foundation's Census of Marine Life programme. TOPP research was funded by the Sloan, Packard and Moore foundations. FMAP was funded by the Sloan Foundation. Electronic tagging and tracking in TOPP was also supported by the Office of Naval Research, the NOAA, the E&P Sound and Marine Life JIP under contract from the OGP, and the Monterey Bay Aquarium Foundation. We thank the TOPP scientific teams and all those who supported animal tagging efforts, R. Kochevar and D. Kohrs for their dedication and their effort on behalf of the Census of Marine Life. We are grateful to the numerous captains and crews who provided ship time and logistical support, and to the US Fish and Wildlife Service in Hawaii for further logistical support. We thank the Mexican authorities and collaborating TOPP partners (O. Sosa-Nishizki) for permitting and assisting in research in their waters. All animal research was conducted in accordance with IACUC protocols from Stanford University and the University of California.

Author Contributions This synthesis study was initiated by B.A.B. and I.D.J. The TOPP project was designed and coordinated by B.A.B., D.P.C. and S.J.B. B.A.B., S.J.J., H.D. and K.M.S. designed experiments and deployed electronic tags on fish and sharks. D.P.C., S.A.S., R.W.H., M.J.W. and B.R.M. designed experiments and deployed electronic tags on marine mammals and birds. G.L.S., B.A.B. and S.R.B. designed experiments and deployed electronic tags on sea turtles. Tracking data were compiled by S.J.J., S.A.S., G.A.B., A.-L.H., B.A.B., G.L.S. and M.C. Data management was coordinated by A.S. and J.E.G. Oceanographic data were compiled by S.J.B., E.L.H. and D.G.F. Analyses were conducted by A.J.W., S.J.J., I.D.J., G.A.B, E.L.H., D.G.F., A.-L.H., J.E.G. and A.S. Figures were created by B.A.B., M.C., A.-L.H., I.D.J., S.J.J., A.J.W., J.E.G., A.S., E.L.H. and D.G.F. The manuscript was drafted by B.A.B. and edited by I.D.J., D.P.C., S.J.J., S.A.S., S.J.B., E.L.H., A.-L.H., A.J.W., H.D., G.L.S. and B.R.M.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.A.B. (bblock@stanford.edu).

LETTER

doi:10.1038/nature10118

XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast

E. L. van Dijk[1]*, C. L. Chen[1]*, Y. d'Aubenton-Carafa[1], S. Gourvennec[2], M. Kwapisz[2], V. Roche[2], C. Bertrand[2], M. Silvain[1], P. Legoix-Né[3], S. Loeillet[4], A. Nicolas[4], C. Thermes[1] & A. Morillon[1,2]

Non-coding (nc)RNAs are key players in numerous biological processes such as gene regulation, chromatin domain formation and genome stability[1,2]. Large ncRNAs interact with histone modifiers[3–5] and are involved in cancer development[6], X-chromosome inactivation[7] and autosomal gene imprinting[8]. However, despite recent evidence showing that pervasive transcription is more widespread than previously thought[9], only a few examples mediating gene regulation in eukaryotes have been described[10]. In *Saccharomyces cerevisiae*, the bona-fide regulatory ncRNAs are destabilized by the Xrn1 5′–3′ RNA exonuclease[11,12] (also known as Kem1), but the genome-wide characterization of the entire regulatory ncRNA family remains elusive. Here, using strand-specific RNA sequencing (RNA-seq), we identify a novel class of 1,658 Xrn1-sensitive unstable transcripts (XUTs) in which 66% are antisense to open reading frames. These transcripts are polyadenylated and RNA polymerase II (RNAPII)-dependent. The majority of XUTs strongly accumulate in lithium-containing media, indicating that they might have a role in adaptive responses to changes in growth conditions. Notably, RNAPII chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) analysis of Xrn1-deficient strains revealed a significant decrease of RNAPII occupancy over 273 genes with antisense XUTs. These genes show an unusual bias for H3K4me3 marks and require the Set1 histone H3 lysine 4 methyl-transferase for silencing. Furthermore, abolishing H3K4me3 triggers the silencing of other genes with antisense XUTs, supporting a model in which H3K4me3 antagonizes antisense ncRNA repressive activity. Our results demonstrate that antisense ncRNA-mediated regulation is a general regulatory pathway for gene expression in *S. cerevisiae*.

Antisense ncRNAs in yeast control Ty1 mobility[12,13] and *PHO84* transcription through histone modifications[11,12] in an RNA-interference-independent manner[14] (Supplementary Fig. 1). The regulatory ncRNAs are distinct from the canonical cryptic unstable transcripts (CUTs) of shorter size, destabilized by the nuclear exosome[15]. In contrast, the regulatory antisense ncRNAs are mainly degraded by the cytoplasmic 5′–3′ Xrn1 exonuclease[16]. These features suggest that yeast antisense regulatory ncRNAs belong to a separate class of unstable transcripts. Here, we address their systematic identification by characterizing the Xrn1-dependent cryptic transcriptome.

We analysed the whole transcriptome in wild-type and *xrn1Δ* strains (Fig. 1a). Because Xrn1 mainly acts downstream of RNA deadenylation[16], we analysed total RNA depleted for ribosomal RNAs (ribo$^-$) and subsequently compared with fractions enriched for polyadenylated RNAs (polyA$^+$) (Supplementary Fig. 2). Ultra-deep sequencing produced 84 and 134 millions of unique reads for ribo$^-$ preparations from wild-type and *xrn1Δ* strains, respectively (Supplementary Table 1), reaching unprecedented coverage of the yeast cryptic transcriptome[17–19]. Read densities were normalized with transfer RNAs and small nucleolar RNAs insensitive to Xrn1 activity[20,21] (Methods and Supplementary Fig. 3). Similar results were obtained with biological duplicates and with ABI-SOLiD or Illumina sequencing technologies (Supplementary Fig. 4a, b). Transcripts were identified along read density profiles and compared to annotated open reading frames (ORFs) and stable uncharacterized transcripts (SUTs)[22] (Methods and Supplementary Fig. 5). The Xrn1-deficient strain showed an average 4.4- to 3.9-fold increase of ORF and dubious ORF transcripts, respectively (Fig. 1b and Supplementary Fig. 6a, b), further establishing Xrn1 as the major exonuclease responsible for messenger RNA turnover[16]. The data revealed 932 previously uncharacterized ncRNAs (New) presenting a 10.8-fold average increase in the *xrn1Δ* strain (Fig. 1b, c and Supplementary Fig. 6c). Additionally, 75% (543) of the SUTs described in rich media[22] showed an unexpected 9.6-fold increase in the *xrn1Δ* strain, indicating that Xrn1 is also active on these ncRNAs (Fig. 1c and Supplementary Fig. 6d). In contrast to the majority of CUTs, 20% (183) of them were highly sensitive to Xrn1 (7.9-fold increase, Fig. 1c and Supplementary Fig. 6e), showing that a subclass of CUTs escapes from exosome-dependent nuclear degradation, as previously suggested[23]. Thus, we describe 1,658 ncRNAs that are Xrn1-sensitive unstable transcripts (XUTs), of which 66% are antisense to ORFs (Fig. 1b, c). The start sites of three XUTs upstream of the *SUC2* locus (*SUC2uxut*) (Fig. 1d) were identified by 5′ rapid amplification of cDNA ends (5′RACE) experiments (Supplementary Fig. 7). Northern blot analyses (Fig. 1e) showed that the levels of *SUC2uxut1* and *SUC2uxut3* are independent of Trf4, subunit of the TRAMP complex, required for CUT degradation[24]. This confirms that XUTs encompass a distinct class of cryptic transcripts. Further analyses showed that XUTs are polyadenylated and synthesized by RNAPII (Methods and Supplementary Fig. 8a–d).

We asked whether changes in growth conditions would activate XUTs in wild-type cells. Interestingly, lithium toxicity is due to inactivation of Xrn1, Rat1 and RNase MRP[25] and might be associated with XUT overexpression. The kinetics of appearance of selected XUTs in lithium-containing media (Fig. 2a and Supplementary Fig. 9) showed that the Ty1-ncRNA (*RTL*) appears before 30 min, similarly to *SUC2uxut1* and *TPO1axut* (transcriptome details in Supplementary Fig. 10a). After washing and further growth, cells present a dramatic decrease of XUTs levels, indicating a highly dynamic XUT accumulation. RNA-seq analyses after 80 min in lithium media showed XUTs overexpression (Fig. 2b) and 56% of them have similar RNA levels as the *xrn1Δ* strain (Fig. 2c). To note, some XUTs present a milder sensitivity to lithium as exemplified by *SUC2uxut2* and *SUC2uxut3* (Fig. 2a, d). These results indicate that a majority of XUTs rapidly and reversibly accumulate after lithium addition, indicating a role in the early response to this growth media change.

We tested whether antisense ncRNA accumulation controls sense gene transcription as for Ty1 (ref. 12). ChIP-seq analyses identified 996 ORFs with reduced RNAPII occupancy in the *xrn1Δ* strain (Fig. 3a, $P < 0.01$, Methods); among them, 273 (27%) accumulate antisense

[1]Centre de Génétique Moléculaire (CNRS UPR 3404), avenue de la Terrasse, 91198 Gif sur Yvette, France. [2]ncRNA, epigenetic and genome fluidity, Institut Curie, Centre de recherche, CNRS UMR3244, Université Pierre et Marie Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France. [3]NGS Platform, Institut Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France. [4]Recombination and Genome instability, Institut Curie, Centre de recherche, CNRS UMR3244, Université Pierre et Marie Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France.
*These authors contributed equally to this work.

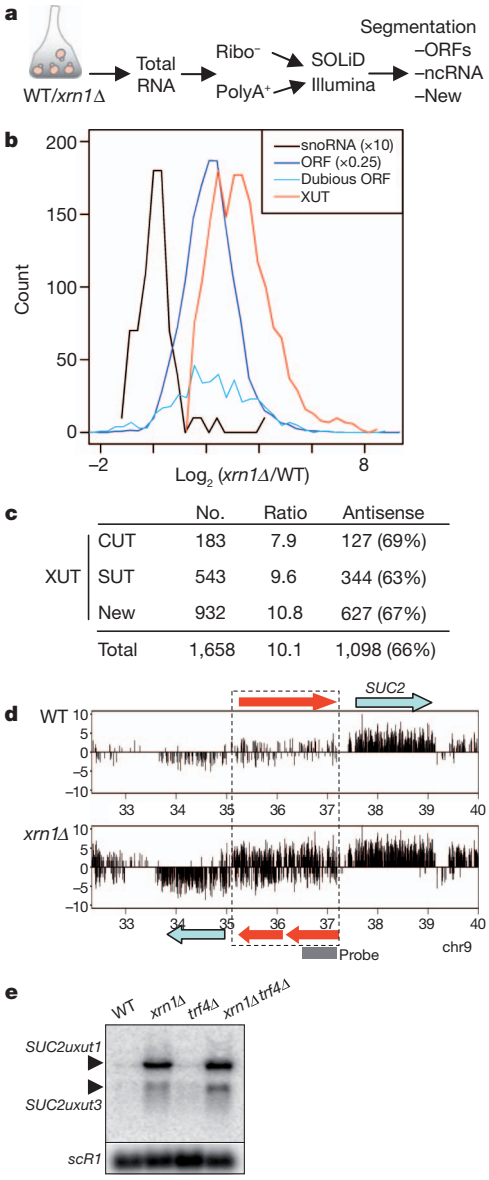


Figure 1 | Cytoplasmic 5′–3′ RNA decay controls a class of cryptic ncRNA. **a**, RNA-seq analyses using wild-type (WT) and *xrn1Δ* strains. Cells were grown in rich media and cDNA sequenced with SOLiD or Illumina technologies (Methods). Segmentation defines transcriptional units (ORFs, ncRNAs, unidentified transcripts (New)). **b**, XUTs define a novel class of ncRNAs. Tag density ratio distributions of *xrn1Δ*/WT for snoRNAs, ORFs and XUTs. **c**, Numbers of XUTs including new ncRNAs (New), most SUTs and few CUTs; average ratios of *xrn1Δ*/WT, numbers and percentage of XUTs antisense to ORFs. **d**, Transcriptome data for *SUC2* locus. Three XUTs (red) were identified within *SUC2* promoter in both orientations; bars, $\log_2$ read numbers (upward and downward, Watson and Crick strands, respectively); grey bar, probe used in **e**. **e**, Northern blot for indicated strains of sense *SUC2uxut1* (2.3 kb) and antisense *SUC2uxut3* (1 kb), using a double-stranded probe (**d**) and control *scR1*.

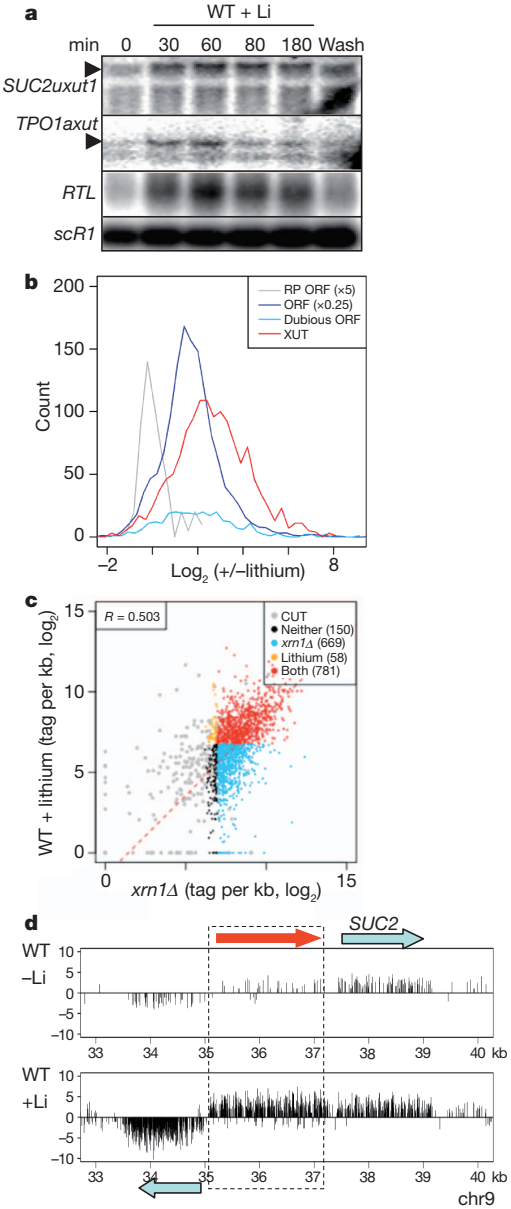


Figure 2 | XUT ncRNAs accumulate in wild-type cells grown in presence of lithium. **a**, XUT levels rapidly increase upon lithium treatment. Northern blot of total RNA from wild-type cells grown without and with 100 mM lithium at different times (Methods); strand-specific probes for *SUC2uxut1* and *SUC2uxut3* as in Fig. 1e, *TPO1axut* and *TY1* XUTs (*RTL*). **b**, XUTs are overexpressed in lithium media. Same as in Fig. 1b with wild type with or without lithium. Ribosomal protein genes were used as reference (grey, Methods). **c**, Tag density of XUTs and CUTs in *xrn1Δ* (abscissa) and WT + lithium (ordinate); 56% of XUTs are similarly expressed in *xrn1Δ* and WT grown in lithium (80 min). XUT expression compared to CUT expression (grey). In the subset of XUTs expressed at higher level than 75% of CUTs, 669 are found in *xrn1Δ* only (blue), 58 in lithium only (yellow) and 781 in both conditions (red). **d**, Same as in Fig. 1d with wild-type cells with or without lithium.

XUTs (Fig. 3a, class 1, Supplementary Table 2). This proportion is larger than for the ORFs that do not present significantly reduced RNAPII levels ($P < 10^{-15}$, chi-squared test; only 15% associate with antisense XUTs), indicating that sense transcriptional silencing correlates with antisense RNA accumulation. Although RNA stabilization in *xrn1Δ* strain could mask potential transcriptional repression, we identified 72 downregulated genes with antisense XUTs (Supplementary Fig. 12) among which 65 showed reduced RNAPII occupancy (Supplementary Fig. 13 and Supplementary Table 3). We conclude that, even though RNA-seq is more suitable to characterize XUT transcripts, ChIP-seq analyses identify extensively downregulated genes like *TIR1* and *TPO1* (Fig. 3b and Supplementary Fig. 10c, transcriptomes in Fig. 4b and Supplementary Fig. 10a). In these cases, RNAPII is almost undetectable within the XUTs' loci in the *xrn1Δ* strain, indicating that either antisense transcription is also silenced or XUT transcription is too low for detection. To establish that XUT accumulation and gene silencing are direct consequences of *XRN1* deletion, we performed *xrn1* conditional inactivation (*xrn1ts*[26], Methods). Not only do 70% of XUTs share identical expression in *xrn1Δ* and *xrn1ts* strains (Supplementary Fig. 14), but both strains also show similar RNAPII occupancies on a selection of class 1 genes (Figure 3c). This suggests

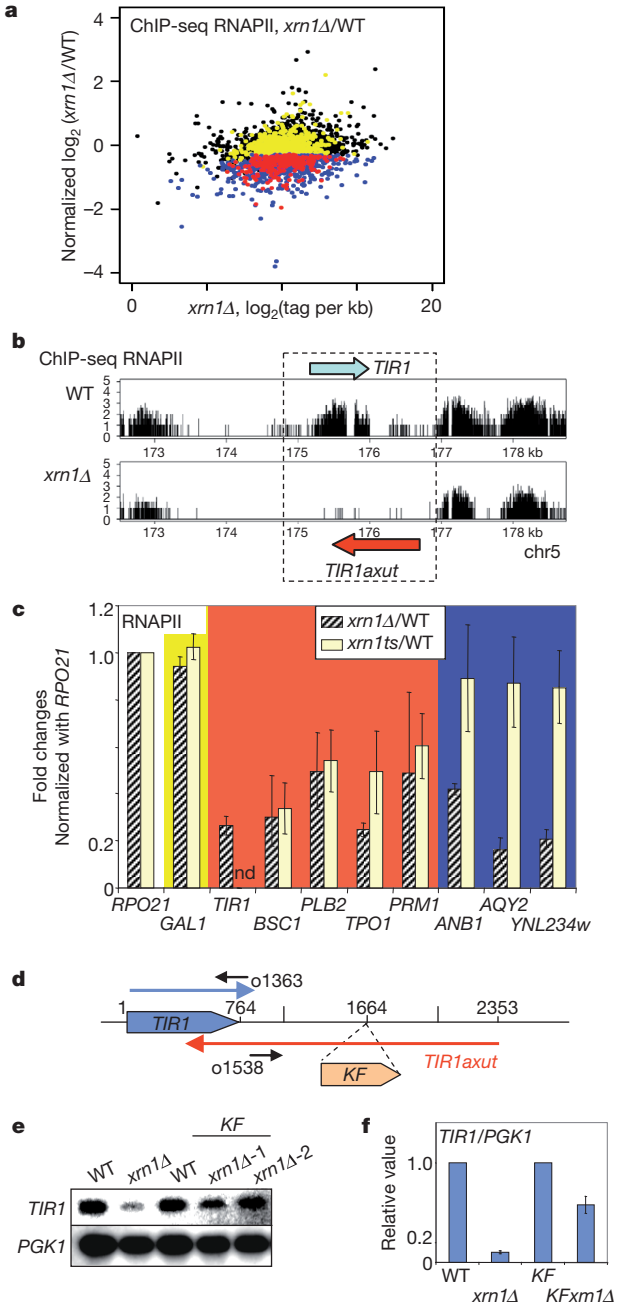


Figure 3 | Transcriptional gene silencing correlates with antisense XUT accumulation. **a**, ORFs with reduced tag densities in *xrn1Δ* strain compared to wild type by RNAPII ChIP-seq analysis ($P < 0.01$, Methods). Selected reduced ORFs associated with (class 1, red) or without antisense XUT (class 2, blue); ORFs with no significant variation with antisense XUT ($P > 0.01$, class 3, yellow) or without (class 4, black). **b**, ChIP-seq density profiles of *TIR1* locus in wild-type and *xrn1Δ* strains. **c**, RNAPII-ChIP analyses by quantitative PCR (mean ± s.d., $n = 3$) using primers within class 4: *RPO21*; class 3: *GAL1*; class 1: *TIR1*, *TPO1*, *BSC1*, *PLB2*, *PRM1* and class 2: *ANB1*, *AYQ2*, *YNL234w*, using wild-type, *xrn1Δ* and *xrn1ts* strains. *xrn1Δ*/*WT* and *xrn1ts*(37 °C)/WT(37 °C) RNAPII ratios normalized by *RPO21*. *xrn1ts* ChIP was performed after 180 min at 37 °C. *TIR1* expression was undetectable upon heat shock (nd). Colours as in **a**. **d**, *TIR1* locus and insertion of the *KANMX* cassette. Probes to detect sense and antisense RNAs are indicated. **e**, *TIR1* RNA levels are directly controlled by *TIR1axut*. Northern-blot with polyA$^+$ RNA using sense-specific probes to detect *TIR1* and *PGK1* mRNAs. Strains: WT, *xrn1Δ*, *KFWT* and *KFxrn1Δ* (two clones). **f**, Quantification of three independent Northern blots using *PGK1* for normalization (mean ± s.d., $n = 3$); respective wild-types were set to 1.

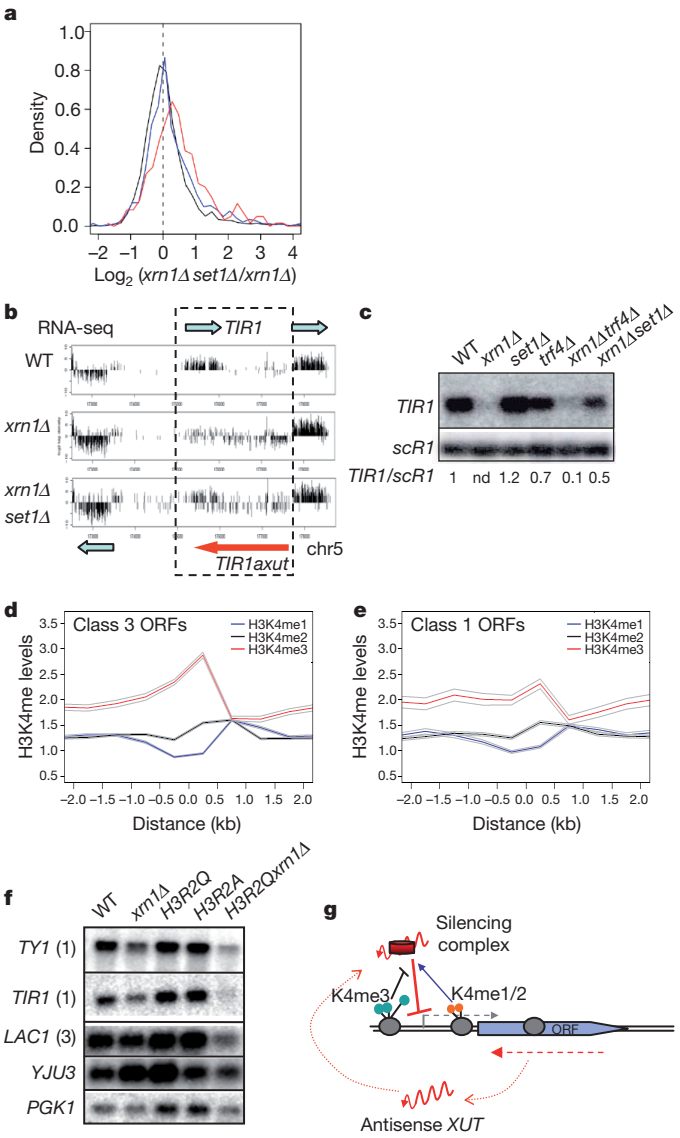


Figure 4 | Antisense XUTs mediate transcriptional gene silencing through Set1-dependent histone methylation. **a**, Repression of class 1 genes partly depends on Set1. Distributions of the $\log_2$ ratios of RNA-seq tag densities *xrn1set1Δ*/*xrn1Δ* for the ORFs; class 1, red; class 2, blue; others, black. **b**, *TIR1* locus transcriptomes in WT, *xrn1Δ* and *xrn1Δset1Δ* strains. Blue arrow, *TIR1* RNA; red arrow, *TIR1axut*; settings as in Fig. 1d. **c**, Northern blot validation of **b** with total RNAs. from indicated strains; probes as in Fig. 3f. **d, e**, Histone H3K4 methylation (mean ± s.e.m.) for genes with antisense XUTs in class 3 ($n = 820$) and class 1 ($n = 273$), respectively. Data from wild type retrieved from previous publication[28] were plotted for H3K4me1 (blue), H3K4me2 (black) and H3K4me3 (red); abscissa, distance to start codon (kb). **f**, Northern blot of polyA$^+$ RNA revealing genes with antisense XUTs: *TIR1*, *TY1* (class 1), *LAC1* (class 3), or without: *YJU3*, *PGK1* (class 4) for indicated strains. Note that *TIR1* and *TY1* mRNAs are less sensitive to *xrn1Δ* due to background differences. **g**, Model depicting XUT-mediated repressive transcriptional activity enhanced by H3K4me1/2 (orange) and antagonized by H3K4me3 (blue).

that class 1 genes are controlled by Xrn1, potentially through their respective antisense ncRNAs. Interruption of the *TIR1axut* in the *xrn1Δ* strain (Fig. 3d and Supplementary Fig. 15) re-established 70% of the *TIR1* mRNA level (Fig. 3e, f). Residual *TIR1axut* detection, due to either alternative initiation start sites of antisense ncRNAs or read-through over the inserted cassette, might explain the residual repressive activity on *TIR1*. Altogether, these data strongly support a direct role for antisense XUTs in gene silencing.

The ncRNA-mediated *TY1* silencing requires the Set1 activity[12] which represses basal transcription of many genes associated with antisense

transcripts[27]. RNA-seq analyses performed in a *set1Δxrn1Δ* strain revealed that only the class 1 genes have higher RNA levels in the absence of Set1 (Fig. 4a; $P < 10^{-6}$, one-tailed two-sample Kolmogorov–Smirnov test), indicating that Set1 has a widespread role in antisense ncRNA-mediated gene repression. Remarkably, Set1 is not involved in XUT expression because the *set1Δxrn1Δ* and *xrn1Δ* strains showed similar levels of XUTs (Supplementary Fig. 16, $P = 0.2$). Furthermore, *TIR1* silencing is partly mediated by Set1-dependent histone H3 lysine 4 methylation (H3K4me) because disrupting the Set1 catalytic activity re-established high levels of *TIR1* RNA as shown in *xrn1Δset1Δ*, *xrn1Δset1G951S* and *xrn1ΔH3K4A* strains (Fig. 4b, c and Supplementary Fig. 17a, b). To dissect the Set1 role, we compared the H3K4me1, H3K4me2 and H3K4me3 levels[28] in the wild type over the genes with antisense XUTs. Class 3 genes show H3K4me1/2/3 patterns (Fig. 4d) equivalent to all ORFs (Supplementary Fig. 18). By contrast, class 1 genes present 60% reduction of H3K4me3 levels but normal H3K4me2/me1 patterns (Fig. 4e). This indicates that although they have similar RNAPII distribution as class 3 genes (Supplementary Fig. 19), class 1 genes are marked and/or sensitized to their respective antisense ncRNA by low H3K4me3 levels. A tantalizing hypothesis is that antisense ncRNA-mediated gene silencing requires H3K4me1 and/or me2 but is antagonized by H3K4me3. Our model predicts that class 1 but also class 3 genes would be further silenced in absence of H3K4me3. To test this possibility, we performed analyses in a strain specifically depleted for H3K4me3 (H3R2Q mutant), without disrupting the Set1 complex[29]. In addition to a synergetic repression of *TY1* and *TIR1* mRNA (class 1 genes), we observed a decrease of the class 3 *LAC1* RNA levels in the *H3R2Qxrn1Δ* strain (Fig. 4f and Supplementary Fig. 20a). H3K4me3 is not involved in the expression of the respective antisense XUTs, similarly expressed in the *xrn1Δ* and *H3R2Qxrn1Δ* strains (Supplementary Fig. 20b). By contrast the *PGK1* and *YJU3* genes (class 4), both devoid of antisense XUTs, were not affected in *H3R2Qxrn1Δ*, supporting a model in which H3K4me3 antagonizes the negative activity of antisense ncRNA potentially associated with unknown silencing factors (Fig. 4g).

In conclusion, our results provide evidence for a novel class of cryptic antisense transcripts degraded by the cytoplasmic 5′–3′ RNA decay pathway. We propose that Set1-dependent histone methylation has a key role in controlling the antisense ncRNA-repressive activity using H3K4me3 as a molecular switch. Given the broad class of XUTs, we already anticipate also a post-transcriptional role(s). Because the key components of this pathway are distributed among the eukaryotic kingdom, the mechanism is likely to be conserved throughout evolution.

METHODS SUMMARY

RNA-seq libraries were generated according to the manufacturer's instructions using ribo$^-$/polyA$^+$ RNAs extracted from *S. cerevisiae* strains. ChIP-seq was performed according to Illumina protocols and sequenced on GAIIx. Tag profiles were analysed using appropriated biocomputational and statistical approaches (Methods). Northern-blot and ChIP (using anti-Rpb1-CTD 8wg16 antibody) were performed using standard techniques (Methods). Sequence data are publicly available at NCBI Sequence Read Archive under accession number SRA030505 and at http://vm-gb.curie.fr/XUT/index.htm.

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

Received 31 December 2010; accepted 14 April 2011.
Published online 22 June 2011.

1. Bernstein, E. & Allis, C. D. RNA meets chromatin. *Genes Dev.* **19,** 1635–1655 (2005).
2. Moazed, D. Small RNAs in transcriptional gene silencing and genome defence. *Nature* **457,** 413–420 (2009).
3. Swiezewski, S., Liu, F., Magusin, A. & Dean, C. Cold-induced silencing by long antisense transcripts of an *Arabidopsis* Polycomb target. *Nature* **462,** 799–802 (2009).
4. Yu, W. *et al.* Epigenetic silencing of tumour suppressor gene *p15* by its antisense RNA. *Nature* **451,** 202–206 (2008).
5. Yap, K. L. *et al.* Molecular interplay of the noncoding RNA *ANRIL* and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of *INK4a*. *Mol. Cell* **38,** 662–674 (2010).
6. Huarte, M. & Rinn, J. L. Large non-coding RNAs: missing links in cancer? *Hum. Mol. Genet.* **19,** R152–R161 (2010).
7. Chow, J. & Heard, E. X inactivation and the complexities of silencing a sex chromosome. *Curr. Opin. Cell Biol.* **21,** 359–366 (2009).
8. Nagano, T. *et al.* The *Air* noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* **322,** 1717–1720 (2008).
9. Amaral, P. P., Dinger, M. E., Mercer, T. R. & Mattick, J. S. The eukaryotic genome as an RNA machine. *Science* **319,** 1787–1789 (2008).
10. Berretta, J. & Morillon, A. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO Rep.* **10,** 973–982 (2009).
11. Camblong, J. *et al. Trans*-acting antisense RNAs mediate transcriptional gene cosuppression in *S. cerevisiae*. *Genes Dev.* **23,** 1534–1545 (2009).
12. Berretta, J., Pinskaya, M. & Morillon, A. A cryptic unstable transcript mediates transcriptional *trans*-silencing of the Ty1 retrotransposon in *S. cerevisiae*. *Genes Dev.* **22,** 615–626 (2008).
13. Matsuda, E. & Garfinkel, D. J. Posttranslational interference of Ty1 retrotransposition by antisense RNAs. *Proc. Natl Acad. Sci. USA* **106,** 15657–15662 (2009).
14. Aravind, L., Watanabe, H., Lipman, D. J. & Koonin, E. V. Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proc. Natl Acad. Sci. USA* **97,** 11319–11324 (2000).
15. Jacquier, A. The complex eukaryotic transcriptome: unexpected pervasive transcription and novel small RNAs. *Nature Rev. Genet.* **10,** 833–844 (2009).
16. Long, R. M. & McNally, M. T. mRNA decay: X (*XRN1*) marks the spot. *Mol. Cell* **11,** 1126–1128 (2003).
17. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320,** 1344–1349 (2008).
18. Neil, H. *et al.* Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* **457,** 1038–1042 (2009).
19. Yassour, M. *et al.* Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome Biol.* **11,** R87 (2010).
20. Chernyakov, I., Whipple, J. M., Kotelawala, L., Grayhack, E. J. & Phizicky, E. M. Degradation of several hypomodified mature tRNA species in *Saccharomyces cerevisiae* is mediated by Met22 and the 5′–3′ exonucleases Rat1 and Xrn1. *Genes Dev.* **22,** 1369–1380 (2008).
21. Fatica, A., Morlando, M. & Bozzoni, I. Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3′-processing apparatus. *EMBO J.* **19,** 6218–6229 (2000).
22. Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457,** 1033–1037 (2009).
23. Thompson, D. M. & Parker, R. Cytoplasmic decay of intergenic transcripts in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **27,** 92–101 (2007).
24. Wyers, F. *et al.* Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* **121,** 725–737 (2005).
25. Dichtl, B., Stevens, A. & Tollervey, D. Lithium toxicity in yeast is due to the inhibition of RNA processing enzymes. *EMBO J.* **16,** 7184–7195 (1997).
26. Johnson, A. W. Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. *Mol. Cell. Biol.* **17,** 6122–6130 (1997).
27. Pinskaya, M. & Morillon, A. Histone H3 lysine 4 di-methylation: a novel mark for transcriptional fidelity? *Epigenetics* **4,** 302–306 (2009).
28. Pokholok, D. K. *et al.* Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* **122,** 517–527 (2005).
29. Kirmizis, A. *et al.* Arginine methylation at histone H3R2 controls deposition of H3K4 trimethylation. *Nature* **449,** 928–932 (2007).

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

Acknowledgements We thank B. Séraphin, L. Bénard and J. O'Sullivan for support and advice; B. Dichtl for insights into lithium treatment data normalization; M. Wéry and A. Taddei for helpful discussions; A. Johnson, T. Kouzarides and V. Géli for generous gift of plasmids and strains. Special thanks to M. Descrimes, C. Jubin and S. Lair for technical assistance. We thank L. Steinmetz and M. Chodder for sharing unpublished results. E.L.V.D. benefits from an FRM fellowship. This work has benefited from facilities and expertise of the IMAGIF sequencing platform (Centre de Recherche de Gif). This work was financially supported by the Canceropole Ile de France, the ANR "REGULncRNA" and ERC "EPIncRNA" starting grant.

Author Contributions E.L.V.D. performed molecular biology experiments, RNA-seq, ChIP-seq libraries and sequencing on the ILLUMINA platform. C.L.C., Y.D.-C., M.S. and C.T. performed statistical and bioinformatic analyses. S.G., M.K., V.R. and C.B. provided technical assistance to molecular biology experiments. A.M., C.T., E.L.V.D. and C.L.C. designed the experiments. P.L.-N. and S.L. performed RNA-seq libraries and NGS sequencing on the SOLiD platform; A.N. managed sequencing on the SOLiD platform. E.L.V.D., C.L.C. and A.N. contributed to the writing. C.T. and A.M. wrote the paper. C.T. and A.M. planned the project.

Author Information Sequence data are publicly available at NCBI Sequence Read Archive under accession number SRA030505 and at http://vm-gb.curie.fr/XUT/index.htm. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.M. (antonin.morillon@curie.fr) or C.T. (thermes@cgm.cnrs-gif.fr).

METHODS

Yeast strains and growth media. Strains used in this study are from the Euroscarf collection (S288C, BY4741 background), except for the RNAPII and histone H3 studies. For RNAPII, we used the strains *rpb1-1* and *rpb1-1xrn1Δ* (ref. 12). For histone H3 we used strains WT-H3, *H3K4A*, *xrn1Δ* and *H3K4Axrn1Δ* previously manipulated[12] and WT-H3, *H3R2A* and *H3R2Q* generously provided by T. Kouzarides[29] in which we deleted *XRN1* gene (this work). The list of BY strains is WT, *xrn1Δ*, *set1Δ*, *set1Δxrn1Δ*, *trf4Δ*, *xrn1Δtrf4Δ*, *set1G951S* and *set1G951Sxrn1Δ*. The *set1G951S* strain was kindly provided by V. Géli. Gene deletions were introduced by transformation of PCR fragments generated with specific primers (sequences can be obtained upon request) using the appropriate plasmids[30].

Growth media were prepared by standards methods using rich YPDA media (yeast peptone dextrose adenine, Gibco) or minimal CSM media (MP biochemical) containing 2% glucose.

For lithium induction, cells were grown in CSM media and then transferred during the indicated time in CSM media with 100 mM lithium as indicated[25].

For Xrn1 conditional mutant experiments, the *xrn1Δ* strain was transformed with a wild-type *XRN1* plasmid (PAM27, pAJ52) or *xrn1ts* plasmid (PAM143, pAJ53), both plasmids from a gift from A. Johnson. Transformed cells were grown in CSM-URA media with glucose at 30 °C. Heat shock was performed at 37 °C and cells collected at the indicated times.

RNA extraction, polyA$^+$ and ribo$^-$ RNA purifications. Total RNA was extracted using the hot phenol extraction procedure. PolyA$^+$ RNAs were purified on oligodT Dynabeads (Invitrogen) and ribosomal RNA were depleted using the RiboMinus kit (Invitrogen). RNA quality was checked on agarose gel (Supplementary Fig. 2) and with the Bioanalyzer (Agilent) ensuring the quantification of RNA peaks recommended by the manufacturer.

Northern blotting. RNAs were loaded on denaturing 1% agarose gels containing formaldehyde and transferred to nitrocellulose membranes (Hybond XL). Membranes were cross-linked by ultraviolet irradiation and hybridized over-night at 65 °C with either ^{32}P-labelled DNA probes or oligonucleotide probes in PERFECT-HYB PLUS buffer (Sigma). Blots were washed at 65 °C for 10 min once with 2× SSC, 1% SDS and twice with 0.1× SSC, 1% SDS. DNA probes were obtained by random primed labelling (Stratagene) of specific DNA fragments generated by PCR. PCR primers are available upon request. *PGK1* and *scR1* RNAs were used as loading controls. Northern signals were quantified using ImageQuant software, normalized with *scR1* or *PGK1* intensities and wild-type ratios were set arbitrarily to 1. Error bars correspond to the standard deviation over three independent cultures.

Reverse transcription. Reverse transcription was performed according to the manufacturer's instructions (Invitrogen, SuperScript) on total RNA. Specific primers were used to amplify *TIR1axut*, *LAC1axut* and *scR1* loading control. Quantitative PCR were performed with the LightCycler 480 (Roche) using SYBR Green. PCR were normalized with *scR1* signals and wild-type ratios adjust to 1 for comparison. Error bars correspond to standard deviation over three independent reverse transcription reactions.

5′ RACE analysis. 5′ RACE experiments were done using a Firstchoice RLM-RACE kit from Ambion following the manufacturer's instructions. Total RNA from wild-type or *xrn1Δ* cells was used. Reverse transcription was performed with primers A, B, and C to detect the 5′ ends of *SUC2uxut* (Supplementary Fig. 7). Subsequently, PCR was done using nested primers (sequences available upon request). PCR products were cloned and sequenced.

Chromatin immunoprecipitation. Chromatin immunoprecipitations were performed essentially as described previously[12]. Yeast strains were grown to $D_{600} = 0.5$ in YPDA at 30 °C, and cross-linked 20 min by the addition of formaldehyde to a final concentration of 1.2%. Crosslinked reaction was quenched by adding glycine at 0.5 M final concentration. Chromatin was sonicated to obtain 400–500 nucleotides DNA fragment and 200 µg of sonicated chromatin was immunoprecipitated for 3 h at 21 °C on Pan mouse Dynabeads (Invitrogen) coated with specific antibody against the carboxy-terminal domain of Rpb1 (8WG16, Millipore). All immunoprecipitations were repeated at least three times with different chromatin extracts. Immunoprecipitated DNA was quantified by real-time PCR using the LightCycler 480 (Roche) with primer pairs (sequences available upon request). Signals are expressed as percentage of input DNA relatively to *RPO21* (coordinate: chromosome 4: 210562 to 205361). Error bars correspond to standard deviations of three independent experiments.

Library preparation for RNA-seq. The extracted RNAs (polyA$^+$ or ribo$^-$) were submitted to RNA fragmentation as recommended in the ABI whole transcriptome library preparation kit. RNA adaptors were ligated before cDNA preparation to ensure strand recognition. Typically, DNAs ranging from 150 to 250 nucleotides were size-selected on polyacrylamide gel followed by PCR amplification (15 cycles). Denatured single-stranded DNA was then hybridized on beads before emulsion PCR and deposited on SOLiD slides for sequencing. Runs were performed on the SOLiD V3 sequencing machine and 50 nucleotides single reads were generated on colour code format.

For Illumina sequencing, RNAs were subjected to sequencing on the Genome Analyzer IIx sequencing machine and 38 nucleotides reads were generated. The same polyA$^+$ samples as for the SOLiD libraries (but with Illumina adapters added by PCR) were used to confirm that sequencing technologies do not introduce bias in the reads. The duplicates libraries for wild type and *xrn1Δ*, lithium and *ts* were sequenced with Illumina technology (Supplementary Table 1).

Library preparation for ChIP-seq. For ChIP-seq, chromatin immunoprecipitation was done as described above, the only modification being that reactions were scaled up to obtain sufficient amounts of material for library preparation. Instead of 200 µg, 800 µg of sonicated chromatin was used. The sequences were obtained on Genome Analyzer IIx sequencer with 38 nucleotides reads (Supplementary Table 1).

Sequencing data. For SOLiD sequencing, 50-nucleotide sequence reads were identified using the standard SOLiD base-calling software and then aligned to the reference genome (*Saccharomyces cerevisiae* S288c retrieved from SGD, http://www.yeastgenome.org/) using Mapreads (v2.4.1) software allowing up to six mismatches. For Illumina/Solexa sequencing, 38-nucleotide sequence reads were identified using the standard Illumina base-calling software and then aligned to the reference genome using the ELAND (CASAVA pipeline) software allowing up to two mismatches within the 32-nucleotide seed. For multiplex samples, reads were partitioned and those lacking an intact barcode or index were discarded. Replicate experiments for ribo$^-$ samples produced read density values that strongly correlated to each other (Supplementary Fig. 4a, Pearson $R = 0.96$, $P < 10^{-15}$). Sequencing data obtained with the polyA$^+$/WT RNA sample using either the SOLiD or the Illumina/Solexa device were highly correlated to each other (Supplementary Fig. 4b, $R = 0.98$, $P < 10^{-15}$).

Transcriptome analysis and normalization. Genome was annotated according to SGD (http://www.yeastgenome.org/, 5 January 2010); SUT (stable unannotated transcript) and CUT (cryptic unstable transcript) annotations were retrieved[22]. Tag densities were computed for each transcript in each sample and in all cases, except for tRNAs, only tags mapping to unique positions were considered. Tag densities for tRNAs were computed (for each anticodon) by using the tags mapping to multiple positions. Except for the lithium and *set1Δ* transcriptomes, the tag densities were normalized in such a way that snoRNAs and tRNAs present the same levels in the wild-type and *xrn1Δ* strains (Supplementary Figure 3). The candidate transcripts obtained by the segmentation process that presented an *xrn1Δ*/WT ratio significantly larger than the background ratio were retained as XUTs. To determine whether the expression level of a gene (ORF) was significantly different in the *xrn1Δ* and wild-type strains, a *P*-value was computed based on negative binomial distribution (edgeR package[31] from Bioconductor, http://www.bioconductor.org/); for this comparison, individual ORF expression levels in the *xrn1Δ* and wild-type strains were compared to the ratio of the corresponding mean ORF expression values; transcripts with a *P*-value < 0.01 were retained. In the wild-type strain, a fraction of XUTs presents expression levels close to the background level of the density profile; to compare XUT expression levels in the ribo$^-$ and polyA$^+$ samples, only those presenting levels 50% higher than the background level of the corresponding profile were retained.

An extensive study was performed to identify the best way to normalize the RNA-seq values in the presence of lithium. As suggested by previous studies[25], tRNAs and snoRNAs levels are affected by lithium addition with mean ratios of 1.8 and 2.1, respectively (Supplementary Fig. 11a–d). In contrast, ribosomal proteins genes-derived transcripts present only a slight increase (1.5 fold) in the *xrn1Δ* strain, and are insensitive to lithium addition (mean ratio of 1.1, Supplementary Fig. 11e, f). The mean ribosomal proteins genes transcript value was thus used as reference to normalize RNA-seq values in the presence of lithium.

For the *set1Δ* transcriptome, levels were normalized by the mean level of all ORFs taken as reference, because the Set1 complex controls snoRNA transcription[32].

Segmentation of massive sequencing data. The RNA-seq procedure generates considerable variation of tag coverage along genes complicating the segmentation process. To address this problem, we designed a crude heuristic to detect candidate transcripts. Segmentation was performed on massive sequencing data obtained from the ribo$^-$ samples extracted from the *xrn1Δ* strain. The tag profile was built by computing the tag density within a sliding 120 base pairs (bp) window, using a 1-bp step size; segments were delimited on the tag profile using a threshold of 0.06 tag per nucleotide (50% of the annotated CUTs present a tag density lower than this threshold). To define transcripts, the extremities of each segment were adjusted to the first and last internal tags. Segments overlapping genes annotated on the same strand (ORFs, stable RNAs) were then removed. The remaining consecutive segments of similar densities (ratio smaller than twofold) distant by less than 250 bp were joined (only the 5′ end of the RNA pieces obtained by fragmentation of the initial transcripts was sequenced); the resulting segments longer than 250 bp were retained as final transcripts corresponding to 543

SUTs, 183 CUTs and 932 new transcripts (New). The extremities of XUTs tested experimentally were manually repositioned according to experimental data. With this procedure, previously defined SUTs were covered at 87% by the detected segments and the segments intersecting the SUTs were covered at 80% by the SUTs; in addition, 90% of the SUTs were covered by a unique detected segment (Supplementary Fig. **5**).

XUTs are polyadenylated and RNAPII dependent. Among the 1,658 XUTs, 683 were detected at low levels in the wild-type ribo$^-$ sample. Interestingly, 493 (72%) were also detected within the polyA$^+$ fraction (Supplementary Fig. 8a, b), strongly supporting that XUTs are polyadenylated, as shown for *PHO84*-, *GAL10*- and *TY1*-associated ncRNAs[12,33,34]. To confirm that XUTs are synthesized by RNAPII, we monitored the steady-state levels of the *SUC2uxut1* and *SUC2uxut3* mRNAs upon depletion of Rpb1, the main subunit of RNAPII (ref. 35). The *SUC2uxut1* mRNA levels were significantly reduced after 60 min of RNAPII inactivation similarly to the class 2 *ACT1* mRNA (Supplementary Fig. 8c, d). In contrast, the levels of *scR1* and rRNA, transcribed by RNA polymerase III and I, respectively, remained constant. Altogether these data support that XUTs RNA are polyadenylated and synthesized by the RNA polymerase II.

Analysis of *xrn1* conditional mutant. To examine possible late secondary effects of *XRN1* deletion, we asked whether a conditional mutant of Xrn1 would rapidly accumulate XUTs after Xrn1 inactivation and show repression of the same group of genes defined in Fig. 3a (class 1) . To answer this question, we determined the kinetics of the ncRNA-mediated Ty1 regulation using an *xrn1* thermosensitive (ts) mutant[26]. Interestingly, *TY1* ncRNA accumulated around 30 min after heat shock (Supplementary Figure 14a). In addition, *TY1* ncRNA accumulation preceded the decrease of the *TY1* mRNA starting approximately at 60 min, confirming the causal link between XUT accumulation and gene repression[12,13]. Interestingly, the *TY1* ncRNA kinetics reflect other XUTs dynamic because *SUC2uxut1*, *SUC2uxut3* and *TPO1axut* also appeared at 30 min, indicating a rapid and general accumulation of the XUTs upon Xrn1 inactivation. Indeed, RNA-seq experiments showed that 75% of the XUTs detected in *xrn1Δ* strain accumulated similarly after 90 min of Xrn1 inactivation (Supplementary Fig. 14b–d), indicating that XUTs are primary targets of Xrn1 RNA decay pathway. To determine whether the regulation of class 1 genes is a late or early event, we performed RNAPII-ChIP experiments in the *xrn1Δ* and *xrn1ts* strains (after 180 min in 37 °C) on a selection of genes belonging to the different classes defined in Figure 3a. Strikingly, the class 1 genes showed a similar decrease of RNAPII occupancy in the two *XRN1*-defective strains (Figure 3c), indicating that Xrn1 inactivation has an early effect on RNAPII on those genes. By contrast, the class 2 genes are only suppressed in *xrn1Δ* strain supporting a late and likely indirect effect of Xrn1 inactivation on those genes. We conclude that the class 1 genes are directly controlled by Xrn1, potentially though their respective antisense ncRNAs.

30. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14,** 953–961 (1998).
31. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression. *Bioinformatics* **26,** 139–140 (2010).
32. Dichtl, B., Aasland, R. & Keller, W. Functions for *S. cerevisiae* Swd2p in 3′ end formation of specific mRNAs and snoRNAs and global histone 3 lysine 4 methylation. *RNA* **10,** 965–977 (2004).
33. Pinskaya, M., Gourvennec, S. & Morillon, A. H3 lysine 4 di- and tri-methylation deposited by cryptic transcription attenuates promoter activation. *EMBO J.* **28,** 1697–1707 (2009).
34. Camblong, J., Iglesias, N., Fickentscher, C., Dieppois, G. & Stutz, F. Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in *S. cerevisiae*. *Cell* **131,** 706–717 (2007).
35. Nonet, M., Scafe, C., Sexton, J. & Young, R. Eucaryotic RNA polymerase conditional mutant that rapidly ceases mRNA synthesis. *Mol. Cell. Biol.* **7,** 1602–1611 (1987).

LETTER

doi:10.1038/nature10134

Intravenous gammaglobulin suppresses inflammation through a novel T_H2 pathway

Robert M. Anthony[1]*, Toshihiko Kobayashi[1]*, Fredrik Wermeling[1] & Jeffrey V. Ravetch[1]

High-dose intravenous immunoglobulin is a widely used therapeutic preparation of highly purified immunoglobulin G (IgG) antibodies. It is administered at high doses (1–2 grams per kilogram) for the suppression of autoantibody-triggered inflammation in a variety of clinical settings[1]. This anti-inflammatory activity of intravenous immunoglobulin is triggered by a minor population of IgG crystallizable fragments (Fcs), with glycans terminating in α2,6 sialic acids (sFc) that target myeloid regulatory cells expressing the lectin dendritic-cell-specific ICAM-3 grabbing non-integrin (DC-SIGN; also known as CD209)[2–4]. Here, to characterize this response in detail, we generated humanized DC-SIGN mice (hDC-SIGN), and demonstrate that the anti-inflammatory activity of intravenous immunoglobulin can be recapitulated by the transfer of bone-marrow-derived sFc-treated hDC-SIGN$^+$ macrophages or dendritic cells into naive recipients. Furthermore, sFc administration results in the production of IL-33, which, in turn, induces expansion of IL-4-producing basophils that promote increased expression of the inhibitory Fc receptor FcγRIIB on effector macrophages. Systemic administration of the T_H2 cytokines IL-33 or IL-4 upregulates FcγRIIB on macrophages, and suppresses serum-induced arthritis. Consistent with these results, transfer of IL-33-treated basophils suppressed induced arthritic inflammation. This novel DC-SIGN–T_H2 pathway initiated by an endogenous ligand, sFc, provides an intrinsic mechanism for maintaining immune homeostasis that could be manipulated to provide therapeutic benefit in autoimmune diseases.

Binding of intravenous immunoglobulin (IVIG) or sFc to specific ICAM-3 grabbing non-integrin-related 1 (SIGN-R1) on splenic marginal zone macrophages suppresses autoantibody-mediated inflammation[4]. Although the human orthologue of SIGN-R1, DC-SIGN, showed similar binding specificity for sFc as mouse SIGN-R1, its expression pattern is broader, as it is detected systemically on myeloid-derived cells, including dendritic cells, macrophages and some monocytes[5,6]. DC-SIGN recognizes high-mannose glycans from a variety of pathogens, and acts as a pattern recognition receptor bridging innate and adaptive immunity[7]. Ligation of DC-SIGN by bacteria-derived mannosylated glycans can induce their internalization, and also synergize with other innate receptor pathways promoting inflammation and resistance to infection. In contrast, binding of sFc to DC-SIGN requires both carbohydrate and protein determinants, and results in an anti-inflammatory response[2,4]. The immunosuppressive potential of DC-SIGN has been documented following ligation by HIV-derived gp120 or anti-DC-SIGN antibody, which promotes the development of tolerogenic, IL-10-producing dendritic cells, and interferes with Toll-like receptor (TLR) signalling[8,9].

To study human DC-SIGN in the context of IVIG anti-inflammatory activity, we expressed hDC-SIGN—driven by its endogenous promoter to reproduce the characteristically broad *in vivo* expression pattern of hDC-SIGN—in a mouse. Human bacterial artificial chromosome (BAC) clones encoding the *DC-SIGN* gene and its regulatory regions were introduced as a transgene into mice (Fig. 1a). Transgenic mice showed surface expression of this human lectin on dendritic cells, macrophages, and monocytes, in the peripheral blood, bone marrow and spleen, resembling the human expression pattern of DC-SIGN (Supplementary Fig. 2a–c), although a higher percentage of murine monocytes were found to express DC-SIGN.

To determine if hDC-SIGN could substitute for SIGN-R1 in mediating IVIG protection, hDC-SIGN$^+$ mice were crossed to SIGN-R1-deficient animals (hDC-SIGN$^+$/SIGN-R1$^{-/-}$) and challenged with arthritogenic K/BxN serum[10]. Both induction of arthritis and responsiveness to IVIG and sFc were similar in wild-type mice and hDC-SIGN$^+$/SIGN-R1$^{-/-}$ mice (Fig. 1b and Supplementary Fig. 3a). In contrast, induced arthritis was not suppressed by IVIG or sFc in

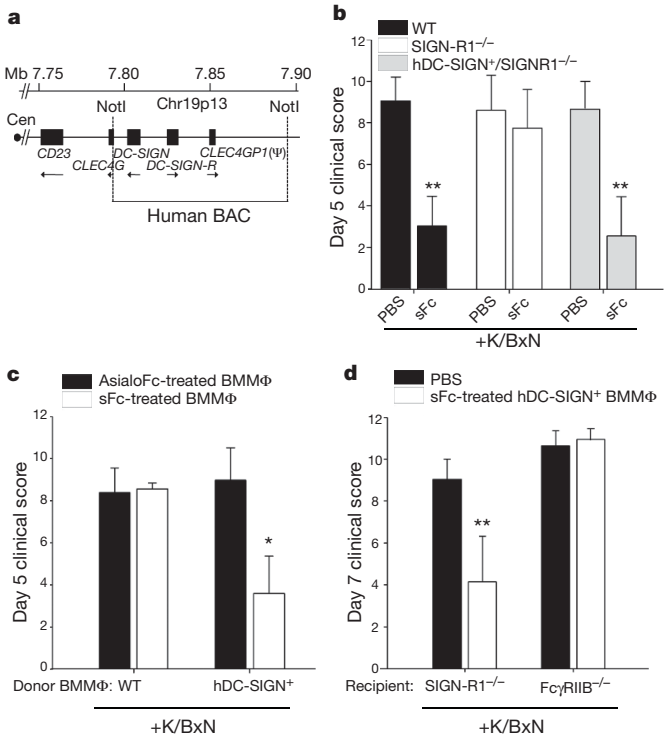


Figure 1 | Human DC-SIGN conveys sFc anti-inflammatory activity. **a**, A map of the human chromosome 19 BAC clone containing DC-SIGN and DC-SIGN-R genes. Cen, centromere. **b**, Wild-type (WT; black bars), SIGN-R1$^{-/-}$ (white bars), or hDC-SIGN$^+$/SIGN-R1$^{-/-}$ (grey bars) mice were administered K/BxN sera and sFc. $^{**}P < 0.001$ determined by a Fisher least significant difference (LSD) post-hoc test. **c**, AsialoFc- (black bars) or sFc- (white bars) treated wild-type and hDC-SIGN$^+$ bone-marrow derived macrophages (BMMΦ) were administered to K/BxN-sera-treated wild-type recipients. $^{*}P < 0.05$ determined by Tukey's post-hoc test. **d**, PBS (black bars) or sFc-treated hDC-SIGN$^+$ bone-marrow-derived macrophages (white bars) were administered to SIGN-R1$^{-/-}$ and FcγRIIB$^{-/-}$ recipients. Means and standard deviations are plotted; $^{**}P < 0.001$ determined by Tukey's post-hoc test.

[1]Laboratory of Molecular Genetics and Immunology, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA.
*These authors contributed equally to this work.

SIGN-R1$^{-/-}$ mice. Thus, hDC-SIGN expression was sufficient to trigger the IVIG and sFc anti-inflammatory response.

A related lectin, DC-SIGN-R, is linked to DC-SIGN on the BAC transgene (Fig. 1a). hDC-SIGN-R has reduced affinity to sFc as compared to hDC-SIGN (Supplementary Fig. 3b). To define the contribution of DC-SIGN-R to sFc anti-inflammatory activity, mice that express hDC-SIGN alone as a transgene[11] were crossed with SIGN-R1$^{-/-}$ mice (CD11c-DC-SIGN/SIGN-R1$^{-/-}$). These mice were protected from inflammatory arthritis by IVIG (Supplementary Fig. 3c). Further, selective blockade of hDC-SIGN in transgenic hDC-SIGN^{+}/SIGN-R1$^{-/-}$ mice expressing both hDC-SIGN and hDC-SIGN-R resulted in a loss of IVIG protection *in vivo* (Supplementary Fig. 3d). These results support a requirement for hDC-SIGN but not hDC-SIGN-R in this anti-inflammatory response triggered by sFc.

Next, we sought to determine if stimulation of hDC-SIGN^{+} cells matured from bone marrow with sFc was sufficient to induce an anti-inflammatory response. Bone-marrow-derived macrophages and dendritic cells cultured from hDC-SIGN^{+} transgenic animals expressed hDC-SIGN, but not hDC-SIGN-R or SIGN-R1 (Supplementary Fig. 4a, b, c). Bone-marrow-derived cells cultured from hDC-SIGN^{+} transgenic or wild-type mice were pulsed for 30 min with sFc or asialylated Fcs (asialoFc) at a concentration representative of *in vivo* treatments. The treated cells were collected, washed and administered to wild-type mice, which were then challenged with K/BxN serum (Supplementary Fig. 4d). Mice receiving hDC-SIGN^{+} bone-marrow-derived macrophages or dendritic cells pulsed with sFc or IVIG showed reduced joint inflammation as compared to recipient mice that received wild-type cells, or hDC-SIGN^{+} cells pulsed with asialoFc (Fig. 1c and Supplementary Fig. 4e, f.). The anti-inflammatory response triggered by transferred sFc-stimulated hDC-SIGN^{+} bone-marrow-derived macrophages required the expression of the inhibitory Fc receptor (FcR) FcγRIIB, as FcγRIIB$^{-/-}$ recipient mice were not protected from inflammation induced by K/BxN serum (Fig. 1d). Collectively, these results were consistent with the *in vivo* requirements for IVIG protection previously defined[1–4], and demonstrated that ligation of hDC-SIGN by sFc on bone-marrow-derived myeloid cells is sufficient to induce an anti-inflammatory cellular response.

DC-SIGN engagement has been reported to result in dendritic cell production of IL-10 (refs 7–8), making this anti-inflammatory cytokine an appealing candidate responsible for mediating IVIG anti-inflammatory activity. However, IL-10$^{-/-}$ mice were protected from induced arthritis by IVIG similarly to wild-type controls (Supplementary Fig. 5). We next addressed other cytokines that could be responsible for this response. The T_H2 cytokine IL-4 has been shown to upregulate FcγRIIB surface expression on peripheral monocytes[12], and increase the threshold for activation by pathogenic immune complexes, consistent with the FcγRIIB requirement of IVIG[1,13,14]. Therefore, sFc-treated DC-SIGN^{+} bone-marrow-derived macrophages were administered to naive wild-type mice or IL-4$^{-/-}$ mice, and the recipient mice challenged with K/BxN serum. Wild-type recipients were protected from induced arthritis, whereas IL-4$^{-/-}$ recipients were not (Fig. 2a).

These results led us to predict that IVIG anti-inflammatory activity would require IL-4 signalling. Indeed, mice deficient in IL-4 (IL-4$^{-/-}$; Fig. 2b), the IL-4 receptor (IL-4Rα$^{-/-}$; Fig. 2c), or the IL-4R signalling adaptor (Stat6$^{-/-}$; Fig. 2c), were not protected from K/BxN-induced inflammation by IVIG or sFc. Further, monocytes in the peripheral blood and bone marrow of wild-type mice, but not IL-4Rα$^{-/-}$ mice, upregulated FcγRIIB after sFc administration (Supplementary Fig. 6). Next we examined whether exogenous T_H2 cytokines could also suppress autoantibody-induced inflammation. Mice were treated with cytokine immune complexes (ic)[15] of T_H2 cytokines IL-4 (IL-4ic), and IL-13 (IL-13ic), or a non-T_H2 cytokine complex of IL-3 (IL-3ic), and challenged with K/BxN serum. Inflammation was significantly attenuated after single administration of IL-4ic or IL-13ic, but not after IL-3ic treatment (Fig. 2d). However, IL-4ic treatment did not attenuate

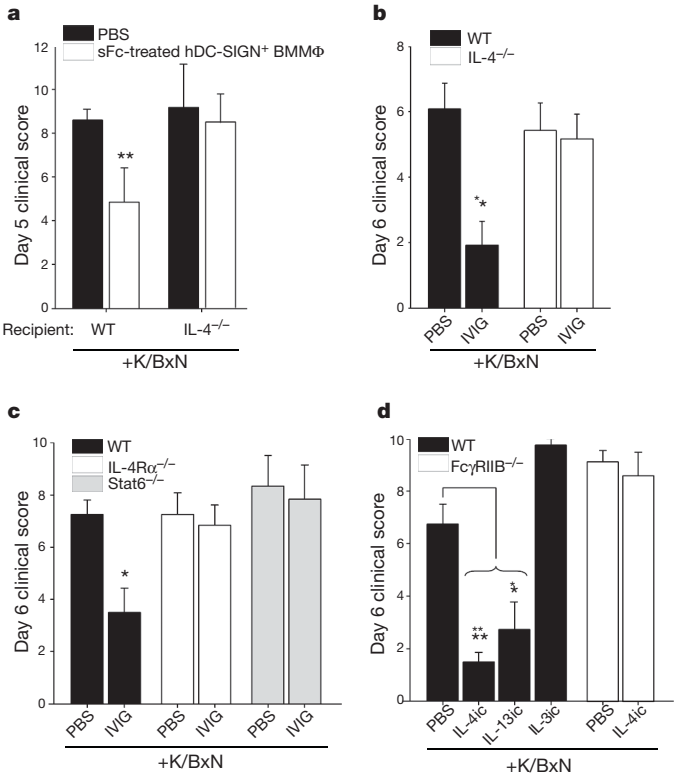


Figure 2 | IL-4 requirements of sFc anti-inflammatory activity. **a**, sFc-treated hDC-SIGN^{+} bone-marrow-derived macrophages (BMMΦ; white bars) or PBS was administered to K/BxN-treated wild-type or IL-4$^{-/-}$ recipient mice. $^{**}P < 0.002$ determined by Fisher LSD post-hoc test. **b**, Wild-type (black bars) and IL-4$^{-/-}$ (white bars) mice were treated with K/BxN sera and IVIG. $^{*}P < 0.01$ determined by Mann–Whitney's U test. **c**, Wild-type (black bars), IL-4Rα$^{-/-}$ (white bars), or Stat6$^{-/-}$ mice (grey bars) were given K/BxN sera and IVIG. $^{*}P < 0.01$ determined by Tukey's post-hoc test. **d**, Wild-type (black bars) and FcγRIIB$^{-/-}$ mice (white bars) were administered cytokine immune complexes (IL-4ic, IL-3ic, IL-13ic) and K/BxN sera. Means and standard deviations are plotted; $^{*}P < 0.01$, $^{**}P < 0.001$ determined by Mann–Whitney's U test.

inflammation in FcγRIIB$^{-/-}$ mice, consistent with IL-4ic also requiring the FcγRIIB to suppress inflammation (Fig. 2d).

We next examined whether T_H2 cytokines were induced after IVIG or sFc administration. No changes in *Il4* messenger RNA levels were observed (Fig. 3a and Supplementary Fig. 7a–c), so we surveyed cytokines known to induce *Il4* expression, including IL-33 (refs 16–18), IL-25 (refs 18–20), and thymic stromal lymphopoietin (TSLP)[18,21,22]. Interestingly, *Il33* mRNA was upregulated in wild-type mice after IVIG and sFc administration, but remained unchanged SIGN-R1$^{-/-}$ mice. We then administered IL-33, IL-25, or TSLP to mice challenged with K/BxN serum. Exogenous IL-33 fully suppressed K/BxN arthritogenic activity and induced IL-4 production *in vivo*, whereas IL-25 promoted only modest protection, and TSLP provided no protection (Fig. 3b), all of which correlated with systemic IL-4 and IL-13 levels (Fig. 3b and Supplementary Fig. 7d, e). Further, exogenous IL-33 and IL-4ic were unable to ameliorate serum-induced arthritis in IL-4Rα$^{-/-}$ mice (Fig. 3c and Supplementary Fig. 7e), indicating that IL-4Rα acts downstream of IL-33 in this pathway.

Our results support an anti-inflammatory cascade where DC-SIGN ligation by sFc promotes IL-33 production, IL-33 induces IL-4 expression, culminating in FcγRIIB upregulation on monocytes and macrophages (Supplementary Fig. 1). To confirm this, hDC-SIGN^{+}/SIGN-R1$^{-/-}$ mice were treated with arthritogenic sera and sFc, in combination with a blocking antibody to the IL-33 receptor (anti-IL-33Rα). This intervention ablated the ability of sFc to protect

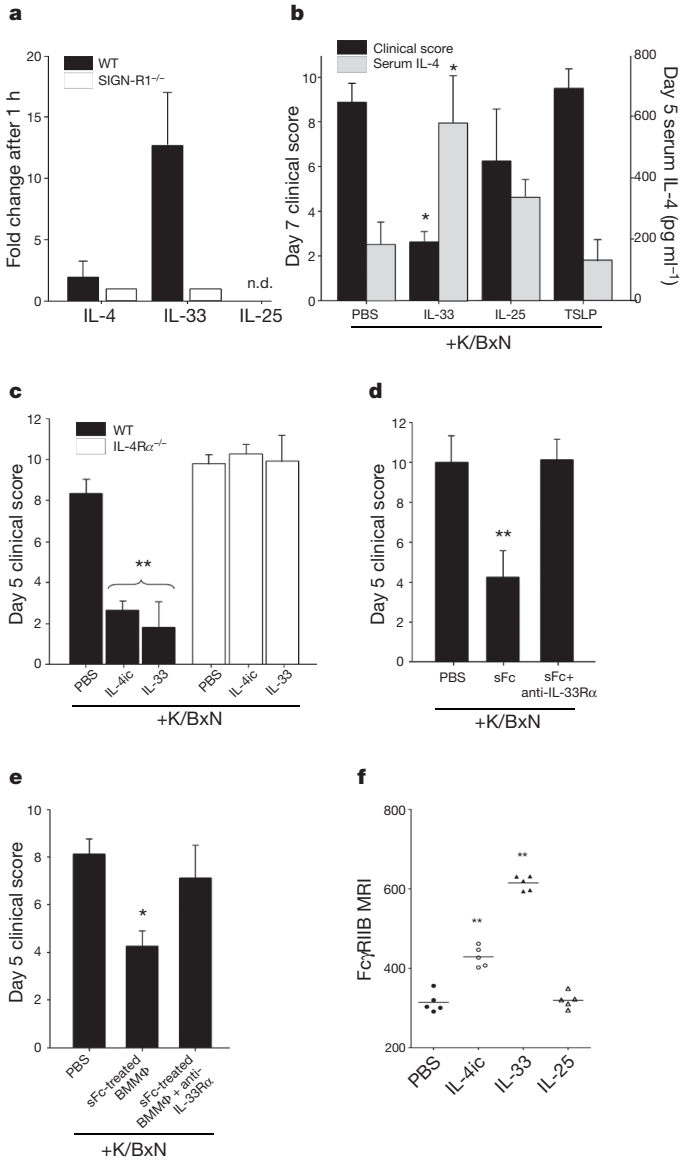


Figure 3 | IL-33 triggers IL-4 anti-inflammatory activity. **a**, Cytokine expression 1 h after IVIG administration in wild-type (black bars) or SIGN-R1$^{-/-}$ mice (white bars) determined by quantitative polymerase chain reaction (qPCR). n.d., not detected. **b**, K/BxN-treated wild-type mice received PBS, IL-33, IL-25, or TSLP. $*P < 0.05$ determined by Tukey's test. **c**, K/BxN-treated wild-type (black bars) or IL-4Rα$^{-/-}$ (white bars) mice received PBS, IL-4ic, or IL-33. $**P < 0.001$ determined by Tukey's test. **d**, hDC-SIGN^{+}/SIGN-R1$^{-/-}$ mice received K/BxN sera, sFc and anti-IL-33Rα. $**P < 0.001$ determined by Fisher LSD test. **e**, sFc-treated hDC-SIGN^{+} bone-marrow-derived macrophages were administered to wild-type mice, K/BxN- and anti-IL-33Rα-treated wild-type mice. Means and standard deviations are plotted; $*P < 0.05$ determined by Tukey's test. **f**, Individual mean fluorescence intensities (MFI) of bone marrow monocyte (CD11b^{+} Ly6G^{-}) FcγRIIB surface expression 24 h after PBS, IL-4, IL-33, or IL-25 treatment by FACS. $**P < 0.01$ determined by Tukey's test.

hDC-SIGN^{+}/SIGN-R1$^{-/-}$ mice (Fig. 3d and Supplementary Fig. 7f). Protection of transferred sFc-treated hDC-SIGN^{+} bone-marrow-derived macrophages was also diminished by anti-IL-33Rα treatment (Fig. 3e). Further, administration of exogenous IL-4ic or IL-33 increased FcγRIIB surface expression on monocytes, whereas IL-25 had no effect (Fig. 3f and Supplementary Fig. 8). IL-4 treatment down-regulated FcγRIIB expression on B cells (Supplementary Fig. 8), consistent with the diverse effects of this cytokine on different leukocyte types.

IL-4 can be produced by T cells and several innate immune cell populations[18], including basophils[23,24], mast cells[25,26], eosinophils[27] and progenitor cells[19]. IVIG activity is T-cell independent[4], thus eliminating these cells as a source of sFc-induced IL-4. To determine whether basophils were involved in this response, we selectively depleted these cells *in vivo*[28] (Fig. 4a and Supplementary Figs 9–11). Arthritis could be induced in basophil-depleted hDC-SIGN^{+}/SIGN-R1$^{-/-}$ mice (Supplementary Fig. 9a) but the protective capacity of sFc and IVIG was lost (Fig. 4a and Supplementary Fig. 11), indicating that these cells have a pivotal role . We next administered sFc and K/BxN sera to IL-4–GFP reporter mice (4get[29]), and found a twofold increase in GFP^{+} basophils (DX5^{+} FcεRI^{+} c-Kit^{-}) in the circulation of protected sFc-treated mice, indicating that basophils produced IL-4 in response to sFc (Fig. 4b). To determine whether basophils were ultimately responsible for the anti-inflammatory activity induced by sFc through IL-33, we transferred PBS- or IL-33-treated basophils to K/BxN-treated recipient mice (Fig. 4c, d and Supplementary Fig. 12a–c). IL-33-treated basophils derived from wild-type or FcγRIIB$^{-/-}$ mice were equally effective at suppressing arthritic inflammation, reducing serum IL-6 levels, and curbing leukocyte infiltration to arthritic paws (Fig. 4c, d and Supplementary Fig. 12d, e). These results confirm the anti-inflammatory potential of these cells, and support a model whereby IL-4 produced by basophils increases FcγRIIB expression on inflammatory macrophages (Supplementary Fig. 1).

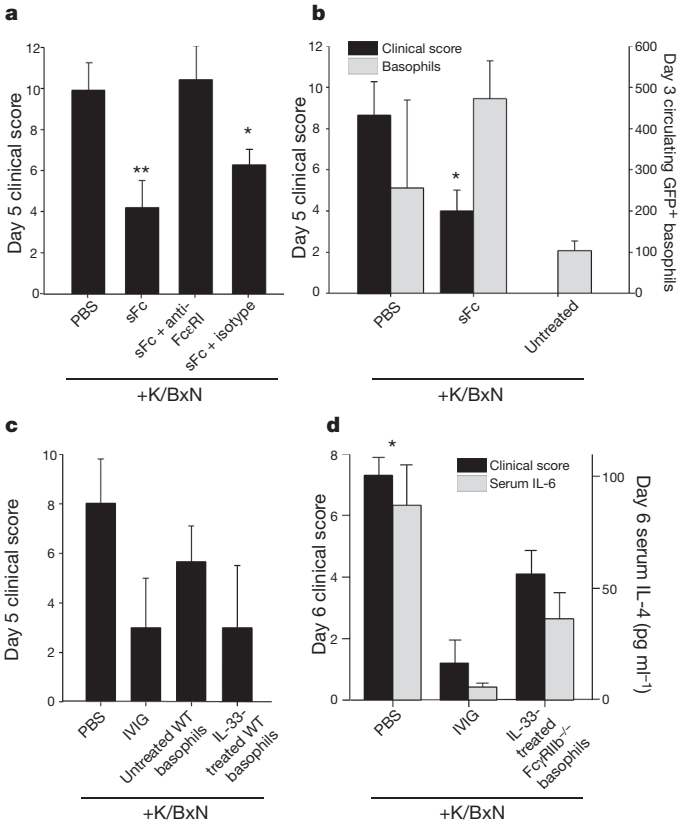


Figure 4 | Anti-inflammatory activity mediated by basophils. **a**, hDC-SIGN^{+}/SIGN-R1$^{-/-}$ mice were administered K/BxN sera, sFc and anti-FcεRI or an isotype control. $**P < 0.001$, $*P < 0.05$ determined by Fisher LSD test. **b**, 4get mice were administered K/BxN sera and sFc. Circulating IL-4^{+} basophils (grey bars, DX5^{+} FcεRI^{+} GFP^{+}) and clinical scores (black bars) are plotted. $*P < 0.05$ determined by Tukey's test. **c**, PBS or IL-33-treated basophils (DX5^{+} FcεRI^{+} c-Kit^{-}) were administered to K/BxN-treated wild-type mice. Control mice received PBS or IVIG. **d**, Basophils from IL-33-treated FcγRIIB$^{-/-}$ mice were administered to K/BxN-treated wild-type recipients. Clinical scores (black) and serum IL-6 levels (grey) are plotted. Means and standard deviations are plotted;$*P < 0.05$ determined by Mann–Whitney's U test.

Analysing the anti-inflammatory activity of IVIG led to identification of an endogenous, innate pathway in which sialylated IgG, a minor component of serum IgG antibodies, binds DC-SIGN, promoting production of IL-33, which expands IL-4$^+$ basophils. These cytokines are capable of suppressing autoantibody-mediated inflammation by modulating FcγRIIB expression on effector cells (Supplementary Fig. 1). IL-4 and IL-33 have pleiotropic activities, and mediate T_H2 responses to helminth parasites and allergens[18], as well as enhance inflammatory arthritis[30,31], in addition to their activities reported here. Cytokine concentration, cellular environment, and differential responses of individual cell types are likely to explain these distinct effector functions.

Various stimuli have been reported to modulate the level of IgG Fc sialylation, and could regulate this intrinsic pathway. Antigenic stimulation results in the production of pro-inflammatory, antigen-specific, asialylated IgG antibodies[2]. Pathogenic autoantibodies, such as those produced during rheumatoid arthritis that recognize citrullinated peptides, similarly show reduced sialic acid as compared to other serum antibodies[32]. Conversely, increases in sialylated IgG antibodies occur during pregnancy[33], which could contribute to the remission in arthritis seen in pregnant women. Therefore, affecting the sialylation of IgG antibodies could provide an intrinsic mechanism for regulating T_H2 cytokine production by innate myeloid cells in a DC-SIGN-dependent manner, provide a means for maintaining homeostasis, and is an attractive therapeutic approach to suppressing inflammation in autoimmune diseases.

METHODS SUMMARY

hDC-SIGN BAC transgenic mice were created using the BAC clone CTD2102F19 (Invitrogen) containing the human *DC-SIGN* gene. The BAC was linearized by the NotI restriction endonuclease. The human *DC-SIGN* gene fragment was purified and injected into one-day-old C57BL/6 embryos via pronuclear microinjection. The embryos were then implanted into ICR surrogate females and the resulting progeny were screened by PCR for the presence of the human *DC-SIGN* transgene. hDC-SIGN$^+$ mice were crossed to SIGN-R1$^{-/-}$ mice to generate hDC-SIGN$^+$/SIGN-R1$^{-/-}$ lines.

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

Received 1 October 2010; accepted 20 April 2011.
Published online 19 June 2011.

1. Nimmerjahn, F. & Ravetch, J. V. Anti-inflammatory actions of intravenous immunoglobulin. *Annu. Rev. Immunol.* **26,** 513–533 (2008).
2. Kaneko, Y., Nimmerjahn, F. & Ravetch, J. V. Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science* **313,** 670–673 (2006).
3. Anthony, R. M. *et al.* Recapitulation of IVIG anti-inflammatory activity with a recombinant IgG Fc. *Science* **320,** 373–376 (2008).
4. Anthony, R. M., Wermeling, F., Karlsson, M. C. & Ravetch, J. V. Identification of a receptor required for the anti-inflammatory activity of IVIG. *Proc. Natl Acad. Sci. USA* **105,** 19571–19578 (2008).
5. Granelli-Piperno, A. *et al.* Dendritic cell-specific intercellular adhesion molecule 3-grabbing nonintegrin/CD209 is abundant on macrophages in the normal human lymph node and is not required for dendritic cell stimulation of the mixed leukocyte reaction. *J. Immunol.* **175,** 4265–4273 (2005).
6. Soilleux, E. J. *et al.* Constitutive and induced expression of DC-SIGN on dendritic cell and macrophage subpopulations *in situ* and *in vitro*. *J. Leukoc. Biol.* **71,** 445–457 (2002).
7. Geijtenbeek, T. B. & Gringhuis, S. I. Signalling through C-type lectin receptors: shaping immune responses. *Nature Rev. Immunol.* **9,** 465–479 (2009).
8. Gringhuis, S. I. *et al.* C-type lectin DC-SIGN modulates Toll-like receptor signaling via Raf-1 kinase-dependent acetylation of transcription factor NF-κB. *Immunity* **26,** 605–616 (2007).
9. Hodges, A. *et al.* Activation of the lectin DC-SIGN induces an immature dendritic cell phenotype triggering Rho-GTPase activity required for HIV-1 replication. *Nature Immunol.* **8,** 569–577 (2007).
10. Korganow, A. S. *et al.* From systemic T cell self-reactivity to organ-specific autoimmune disease via immunoglobulins. *Immunity* **10,** 451–461 (1999).
11. Schaefer, M. *et al.* Decreased pathology and prolonged survival of human DC-SIGN transgenic mice during mycobacterial infection. *J. Immunol.* **180,** 6836–6845 (2008).
12. Pricop, L. *et al.* Differential modulation of stimulatory and inhibitory Fcγ receptors on human monocytes by Th1 and Th2 cytokines. *J. Immunol.* **166,** 531–537 (2001).
13. Bruhns, P., Samuelsson, A., Pollard, J. W. & Ravetch, J. V. Colony-stimulating factor-1-dependent macrophages are responsible for IVIG protection in antibody-induced autoimmune disease. *Immunity* **18,** 573–581 (2003).
14. Samuelsson, A., Towers, T. L. & Ravetch, J. V. Anti-inflammatory activity of IVIG mediated through the inhibitory Fc receptor. *Science* **291,** 484–486 (2001).
15. Finkelman, F. D. *et al.* Anti-cytokine antibodies as carrier proteins. Prolongation of *in vivo* effects of exogenous cytokines by injection of cytokine-anti-cytokine antibody complexes. *J. Immunol.* **151,** 1235–1244 (1993).
16. Schmitz, J. *et al.* IL-33, an interleukin-1-like cytokine that signals via the IL-1 receptor-related protein ST2 and induces T helper type 2-associated cytokines. *Immunity* **23,** 479–490 (2005).
17. Neill, D. R. *et al.* Nuocytes represent a new innate effector leukocyte that mediates type-2 immunity. *Nature* **464,** 1367–1370 (2010).
18. Paul, W. E. & Zhu, J. How are T_H2-type immune responses initiated and amplified? *Nature Rev. Immunol.* **10,** 225–235 (2010).
19. Saenz, S. A. *et al.* IL25 elicits a multipotent progenitor cell population that promotes T_H2 cytokine responses. *Nature* **464,** 1362–1366 (2010).
20. Fort, M. M. *et al.* IL-25 induces IL-4, IL-5, and IL-13 and Th2-associated pathologies *in vivo*. *Immunity* **15,** 985–995 (2001).
21. Soumelis, V. *et al.* Human epithelial cells trigger dendritic cell mediated allergic inflammation by producing TSLP. *Nature Immunol.* **3,** 673–680 (2002).
22. Wang, Y. H. *et al.* IL-25 augments type 2 immune responses by enhancing the expansion and functions of TSLP-DC-activated Th2 memory cells. *J. Exp. Med.* **204,** 1837–1847 (2007).
23. Min, B. *et al.* Basophils produce IL-4 and accumulate in tissues after infection with a Th2-inducing parasite. *J. Exp. Med.* **200,** 507–517 (2004).
24. Seder, R. A. *et al.* Mouse splenic and bone marrow cell populations that express high-affinity Fc epsilon receptors and produce interleukin 4 are highly enriched in basophils. *Proc. Natl Acad. Sci. USA* **88,** 2835–2839 (1991).
25. Seder, R. A. *et al.* Production of interleukin-4 and other cytokines following stimulation of mast cell lines and *in vivo* mast cells/basophils. *Int. Arch. Allergy Appl. Immunol.* **94,** 137–140 (1991).
26. Wang, M., Saxon, A. & Diaz-Sanchez, D. Early IL-4 production driving Th2 differentiation in a human *in vivo* allergic model is mast cell derived. *Clin. Immunol.* **90,** 47–54 (1999).
27. Shinkai, K., Mohrs, M. & Locksley, R. M. Helper T cells regulate type-2 innate immunity *in vivo*. *Nature* **420,** 825–829 (2002).
28. Sokol, C. L., Barton, G. M., Farr, A. G. & Medzhitov, R. A mechanism for the initiation of allergen-induced T helper type 2 responses. *Nature Immunol.* **9,** 310–318 (2008).
29. Mohrs, M., Shinkai, K., Mohrs, K. & Locksely, R. M. Analysis of type 2 immunity *in vivo* with a bicistronic IL-4 reporter. *Immunity* **15,** 303–311 (2001).
30. Ohmura, K. *et al.* Interleukin-4 can be a key positive regulator of inflammatory arthritis. *Arthritis Rheum.* **52,** 1866–1875 (2005).
31. Xu, D. *et al.* IL-33 exacerbates autoantibody-induced arthritis. *J. Immunol.* **184,** 2620–2626 (2010).
32. Scherer, H. U. *et al.* Glycan profiling of anti-citrullinated protein antibodies isolated from human serum and synovial fluid. *Arthritis Rheum.* **62,** 1620–1629 (2010).
33. van de Geijn, F. E. *et al.* Immunoglobulin G galactosylation and sialylation are associated with pregnancy-induced improvement of rheumatoid arthritis and the postpartum flare: results from a large prospective cohort study. *Arthritis Res. Ther.* **11,** R193 (2009).

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

Acknowledgements The authors thank P. Smith, M. Kibe, J. Brown and K. Velinzon for technical support, K. L. Jeffrey, C. Cheong, R. Steinman and J. J. Lee for discussions, M. Pack for providing human spleen sections, A. McKenzie for providing SIGN-R1$^{-/-}$ mice, T. Sparwasser for providing CD11c-hDC-SIGN mice, C. G. Park for providing hDC-SIGN and hDC-SIGN-R expressing cell lines, and H. Watarai for providing anti-IL-25R antibodies. R.M.A. is an Irvington Institute fellow of the Cancer Research Institute. F.W. is supported by the Wenner-Gren Foundations, Sweden. This work was performed with support from Virdante Pharmaceuticals and NIH grants to J.V.R.

Author Contributions R.M.A., T.K., F.W. and J.V.R. designed the experiments and interpreted the results. R.M.A., T.K. and F.W. performed the experiments, and R.M.A. and J.V.R. wrote the manuscript.

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. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.V.R. (ravetch@rockefeller.edu).

METHODS

Mice. Eight-to-twelve-week old, sex and age matched mice were used for all experiments in compliance with federal laws, institutional guidelines and have been approved by the Rockefeller University. Wild-type C57BL/6, wild-type BALB/c, NOD, IL-4$^{-/-}$, IL-4Rα$^{-/-}$, Stat6$^{-/-}$, IL-10$^{-/-}$, 4get (IL-4–GFP reporter) mice were purchased from Jackson Laboratories, and maintained at the Rockefeller University animal facility. FcγRIIB$^{-/-}$ mice[34] were generated previously in the laboratory. SIGN-R1$^{-/-}$ mice[35] were provided by A. McKenzie. CD11c-hDCSIGN^{+} mice[11] were provided by T. Sparwasser. KRN T-cell receptor transgenic mice on a C57BL/6 background (K/B) were gifts from D. Mathis and C. Benoist and were bred to NOD mice to generate K/BxN mice[10]. K/BxN serum was prepared as described previously[13]. Briefly, serum was separated from blood collected from the K/BxN mice (6–12-weeks old). Several weeks of serum collection were pooled together and frozen in aliquots to be used in the experiments described here. One intravenous injection of 200 μl K/BxN serum was used to induce arthritis. Severity of arthritis was scored by clinical examination by adding the index of all four paws, where 0 is unaffected, 1 is swelling of one joint, 2 is swelling of more than one joint, and 3 is severe swelling of the entire paw. All experiments shown yielded similar results at least 3 times with treatment groups of 4–5 mice, and means and standard deviations are plotted in bar graphs.

hDC-SIGN BAC transgenic mice were created using the BAC clone CTD2102F19 (Invitrogen) containing the human *DC-SIGN* gene. The BAC was linearized by the NotI restriction endonuclease. The human *DC-SIGN* gene fragment was purified and injected into one-day-old C57BL/6 embryos via pronuclear microinjection. The embryos were then implanted into ICR surrogate females and the resulting progeny were screened by PCR for the presence of the human *DC-SIGN* transgene. hDC-SIGN^{+} mice were crossed to SIGN-R1$^{-/-}$ mice to generate hDC-SIGN^{+}/SIGN-R1$^{-/-}$ lines.

Reagents and treatments. IVIG (Octagam, Octapharma) or IVIG-derived Fcs was enriched for terminal sialic acid using SNA-agarose[2] (Vector Laboratories) or hypersialyated *in vitro* as previously described[3] to generate sFc. AsialoFc was generated by treating Fcs with neuraminidase (NEB) as per the manufacturer's directions. Sialic acid content was verified by lectin blotting with SNA-biotin (Vector Laboratories). IVIG and IVIG derivations were administered intravenously (i.v.) at 1 g kg^{-1}, SNA-enriched IVIG at 0.1 g kg^{-1}, and sFc at 0.03 g kg^{-1} one hour before K/BxN sera administration.

Mice receiving cytokine:immune complexes (ic) with prolonged half-life were treated with a single i.v. injection 2.5 μg of cytokine (IL-3, IL-4, IL-13; Peprotech) and 12.5 μg of neutralizing antibody at day 0. Neutralizing antibodies used were anti-IL-3 (MP2-8F8, Biolegend), anti-IL4 (11B11, BD Biosciences), and anti-IL-13 (eBio1316H, eBioscience). Other cytokine treatments included intraperitoneal (i.p.) administration of 400 ng (or 800 ng) of IL-25 (R&D), 400 ng of IL-33 (R&D), or 1 μg of TSLP (R&D) on days 0, 1, 2, and 3. Basophils were depleted as described[28] by daily i.p. injection with 10 μg of anti-FcεRI (MAR-1, eBioscience) or hamster IgG isotype control (eBioscience) on days 0–5. Alternatively, mice received a single i.v. injection of 30 μg anti-CD200RL3 (ref. 36) (Ba103, Hycult Biotech) or rat IgG isotype control (BD Biosciences). IL-33Rα was blocked by i.v. injection of 80 μg of anti-IL-33Rα (DT8, MD Biosciences) or rat IgG1 isotype control (BD Biosciences) on day 0. hDC-SIGN was blocked *in vivo* by administration of 125 μg E9E A8 (ref. 37) or isotype control mouse IgG2a (BioLegend).

IL-6 was measured in serum by ELISA as suggested by the manufacturer (BioLegend). Serum IL-4 and IL-13 was measured using an *in vivo* cytokine capture assay as described[38]. Briefly, 10 μg biotinylated anti-IL-4 antibody (clone BVD4-1D11, BioLegend) or biotinylated anti-IL-13 (eBio1316HA, eBioscience) was injected i.v. into treated mice, and sera were collected 24 h later. Cytokine levels were quantified by ELISA assay using anti-IL-4 (BVD6-24G1, BioLegend) or anti-IL-13 (eBio13A, eBioscience) as capture antibodies.

Splenic RNA was purified using RNeasy Mini Kits (Qiagen) and reverse-transcribed using Verso cDNA synthesis kit (Thermo Scientific). Quantitative PCR (qPCR) was conducted in 7300 Real-time PCR System (Life Technologies) with primer-probe sets for mouse IL-4, IL-13, IL-33, IL-25, or rRNA (Life Technologies), and gene expression levels were determined by normalization to rRNA levels.

Saturation binding experiments were performed as previously described[4], comparing CHO and CHO-hDC-SIGN cells or Hep-CD81 and Hep-hDC-SIGN-R cells.

Flow cytometry. Single cell suspensions were prepared from peripheral blood, spleen, bone marrow, or paws from mice. After red blood cell lysis, cells were stained with the indicated monoclonal antibodies, and subjected to analysis using a FACSCalibur or LSR-II cytometer (BD Biosciences). Human leukocytes were obtained from peripheral blood samples (New York Blood Center) after density gradient centrifugation (Ficoll-Paque, GE Healthcare). Antibodies used for murine cell staining were as follows: anti-CD19 (1D3), anti-B220 (RA3-6B2), anti-CD3ε (145-2C11), anti-CD11b (M1/70), anti-Ly6G (1A8), anti-CD11c (HL3), anti-I-A^{b} (AF6-120.1), anti-CD49b (DX5 and HMa2), anti-c-Kit (2B8), anti-CD45.2 (104) from BD Biosciences, anti-NKp46 (29A1.4), anti-SIGN-R1 (22D1), anti-CD123 (5B11) from eBioscience, anti-hDC-SIGN (9E9A8), anti-FceR1 (MAR-1) from Biolegend, anti-FcγRIIB (K9.361), anti-hDC-SIGNR (120604) from R&D systems. Antibodies used for human cell staining were: anti-CD14 (M5E2), anti-CD16 (B73.1), anti-CD3 (UCHT1), anti-CD56 (B159), anti-CD19 (SJ25C1), anti-CD11c (B-Ly6), anti-HLA-DR (L243 (G46-6)), anti-hDC-SIGN (AZND1) from BD Biosciences, and anti-hDC-SIGN (9E9A8, Biolegend). AccuCheck Counting Beads (Invitrogen) were used to quantify cells.

Bone-marrow macrophage and dendritic cell cultures and transfers. Bone-marrow-derived macrophages were cultured as described previously[39]. Briefly, marrow was recovered from tibias and femurs of mice, and seeded in non-tissue culture treated 10-cm plates with DMEM supplemented with 10% fetal bovine serum, 2% penicillin/streptomicin (Invitrogen), 1% glutamine 200 mM (Invitrogen), 0.1% β-mercaptoethanol, IL-3 (5 ng ml^{-1}, Peprotech) and M-CSF (5 ng ml^{-1}, Peprotech) overnight at 37 °C, 5% CO_2. The next day, non-adherent cells were recovered and plated in 10-cm non-tissue culture treated plates in supplemented DMEM with cytokines, and cultured for 5–7 days. Once the cultured cells were mature macrophages (>90% CD11b^{+} F4/80^{+} by FACS), the cells were detached and 2×10^{6} macrophages were plated per well in 6-well plates, and allowed to attach overnight. The next day, the cells were pulsed with IVIG (15 mg ml^{-1}), BSA (15 mg ml^{-1}, Sigma), sFc (0.5 mg ml^{-1}), or asialoFc (0.5 mg ml^{-1}) for 30 min at 37 °C. The cells were recovered, washed thoroughly in cold PBS, and 1×10^{6} macrophages were administered i.v. into naive recipients. One hour later, the recipient mice were treated with K/BxN sera. Dendritic cells were cultured from mouse tibia and femur bone marrow cells as described[40]. Briefly, 1×10^{6} cells ml^{-1} were plated in 24-well plates with DMEM supplemented with 10% FBS and 10 ng ml^{-1} mouse granulocyte macrophage colony stimulating factor (GM-CSF, Peprotech). On day 6, loosely adherent cells were collected by gentle pipetting, and were subjected to flow cytometric analysis or bone marrow cell transfer experiments as described above.

Histology. Human lymph node samples were from ISL Bio, and M. Pack provided human spleen samples. Spleens or lymph nodes embedded in O.C.T. compound (Sakura Finetek) were fixed in ice-cold acetone for 10 min, and stained with anti-SIGN-R1, anti-hDC-SIGN or anti-hDC-SIGN-R for 1 h at 25 °C in combination with antibodies for macrophages or B cells. Antibodies used included anti-F4/80 (BM8, Invitrogen) for mouse red pulp macrophages, anti-B220 for mouse B cells, anti-hCD20 (2H7, Biolegend) for human B cells, and anti-CD68 (Y1/82A, Biolegend) for human macrophages. Sections were visualized by wide-field fluorescence microscope (Zeiss).

Basophil adoptive transfers. Basophils were expanded by administering IL-3ic[41], as described above, to wild-type or FcγRIIB$^{-/-}$ mice. Five days later, IL-3ic-treated mice were administered PBS or IL-33 (400 ng) i.p. The next day, basophils (DX5^{+} FcRI^{+} cKit^{-}) were sorted using a FACSAria II (BD Biosciences). Sorted basophils were washed in cold PBS, and 0.7×10^{6} basophils were administered to naive recipient mice subsequently administered K/BxN sera. A Wright–Giemsa stain of sorted, cytospun basophils was performed as suggested by the manufacturer (Sigma).

34. Takai, T., Ono, M., Hikida, M., Ohmori, H. & Ravetch, J. V. Augmented humoral and anaphylactic responses in FcγRII-deficient mice. *Nature* **379,** 346–349 (1996).
35. Lanoue, A. *et al.* SIGN-R1 contributes to protection against lethal pneumococcal infection in mice. *J. Exp. Med.* **200,** 1383–1393 (2004).
36. Obata, K. *et al.* Basophils are essential initiators of a novel type of chronic allergic inflammation. *Blood* **110,** 913–920 (2007).
37. Cheong, C. *et al.* New monoclonal anti-mouse DC-SIGN antibodies reactive with acetone-fixed cells. *J. Immunol. Methods* **360,** 66–75 (2010).
38. Finkelman, F., Morris, S., Orekhova, T. & Sehy, D. The *in vivo* cytokine capture assay for measurement of cytokine production in the mouse. *Curr. Protoc. Immunol.* 6.28.1–6.28.10 (2003).
39. Jeffrey, K. L. *et al.* Positive regulation of immune cell function and inflammatory responses by phosphatase PAC-1. *Nature Immunol.* **7,** 274–283 (2006).
40. Inaba, K. *et al.* Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor. *J. Exp. Med.* **176,** 1693–1702 (1992).
41. Ohmori, K. *et al.* IL-3 induces basophil expansion *in vivo* by directing granulocyte-monocyte progenitors to differentiate into basophil lineage-restricted progenitors in the bone marrow and by increasing the number of basophil/mast cell progenitors in the spleen. *J. Immunol.* **182,** 2835–2841 (2009).

LETTER

doi:10.1038/nature10175

A salt–water reservoir as the source of a compositionally stratified plume on Enceladus

F. Postberg[1,2], J. Schmidt[3], J. Hillier[4], S. Kempf[2,5,6] & R. Srama[2,7]

The discovery of a plume of water vapour and ice particles emerging from warm fractures ('tiger stripes') in Saturn's small, icy moon Enceladus[1–6] raised the question of whether the plume emerges from a subsurface liquid source[6–8] or from the decomposition of ice[9–12]. Previous compositional analyses of particles injected by the plume into Saturn's diffuse E ring have already indicated the presence of liquid water[8], but the mechanisms driving the plume emission are still debated[13]. Here we report an analysis of the composition of freshly ejected particles close to the sources. Salt-rich ice particles are found to dominate the total mass flux of ejected solids (more than 99 per cent) but they are depleted in the population escaping into Saturn's E ring. Ice grains containing organic compounds are found to be more abundant in dense parts of the plume. Whereas previous Cassini observations were compatible with a variety of plume formation mechanisms, these data eliminate or severely constrain non-liquid models and strongly imply that a salt-water reservoir with a large evaporating surface[7,8] provides nearly all of the matter in the plume.

During three traversals (E4, E5 and E7) of the Enceladian plume (Supplementary Fig. 1), Cassini's Cosmic Dust Analyser (CDA) produced time-of-flight mass spectra generated by high-velocity impacts of individual grains onto a metal target[14]. A specific configuration of CDA at E5 (Supplementary Information) provided a sampling rate high enough to infer a spatial compositional profile in the plume.

Most plume spectra can be assigned to one of three compositional types (Supplementary Fig. 4) previously detected in the E ring[8,15,16]. One of these types (type III, ~6% of E-ring detections) is particularly rich in sodium and potassium salts (0.5–2% by mass). The composition of these salt-rich grains[8] closely resembles the predicted composition of an Enceladian ocean that has been in prolonged contact with the rocky core of the moon[17]. Because of this compelling compositional match, the grains are thought to form from frozen droplets present as spray over a liquid reservoir close to the surface[8].

The E5 fly-by crossed the plume almost from north to south, with a closest approach to Enceladus of 21 km; this closest approach was located at the fringe of the plume (Fig. 3, Supplementary Fig. 1). The proportions of the three main compositional types exhibit significant variations with position in the plume (Fig. 1). Shortly before closest approach, the proportion of salt-rich grains increases steeply from E-ring background levels, reaching >40% just after closest approach, followed by a shallower decrease towards the dense plume. Simultaneously, the proportion of salt-poor type I grains starts to decrease shortly before closest approach. Furthermore, there is a less pronounced increase in the proportion of type II grains (containing organic compounds and/or silicates[16]) after closest approach when Cassini enters the denser parts of the plume. The absolute number of all three types steadily increases until ~20 s after closest approach (Fig. 2a).

The most plausible explanation for the simultaneous increase of type III and decrease of type I proportions is that salt-rich grains become more abundant close to the sources located in the south polar terrain. Salt-poor type I grains can form by homogenous nucleation from the gas phase[7,8]. In contrast, the salt-rich type III particles have to be formed from salt-ice condensation cores, presumably frozen spray of salt water[8]. The latter mechanism naturally forms larger grains, which then receive lower average ejection speeds for a given density and speed of the carrier gas[7]. Indeed, measurements of CDA's High Rate Detector[18] and photometry of the plume in the near-infrared[19] both indicate that the grain ejection velocity decreases with size. This would lead to the compositional stratification of the plume, with an (observed) increase of the proportion of salt-rich grains close to the sources. This idea is supported by the generally larger yields of ions from impacts of salt-rich particles recorded in the E ring (Supplementary Fig. 5). Proportionally fewer of the slower (larger) salt-rich grains escape the moon's gravity and end up in the E ring, explaining the dominance of the smallest of the three types (type I) in the E ring[8].

We use a simple model (Supplementary Information) to test the hypothesis of whether a general size difference between salt-poor and salt-rich grains can cause the observed variations. We use the two-body gravitational interaction[20] of plume particles with Enceladus to construct the particle number density expected along the spacecraft trajectory. For this fly-by, we consider only grains larger than the detector threshold of ~0.2 μm. In reality, the size distributions of salt-rich and salt-poor grains overlap, but to model different average sizes we assume that all grains larger than a certain radius (found to be 0.6 μm) are salt-rich, and those smaller are salt-poor. Before ejection, both species obey a single continuous power-law size distribution. We assume a flux of particles emerging from many sources uniformly distributed across all four tiger stripes. Starting speeds are size-dependent because of the size-dependent frictional force governing acceleration by a gas[7]. This leads to the observed tendency of larger particles to preferentially populate the lower regions of the plume[18,19]. A fit of this simple model to the E5 data (Fig. 2) qualitatively reproduces the rise in the fraction of salt-rich grains around closest approach.

Inspired by the observation in the ultraviolet[21] of both slow, diffuse sources and supersonic, collimated gas jets, we add eight jet-like particle sources (Supplementary Fig. 2), at locations identified in Cassini images[22] and combine them with the uniformly distributed sources in a bi-component model. For the observed high gas speeds in these jets[21] (>1,000 m s^{-1}), abundant vapour condensation is expected in the vents[7], resulting in a large fraction of small salt-poor grains, which is reflected by using a steeper size distribution for the jets. The bi-component model further improves the fit to the observations (Fig. 2). Establishing a link between size, ejection speed and composition, our results indicate not only compositional but dynamical substructures in the ice plume, with larger salt-rich grains ejected mainly from slower sources distributed across the tiger stripes.

The models permit us to extrapolate the measured abundance of salt-rich grains down to the densest, near-surface regions of the plume (Fig. 3), representing the conditions immediately after ejection. We find that slow sources (combining contributions from all tiger stripes) dominate the

[1]Institut für Geowissenschaften, Universität Heidelberg, 69120 Heidelberg, Germany. [2]Max-Planck-Institut für Kernphysik, 69117 Heidelberg, Germany. [3]Institut für Physik und Astronomie, Universität Potsdam, 14476 Potsdam-Golm, Germany. [4]Planetary and Space Sciences Research Institute, The Open University, Milton Keynes MK7 6AA, UK. [5]IGEP,Technische Universität Braunschweig, 38106 Braunschweig, Germany. [6]LASP, University of Colorado, Boulder, Colorado 80303, USA. [7]IRS, Universität Stuttgart, 70569 Stuttgart, Germany.

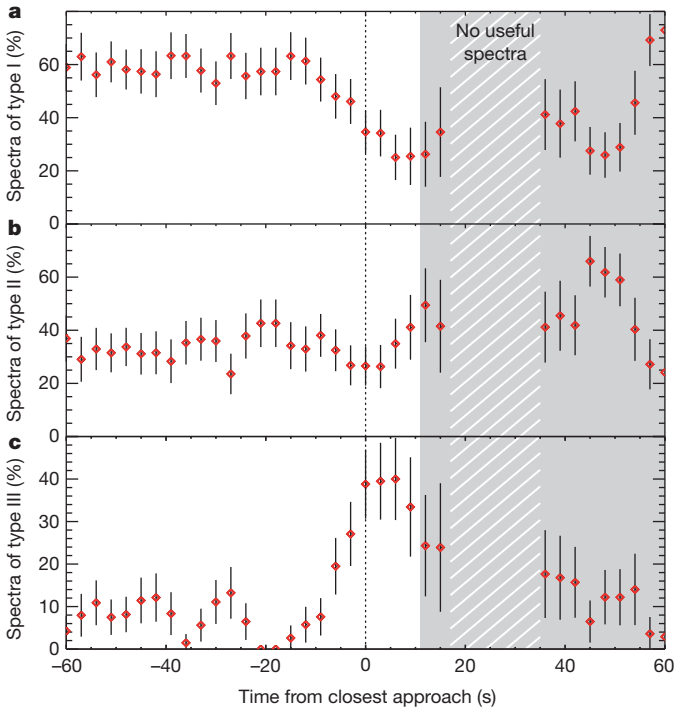


Figure 1 | Proportions of particles of different spectral type detected during the central period of the E5 fly-by. The measured frequencies of particles of different spectral types (**a**, type I; **b**, type II; **c**, type III) along the E5 trajectory reflect proportions and not absolute abundances. Type I spectra imply almost pure water ice grains with an average Na/H_2O mixing ratio of about 10^{-7}, whereas type II spectra exhibit contributions from organic and/or siliceous material[8,16]. Type III exhibit drastically increased alkali salt content. They on average contain 0.5–2% sodium and potassium salts by mass, mainly NaCl, $NaHCO_3$ and/or Na_2CO_3 (ref. 8). Each data point represents an interval of ±4.5 s and includes ~40 spectra. Error bars, standard error of the mean derived from counting statistics. The E-ring background dominates the particle flux until ~15 s before closest approach and is already dense enough to trigger CDA spectrum recording at its maximum rate (in the optimized configuration this is about nine spectra per second, data for five of which could be transmitted), so the detection rate could not increase further during plume crossings. During the period of highest impact rate, between ~18 s and ~35 s after closest approach, too few evaluable spectra were obtained (hatched region) owing to overload of the instrument electronics (Supplementary Information). Similarly, unspecified selection effects may have led to fluctuations in the type statistics starting from ~11 s after closest approach. Although the data obtained during this time interval (grey) may therefore have been affected by instrument performance issues, they exhibit a stable trend matching the model predictions presented in this work (Fig. 3).

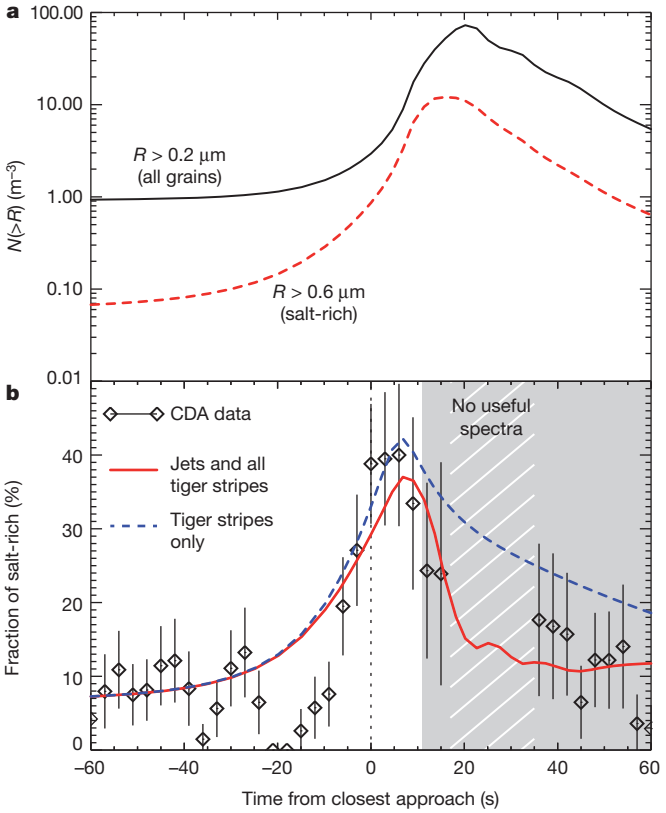


Figure 2 | Modelling of the E5 measurements. **a**, The expected total particle number densities along the E5 fly-by trajectory for particles >0.6 μm (representing salt-rich grains; red dashed line) and >0.2 μm (all particles, including salt-poor; solid black line). In **b**, the resulting compositional profiles are compared to the type III particle profile, measured by CDA (open diamonds). The dashed line shows the compositional profile obtained from a model employing a uniform particle flux emerging from all four tiger stripes. The solid line shows a model including eight faster and more collimated jet-like particle sources[14]. Although both models fit the observations well, the contribution from the jet particles helps fit the rapid decrease of salt-rich grains about five seconds after closest approach, as well as the relatively low level of these grains after 40 s from closest approach, when the spacecraft was still within range of the jets (Fig. 3 and Supplementary Figs 1, 2). For the uniform tiger stripe sources, we use a gas speed[26] of 500 m s^{-1} and a relatively broad distribution of ejection angles, with a mean ejection angle of 35° measured from the surface normal. For the fast jets, we use a gas speed of 1,200 m s^{-1} and a mean ejection angle of 15° (Supplementary Information). Comparison of the model particle number densities to measurements performed with the CDA's High Rate Detector[1] allows us to include a background flux of E-ring particles, dominating the signal of particles >0.2 μm before ten seconds before closest approach. Error bars, standard error of the mean derived from counting statistics.

flux of particles above the detection threshold of the CDA. In both model versions, ~70% of ejected grains are salt-rich. Owing to their larger sizes, they completely dominate the mass flux (>99%). The increased abundance of type II ice grains in the core region of the plume, even at high altitudes, cannot be reproduced with a size-dependent model, and implies a general enrichment of ice grains containing organic material, probably associated with fast collimated jets.

Our conclusion that a stratified plume, which primarily produces salt-rich grains, is feeding a salt-poor E ring is a direct consequence of the measurements and independent of model parameters. If the 6% of salt-rich grains detected in the E ring were also typical of the plume production region, then contributions from 'dry' sources (such as ice sublimation or clathrate decomposition) would be viable. In contrast, a plume source dominated by salt-rich grains, as reported here, eliminates the possibility of significant contributions from dry, sodium-poor sources and severely constrains or rules out non-liquid models in their present form. In a clathrate decomposition scenario[9,10], a substantial part of the water vapour in the plume gas forms from the sublimation of ice grains entrained in a flow of initially released volatile gases[9,23] (CO_2, N_2, CO, CH_4). However, sublimation of sodium-rich grains would lead to the release of sodium into both the plume's gas phase and the E ring, which is not observed[13]. Models employing sublimation from warm ice currently cannot account for the predominant formation of salt-rich ice grains. As all sodium compounds should have been originally bound in rock after the moon's formation, the salt in plume grains must have been extracted from Enceladian rocky material by water. To preserve high salt concentrations in ice, rapid freezing would be necessary. Thus, to reconcile our measurements with ice sublimation as a dominant plume source, it is necessary to invoke a flash-frozen salty ice layer across the entire active region. We regard this scenario as physically unlikely. Sublimation scenarios in which ice grains are predominantly generated by the recondensation of a fraction of the sublimated water vapour (see, for example, refs 11, 12) can be entirely eliminated, as such a process cannot form salt-rich grains. A geyser-like scenario[6] with violently boiling salt-water is excluded by the absence of sodium in the vapour phase[13].

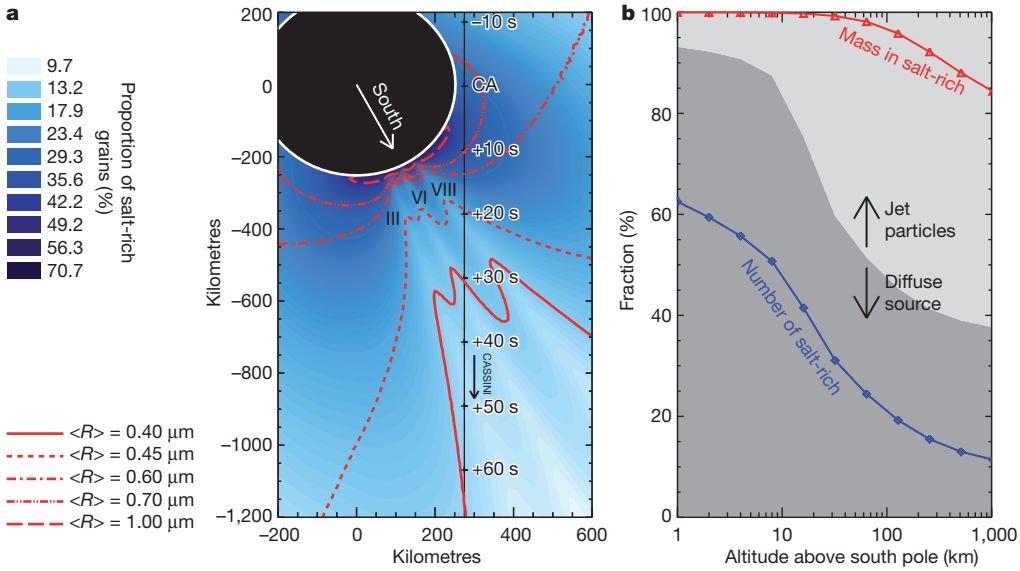


Figure 3 | Compositional and size profile of the ice plume. **a**, Background colours show the proportion of salt-rich grains obtained from the model. Overlaid are contours of constant mean particle radius, $<R>$. In contrast to the compositional profile, the size contours show the pure plume emission without the E-ring background. The projection used is in the plane of the E5 spacecraft trajectory. We expect to see both the largest particles and the highest fraction of salt-rich grains a few seconds after closest approach (CA) to Enceladus. Structures of the three most relevant localized supersonic jets for this projection (III, VI, VIII in ref. 22; Supplementary Fig. 2) are clearly visible in both the compositional profile and size contours. Note that the model only considers particle sizes above the instrument's detection threshold (0.2 µm). **b**, The modelled contribution from jets versus slow, diffuse sources and the fraction of salt-rich ice as a function of logarithmic altitude above the south pole. Near the surface salt-rich particles account for about 70% of all grains >0.2 µm, but owing to their larger average size they account for almost all the solid mass created by Enceladus' active region.

The composition and dynamics of the plume's solid phase inferred in this work and the low concentration of insoluble gases (N_2, CO, CH_4) in the gas phase[21,24,25] strongly imply that one or more salt-water reservoirs with large, potentially non-contiguous, evaporating surfaces[8] provide nearly all the matter forming the plume.

Received 18 January; accepted 5 May 2011.
Published online 22 June 2011.

1. Spahn, F. *et al.* Cassini dust measurements at Enceladus and implications for the origin of the E ring. *Science* **311,** 1416–1418 (2006).
2. Dougherty, M. K. *et al.* Identification of a dynamic atmosphere at Enceladus with the Cassini magnetometer. *Science* **311,** 1406–1409 (2006).
3. Hansen, C. J. *et al.* Enceladus' water vapor plume. *Science* **311,** 1422–1425 (2006).
4. Waite, J. H. *et al.* Cassini Ion and Neutral Mass Spectrometer: Enceladus plume composition and structure. *Science* **311,** 1419–1422 (2006).
5. Spencer, J. R. *et al.* Cassini encounters Enceladus: background and the discovery of a south polar hot spot. *Science* **311,** 1401–1405 (2006).
6. Porco, C. C. *et al.* Cassini observes the active south pole of Enceladus. *Science* **311,** 1393–1401 (2006).
7. Schmidt, J., Brillantov, N., Spahn, F. & Kempf, S. Slow dust in Enceladus' plume from condensation and wall collisions in tiger stripe fractures. *Nature* **451,** 685–688 (2008).
8. Postberg, F. *et al.* Sodium salts in E ring ice grains from an ocean below the surface of Enceladus. *Nature* **459,** 1098–1101 (2009).
9. Kieffer, S. W. *et al.* A clathrate reservoir hypothesis for Enceladus' south polar plume. *Science* **314,** 1764–1766 (2006).
10. Gioia, G., Chakraborty, P., Marshak, S. & Kieffer, W. Unified model of tectonics and heat transport in a frigid Enceladus. *Proc. Natl Acad. Sci. USA* **104,** 13578–13581 (2007).
11. Nimmo, F., Spencer, J. R., Pappalardo, R. T. & Mullen, M. E. Shear heating as the origin of the plumes and heat flux on Enceladus. *Nature* **447,** 289–291 (2007).
12. Kieffer, S. *et al.* A redetermination of the ice/vapor ratio of Enceladus' plumes: implications for sublimation and the lack of a liquid water reservoir. *Icarus* **203,** 238–241 (2009).
13. Schneider, N. M. *et al.* No sodium in the vapour plumes of Enceladus. *Nature* **459,** 1102–1104 (2009).
14. Srama, R. *et al.* The Cassini cosmic dust analyzer. *Space Sci. Rev.* **114,** 465–518 (2004).
15. Hillier, J. K. *et al.* The composition of Saturn's E ring. *Mon. Not. R. Astron. Soc.* **388,** 1588–1596 (2007).
16. Postberg, F. *et al.* The E-ring in the vicinity of Enceladus II. Probing the moon's interior — the composition of E-ring particles. *Icarus* **193,** 438–454 (2008).
17. Zolotov, M. Y. An oceanic composition on early and today's Enceladus. *Geophys. Res. Lett.* **34,** L23203 (2007).
18. Kempf, S. *et al.* Enceladus dust production — new insights from Cassini. *Am. Geophys. Union Fall Meet.* abstr. P33A–1562 (2010).
19. Hedman, M. *et al.* Spectral observations of the Enceladus plume with Cassini-VIMS. *Astrophys. J.* **693,** 1749–1762 (2009).
20. Sremcevic, M. *et al.* Impact-generated dust clouds around planetary satellites: asymmetry effects. *Planet. Space Sci.* **51** (issue 7–8), 455–471 (2003).
21. Hansen, C. J. *et al.* The composition and structure of the Enceladus plume. *Geophys. Res. Lett.* **38,** L11202 (2011).
22. Spitale, J. N. & Porco, C. C. Association of jets on Enceladus with the warmest regions on its south-polar fractures. *Nature* **449,** 695–697 (2007).
23. Halevy, I. & Stewart, S. T. Is Enceladus' plume tidally controlled? *Geophys. Res. Lett.* **35,** L12203, doi:10.1029/2008GL034349 (2008).
24. Waite, J. H. *et al.* Liquid water on Enceladus from observations of ammonia and ^{40}Ar in the plume. *Nature* **460,** 487–490 (2009).
25. Waite, J. H. Jr, Magee, B. & Brockwell, T. The effect of flyby velocity on the composition of the Enceladus gas torus as measured by Cassini INMS. *Proc. Lunar Planet. Sci. Conf.* **42,** 2818 (2011).
26. Tian, F., Stewart, A. I. F., Toon, O. B., Larsen, K. W. & Esposito, L. W. Monte Carlo simulations of water vapour plumes on Enceladus. *Icarus* **188,** 154–161 (2007).

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

Acknowledgements We thank G. Moragas-Klostermeyer and S. Helfert for their efforts in preparing the CDA plume observations. We acknowledge the work of the scientists and engineers of the Cassini Team at JPL. This work was supported by Deutsches Zentrum für Luft und Raumfahrt (DLR) and Deutsche Forschungs Gemeinschaft (DFG programme 'The first 10 million years of the solar system'). J.H. was supported by the UK Science and Technology Facilities Council.

Author Contributions F.P. led the analysis and write-up of the manuscript. J.S. led the dynamical modelling and contributed to data analysis and text. J.H. and S.K. contributed to data analysis and text. S.K. and R.S. designed the CDA plume measurement.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.P. (Frank.Postberg@mpi-hd.mpg.de).

LETTER

doi:10.1038/nature10230

A forty-kilodalton protein of the inner membrane is the mitochondrial calcium uniporter

Diego De Stefani[1,2]*, Anna Raffaello[1,2]*, Enrico Teardo[3], Ildikò Szabò[3] & Rosario Rizzuto[1,2]

Mitochondrial Ca^{2+} homeostasis has a key role in the regulation of aerobic metabolism and cell survival[1], but the molecular identity of the Ca^{2+} channel, the mitochondrial calcium uniporter[2], is still unknown. Here we have identified *in silico* a protein (named MCU) that shares tissue distribution with MICU1 (also known as CBARA1), a recently characterized uniporter regulator[3], is present in organisms in which mitochondrial Ca^{2+} uptake was demonstrated and whose sequence includes two transmembrane domains. Short interfering RNA (siRNA) silencing of MCU in HeLa cells markedly reduced mitochondrial Ca^{2+} uptake. MCU overexpression doubled the matrix Ca^{2+} concentration increase evoked by inositol 1,4,5-trisphosphate-generating agonists, thus significantly buffering the cytosolic elevation. The purified MCU protein showed channel activity in planar lipid bilayers, with electrophysiological properties and inhibitor sensitivity of the uniporter. A mutant MCU, in which two negatively charged residues of the putative pore-forming region were replaced, had no channel activity and reduced agonist-dependent matrix Ca^{2+} concentration transients when overexpressed in HeLa cells. Overall, these data demonstrate that the 40-kDa protein identified is the channel responsible for ruthenium-red-sensitive mitochondrial Ca^{2+} uptake, thus providing a molecular basis for this process of utmost physiological and pathological relevance.

The past two decades have seen the re-appraisal of the key role of mitochondria in decoding the highly pleiotropic Ca^{2+} signals evoked by physiological and pathological stimuli[4]. Mitochondria undergo rapid changes in matrix Ca^{2+} concentration ($[Ca^{2+}]_{mt}$) upon cell stimulation, because their low affinity uptake systems are exposed to microdomains of high $[Ca^{2+}]$ in proximity to endoplasmic reticulum or plasma membrane Ca^{2+} channels[5–9]. In turn, $[Ca^{2+}]_{mt}$ increases upregulate aerobic metabolism[10,11] and sensitize mitochondria to apoptotic challenges, favouring the release of caspase cofactors[12,13]. Mitochondrial Ca^{2+} buffering also shapes the amplitude and spatio-temporal patterns of cytosolic Ca^{2+} concentration ($[Ca^{2+}]_{cyt}$) increases[14–16]. The properties of the Ca^{2+} transporter are still largely those characterized in the 1960s: (1) electrogenic transport (hence the name calcium 'uniporter'), then shown to be a highly selective channel[17], (2) sensitivity to ruthenium red, and (3) low affinity for the cation[1,2,18]. Its molecular identity, however, has remained elusive.

Recently, a uniporter regulator (named MICU1), necessary for rapid mitochondrial Ca^{2+} uptake in intact cells, was identified *in silico* in the MitoCarta database[3]. We thus looked in the same database for the actual transporter, that is, an integral inner membrane protein fulfilling the criteria for being the bona fide MCU. Among 529 candidates with ubiquitous expression in mammalian tissues, we restricted the analysis to 89 proteins with two or more predicted transmembrane domains in the primary sequence. Of these, 20 were absent in *Saccharomyces cerevisiae* (lacking a RuR-sensitive mitochondrial Ca^{2+} uptake route[19]), and, among those, 14 were significantly conserved also in kinetoplastids (*Trypanosoma* and *Leishmania*), in which rapid, uncoupler-sensitive Ca^{2+} fluxes to mitochondria were measured[20,21]. This list (Supplementary Table 1) included 13 proteins with either already characterized functions (complex I subunits, assembly regulators, and so on) or poor alignment scores. The remaining protein (NP_001028431) includes a highly conserved domain comprising two transmembrane regions and an intervening loop enriched in acidic residues (Fig. 1a), indicative of a possible common role in cation permeation. This domain was also detected by hidden Markov model-based analysis of evolutionary conserved proteins (Pfam PF04678 and Panther PTHR1346 domains, with undefined function), further supporting the significance of the hit. We thus considered this protein a putative mitochondrial Ca^{2+} channel, named it MCU (mitochondrial calcium uniporter) and characterized it in detail. Tissue expression profiling by real-time PCR showed ubiquitous presence in all investigated tissues (Fig. 1b), in agreement with a housekeeping role, with expression levels correlating with those of MICU1.

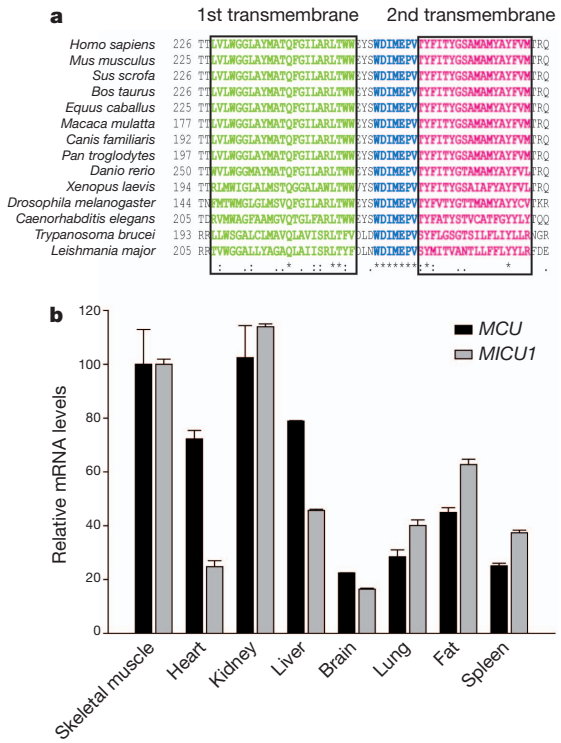
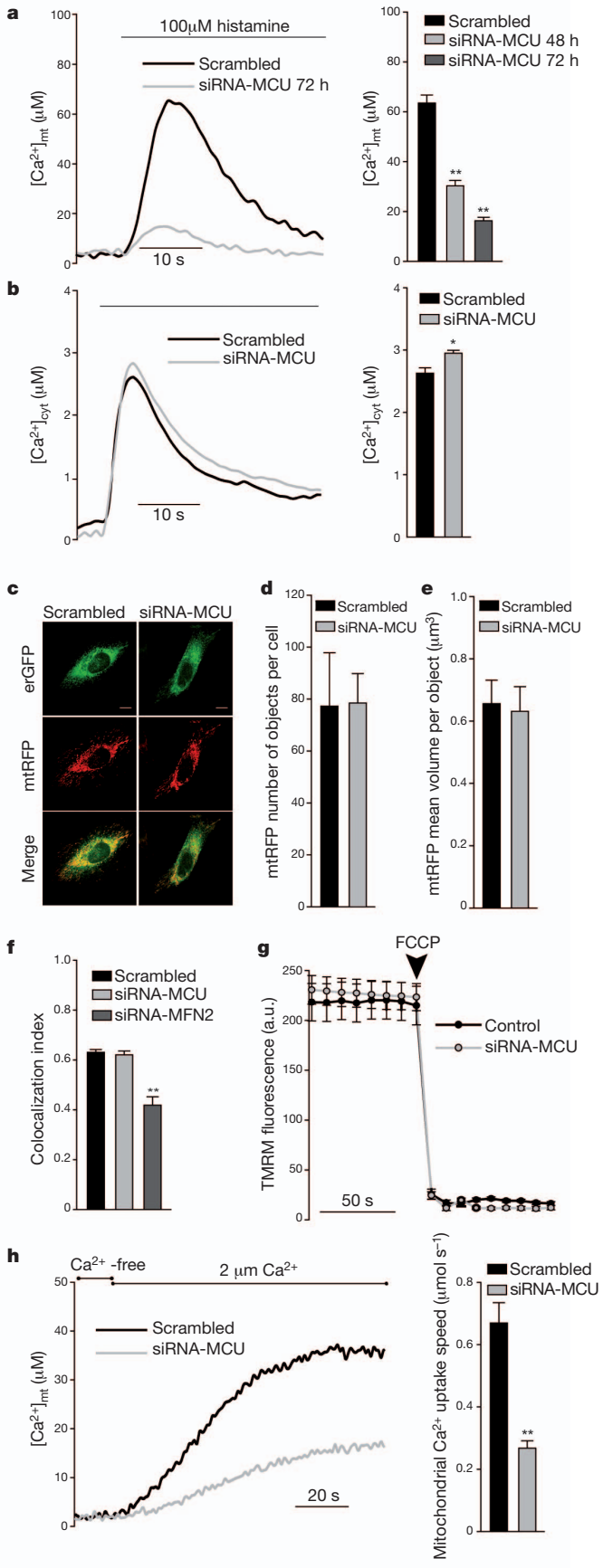


Figure 1 | MCU includes two highly conserved transmembrane domains and is ubiquitously expressed in mammals, similarly to its putative regulator MICU1. a, Alignment of the putative transmembrane domain and pore region of MCU proteins from 14 different species . **b**, Real-time PCR analysis of mouse tissues. mRNA extraction and real-time PCR was as described in the Methods section. Expression levels are normalized to skeletal muscle, and presented as means ± s.d. ($n = 3$).

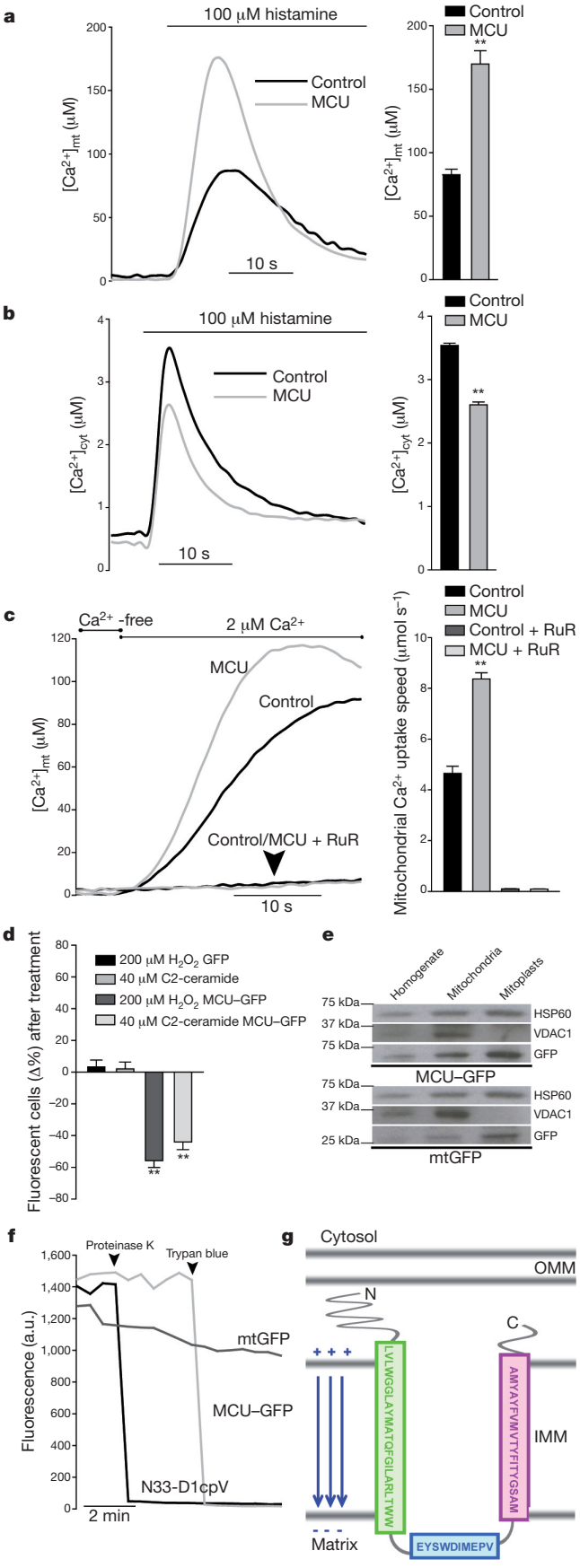
[1]Departments of Biomedical Sciences, University of Padua, 35121 Padua, Italy. [2]CNR Institute of Neuroscience, University of Padua, 35121 Padua, Italy. [3]Department of Biology, University of Padua, 35121 Padua, Italy.
*These authors contributed equally to this work



To verify the role of MCU in mitochondrial Ca^{2+} handling, siRNA oligonucleotides were synthesized and tested (Supplementary Fig. 1), Ca^{2+} measurements were then carried out with aequorin-based mitochondrial and cytosolic Ca^{2+} probes (mtAEQ and cytAEQ) (Fig. 2a and Supplementary Fig. 1). MCU-silenced and control HeLa cells, perfused with modified Krebs–Ringer buffer supplemented with 1 mM $CaCl_2$ (KRB), were challenged, where indicated, with 100 μM histamine, an inositol 1,4,5-trisphosphate-generating agonist causing Ca^{2+} release from the endoplasmic reticulum. In control cells, the $[Ca^{2+}]_{mt}$ rise evoked by histamine stimulation was ~63 μM (Fig. 2a). In MCU-silenced cells mitochondrial Ca^{2+} accumulation was markedly reduced (peak value ~30 μM and ~16 μM 48 h and 72 h after silencing, respectively). The marked reduction was specific of mitochondria, as $[Ca^{2+}]_{cyt}$ rises were almost unaffected (rather, a small but significant increase was detected, possibly due to reduced mitochondrial buffering) (Fig. 2b). Overexpression of MICU1 did not rescue the $[Ca^{2+}]_{mt}$ reduction in MCU-silenced cells, and had a marginal effect in control cells (Supplementary Fig. 2). Experiments were then carried out to exclude that the observed effect was secondary to changes in mitochondrial morphology (that is, fragmentation, swelling or redistribution away from the endoplasmic reticulum) or in the driving force for Ca^{2+} accumulation (that is, a collapse of mitochondrial membrane potential, $\Delta\Psi$). On the former aspect, mitochondria and endoplasmic reticulum were labelled in living cells by co-expressing red and green fluorescent protein specifically targeted to the mitochondria and endoplasmic reticulum compartments (mtRFP and erGFP; Fig. 2c and Supplementary Fig. 3). mtRFP imaging showed the typical three-dimensional interconnected network, with no difference in mitochondrial number and volume between control and MCU-silenced cells (Fig. 2d, e). Also the endoplasmic reticulum showed a similar distribution in control and MCU-silenced cells, and no difference was detected in the number of mtRFP/erGFP fluorescence overlaps, whereas in cells in which the endoplasmic reticulum/mitochondria tether mitofusin 2 (*MFN2*)[22] was silenced a marked reduction was observed (Fig. 2f and Supplementary Fig. 3). As to the driving force for Ca^{2+} uptake, no difference was detected in the loading of the fluorescent dye tetramethyl rhodamine methyl ester (TMRM) (Fig. 2g), thus ruling out a $\Delta\Psi$ drop in MCU-silenced cells. In brief, mitochondria simply seem to have an intrinsically reduced ability of accumulating Ca^{2+}. Direct evidence was obtained by permeabilizing MCU-silenced and control cells in intracellular buffer, containing 100 μM EGTA (Ca^{2+}-free) and then imposing an EGTA-buffered fixed $[Ca^{2+}]$ of 2 μM (2 μM Ca^{2+}). Under those conditions, in MCU-silenced cells Ca^{2+} uptake was initiated at a speed that was threefold lower than controls (Fig. 2h).

To demonstrate that MCU promotes Ca^{2+} uptake, overexpression experiments were carried out. The full-length complementary DNA of the protein, obtained by PCR after reverse transcription from skeletal muscle RNA, was identical to that deposited in the NCBI databank (NM_001033259). The cDNA was cloned in pcDNA3.1 and co-expressed with aequorin probes. MCU-overexpressing cells showed a marked increase in the histamine-evoked $[Ca^{2+}]_{mt}$ rise (106%, ~169 μM in MCU-overexpressing vs ~82 μM in control cells, Fig. 3a). The greater mitochondrial response was not secondary to

Figure 2 | MCU silencing strongly inhibits mitochondrial Ca^{2+} uptake without causing morphological rearrangement or changes in the electrochemical gradient. **a**, **b**, Measurement of $[Ca^{2+}]_{mt}$ (**a**) and $[Ca^{2+}]_{cyt}$ (**b**) in MCU-silenced cells. **c**, Fluorescence images of mtRFP- and erGFP-labelled mitochondria and endoplasmic reticulum, respectively. **d**, **e**, Mitochondrial number and volume, as deduced by calculating object number (**d**) and size (**e**). **f**, endoplasmic reticulum/mitochondria colocalization, estimated by Pearson's correlation coefficient. **g**, TMRM fluorescence measurements. a.u., arbitrary units. **h**, $[Ca^{2+}]_{mt}$ measurements in permeabilized cells. In this and following figures, experiments are representative of more than five trials, conditions are in the Methods section, and statistics in Supplementary Table 2. $^*P < 0.05$, $^{**}P < 0.001$. Error bars correspond to mean ± s.e.m.



alterations of the cytosolic response. Rather, a significant reduction in the $[Ca^{2+}]_{cyt}$ transient was observed, most probably because of increased Ca^{2+} clearance by mitochondria (Fig. 3b). Enhanced mitochondrial Ca^{2+} uptake was confirmed also upon capacitative Ca^{2+} influx (Supplementary Fig. 4) and in permeabilized cells (Fig. 3c). As in Fig. 2h, mitochondrial Ca^{2+} uptake was initiated by switching the medium, after digitonin permeabilization, from intracellular buffer/EGTA (Ca^{2+}-free) to intracellular buffer/Ca^{2+} (3 μM Ca^{2+}). MCU-overexpressing cells showed a faster rate of Ca^{2+} uptake and reached a higher plateau level (Fig. 3c). Finally, we verified whether the increased intrinsic ability of mitochondria to accumulate Ca^{2+} correlated with sensitivity to apoptotic challenges (Fig. 3d). Microscopy counts of cell viability after treatment with C2-ceramide or H_2O_2 showed that MCU-expressing cells were more efficiently killed, thus confirming the notion that mitochondrial Ca^{2+} loading synergizes with these apoptotic stimuli[12,23].

We then investigated the subcellular distribution of MCU. For this purpose, tagged versions of MCU were generated (MCU–GFP and MCU–Flag), that had the same effect on $[Ca^{2+}]_{mt}$ as MCU (Supplementary Fig. 5). GFP fluorescence and immunocytochemistry labelling with anti-Flag antibody completely overlapped with mitochondrial markers (Supplementary Fig. 6). Then, intra-mitochondrial distribution and topology was investigated. HeLa cells were transfected with MCU–GFP (or the matrix probe mtGFP) and harvested. The mitochondrial fraction was obtained by differential centrifugation and either blotted directly, or subjected to osmotic swelling, thus obtaining the mitoplast fraction devoid of the outer membrane. Both MCU and mtGFP were progressively more enriched in the mitochondrial and mitoplast fractions (Fig. 3e). Mitochondrial localization was also confirmed by immuno-electron microscopy of HeLa cells expressing MCU–Flag (Supplementary Fig. 7). To show the inner membrane localization and get information on the topology of MCU, we carried out an experiment in digitonin-permeabilized cells, in which GFP fluorescence was visualized before and after treatment with proteinase K or Trypan blue, a fluorescence quencher that crosses the outer, but not the inner mitochondrial membrane[9] (Fig. 3f). Proteinase K had no effect on MCU–GFP fluorescence (but abrogated fluorescence of the cytosolic, outer mitochondrial membrane-anchored N33-D1cpV probe, confirming that MCU is not on the outer mitochondrial membrane), whereas Trypan blue abrogated MCU–GFP (but not mtGFP) fluorescence (Fig. 3f), thus demonstrating that MCU is in the inner mitochondrial membrane, with the carboxy terminus most probably located in the intermembrane space (Fig. 3g).

Finally, we investigated the channel activity of the purified protein reconstituted in a planar lipid bilayer. For this purpose, a His-tagged MCU was generated and expressed in two different heterologous expression systems, *Escherichia coli* and wheat-germ cell-free transcription/translation. The purified protein (Supplementary Fig. 8) was inserted into lipid bilayers, and its electrophysiological activity was assessed in a medium containing only Ca^{2+} as cation. With both preparations, we measured channel activity, with the properties previously reported for the uniporter[17,24]: conductance of 6–7 pS (in the

Figure 3 | MCU overexpression increases mitochondrial Ca^{2+} accumulation in intact and permeabilized cells, buffers cytosolic $[Ca^{2+}]$ rises, and sensitizes to apoptotic stimuli; GFP-tagged MCU demonstrates mitochondrial localization and indicates a putative membrane topology. **a**, **b**, Measurement of $[Ca^{2+}]_{mt}$ (**a**) and $[Ca^{2+}]_{cyt}$ (**b**) in MCU-silenced cells. **c**, $[Ca^{2+}]_{mt}$ measurements in permeabilized cells. RuR, ruthenium red. **d**, Cell viability upon apoptotic challenge. **e**, Subcellular fractionation of MCU–GFP- and mtGFP-expressing cells and western blotting for GFP, matrix (HSP60) and outer mitochondrial membrane (VDAC1) markers. **f**, GFP proteinase K degradation and fluorescence quenching by Trypan blue in permeabilized cells transfected with mtGFP, MCU–GFP or N33-D1cpV. **g**, Schematic representation of the predicted MCU topology. IMM, inner mitochondrial membrane; OMM, outer mitochondrial membrane. $^{**}P < 0.001$. Error bars: **a**–**c**, mean ± s.e.m.; **d**, **f**, mean ± s.d.

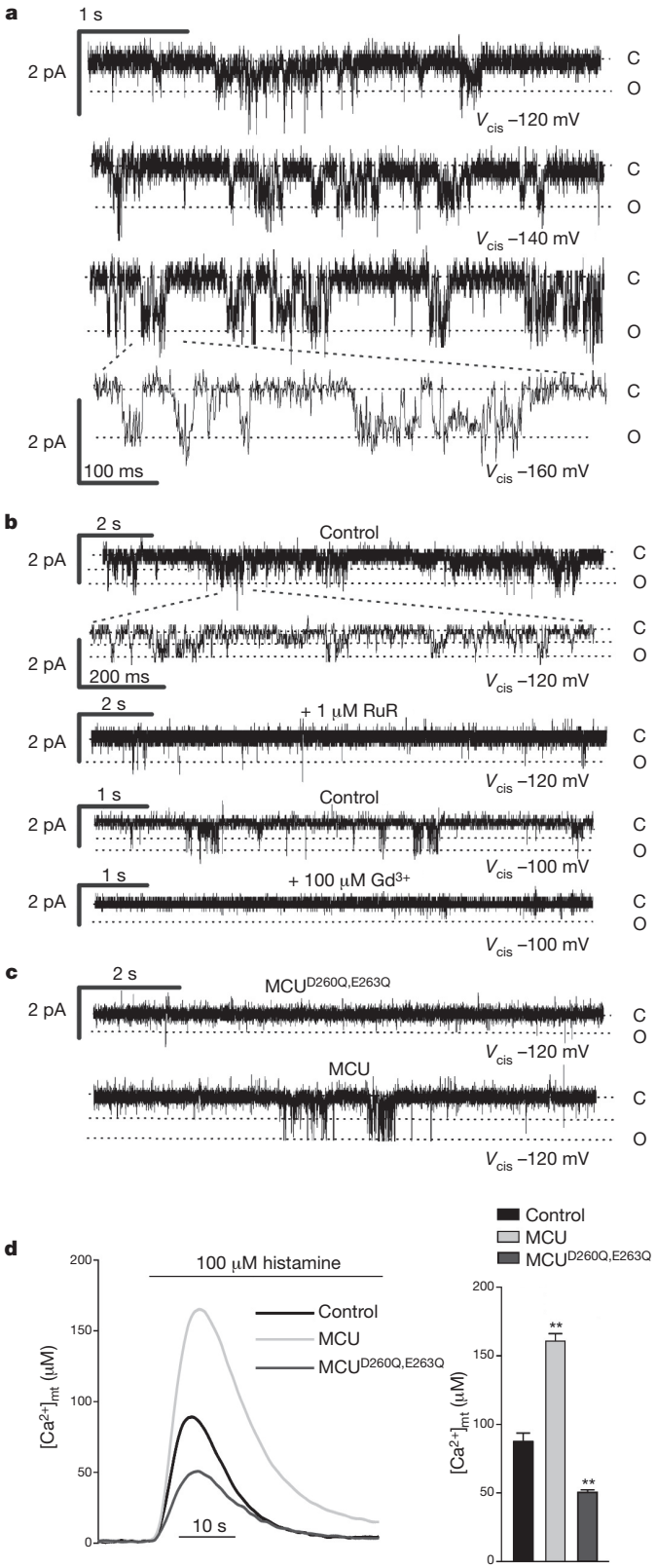


Figure 4 | Purified MCU shows channel activity in lipid bilayers, whereas $MCU^{D260Q,E263Q}$ shows no channel activity and reduces $[Ca^{2+}]_{mt}$ transients in cells. **a**–**c**, Electrophysiological recordings of purified MCU or $MCU^{D260Q,E263Q}$, produced in *E. coli* (**a** and **b**) or *in vitro* (**c**) and reconstituted in planar lipid bilayers. C, closed; O, opened; V_{cis}, voltage applied to the *cis* side. **a**, MCU traces at different applied voltages; **b**, MCU traces before and after addition of RuR or Gd^{3+}; **c**, $MCU^{D260Q,E263Q}$ traces. After 10 min recording with no current, MCU was added and channel activity was detected after 1–2 min (lower trace). **d**, $[Ca^{2+}]_{mt}$ measurements in $MCU^{D260Q,E263Q}$-expressing cells. Error bars correspond to mean ± s.e.m.

negative voltage range applied in the *cis* compartment), fast kinetics, low probability of opening at low voltages (P_o), P_o increase with voltage, and inhibition by ruthenium red and the lanthanide gadolinium (Gd^{3+}) (Fig. 4a and b and Supplementary Fig. 9). Then, to get evidence of cation permeation across the putative pore-forming domain, we mutated into glutamines two negatively charged residues of the conserved region (D260Q and E263Q) and assessed the electrophysiological properties of the mutant protein ($MCU^{D260Q,E263Q}$) produced *in vitro*. $MCU^{D260Q,E263Q}$ failed to give rise to Ca^{2+}-permeable channel activity in bilayer experiments, whereas in the same membrane the subsequent addition of MCU initiated channel activity with the characteristics described above (Fig. 4c). $MCU^{D260Q,E263Q}$, when overexpressed in HeLa cells, showed the same membrane topology as MCU (Supplementary Fig. 10) and caused a marked reduction in histamine-induced $[Ca^{2+}]_{mt}$ rises, both compared to MCU-overexpressing and control cells (Fig. 4d). $MCU^{D260Q,E263Q}$ thus acts as a dominant-negative mutant, most probably by either inhibiting MCU channel activity within an oligomer or competing for docking sites or essential regulators.

In conclusion, we have identified a 40-kDa protein which fulfils the requirements for being the long searched for mitochondrial calcium uniporter: it contains two transmembrane domains and shows channel activity *in vitro* with the previously characterized properties of the uniporter, it is localized in the inner membrane and, finally, it markedly enhances mitochondrial Ca^{2+} uptake upon overexpression. The availability of molecular information opens the mitochondrial checkpoint in cellular Ca^{2+} signalling to a deeper understanding. Informative animal models can now be generated and new drugs developed to influence processes regulated by mitochondrial Ca^{2+} signals, such as aerobic metabolism and cell death.

METHODS SUMMARY

Candidate mouse genes coding for the mitochondrial calcium uniporter were searched starting from the MitoCarta database. Bioinformatic analysis was performed by selecting candidates according to the following parameters: (1) genes expressed in most mammalian tissues, (2) the presence of two or more transmembrane domains, (3) without any homologues in *Saccharomyces cerevisiae* but (4) conserved in kinetoplastids. The resulting 14 candidates were manually screened to search for highly conserved putative channel-forming domains.

All the experiments were performed in HeLa cells. MCU was silenced using specific siRNAs. The MCU cDNA was retro-transcribed and amplified from mouse skeletal muscle messenger RNA and overexpressed after cloning in several expression plasmids.

$[Ca^{2+}]_{cyt}$ and $[Ca^{2+}]_{mt}$ were measured in MCU-silenced and MCU-overexpressing cells using appropriate cytosolic and mitochondrial probes[25].

Morphology of MCU-silenced cells was evaluated using GFP and RFP specifically targeted to the mitochondria and endoplasmic reticulum compartments (erGFP and mtRFP)[6].

MCU mitochondrial localization was demonstrated by confocal imaging of cells transfected with the MCU–GFP expression vector and loaded with MitoTracker Red and by fractionation and western blotting of cells overexpressing MCU–GFP, as previously described[26].

Reconstitution in planar lipid bilayer was carried out with the purified protein from both *E. coli* and *in vitro* transcription/translation using 100 mM $CaCl_2$. Bilayers of approximately 150–200 pF capacity were prepared using purified soybean azolectin.

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

Received 29 December 2010; accepted 27 May 2011.
Published online 19 June 2011.

1. Szabadkai, G. & Duchen, M. R. Mitochondria: the hub of cellular Ca^{2+} signaling. *Physiology (Bethesda)* **23,** 84–94 (2008).
2. Carafoli, E. Historical review: mitochondria and calcium: ups and downs of an unusual relationship. *Trends Biochem. Sci.* **28,** 175–181 (2003).

3. Perocchi, F. *et al. MICU1* encodes a mitochondrial EF hand protein required for Ca^{2+} uptake. *Nature* **467,** 291–296 (2010).
4. Berridge, M. J., Bootman, M. D. & Roderick, H. L. Calcium signalling: dynamics, homeostasis and remodelling. *Nature Rev. Mol. Cell Biol.* **4,** 517–529 (2003).
5. Rizzuto, R., Brini, M., Murgia, M. & Pozzan, T. Microdomains with high Ca^{2+} close to IP3-sensitive channels that are sensed by neighboring mitochondria. *Science* **262,** 744–747 (1993).
6. Rizzuto, R. *et al.* Close contacts with the endoplasmic reticulum as determinants of mitochondrial Ca^{2+} responses. *Science* **280,** 1763–1766 (1998).
7. Csordás, G., Thomas, A. P. & Hajnóczky, G. Quasi-synaptic calcium signal transmission between endoplasmic reticulum and mitochondria. *EMBO J.* **18,** 96–108 (1999).
8. Csordás, G. *et al.* Imaging interorganelle contacts and local calcium dynamics at the ER-mitochondrial interface. *Mol. Cell* **39,** 121–132 (2010).
9. Giacomello, M. *et al.* Ca^{2+} hot spots on the mitochondrial surface are generated by Ca^{2+} mobilization from stores, but not by activation of store-operated Ca^{2+} channels. *Mol. Cell* **38,** 280–290 (2010).
10. Hajnóczky, G., Robb-Gaspers, L. D., Seitz, M. B. & Thomas, A. P. Decoding of cytosolic calcium oscillations in the mitochondria. *Cell* **82,** 415–424 (1995).
11. Jouaville, L. S., Pinton, P., Bastianutto, C., Rutter, G. A. & Rizzuto, R. Regulation of mitochondrial ATP synthesis by calcium: evidence for a long-term metabolic priming. *Proc. Natl Acad. Sci. USA* **96,** 13807–13812 (1999).
12. Pinton, P. *et al.* The Ca^{2+} concentration of the endoplasmic reticulum is a key determinant of ceramide-induced apoptosis: significance for the molecular mechanism of Bcl-2 action. *EMBO J.* **20,** 2690–2701 (2001).
13. Pacher, P. & Hajnóczky, G. Propagation of the apoptotic signal by mitochondrial waves. *EMBO J.* **20,** 4107–4121 (2001).
14. Hajnóczky, G., Hager, R. & Thomas, A. P. Mitochondria suppress local feedback activation of inositol 1,4,5-trisphosphate receptors by Ca^{2+}. *J. Biol. Chem.* **274,** 14157–14162 (1999).
15. Boitier, E., Rea, R. & Duchen, M. R. Mitochondria exert a negative feedback on the propagation of intracellular Ca^{2+} waves in rat cortical astrocytes. *J. Cell Biol.* **145,** 795–808 (1999).
16. Tinel, H. *et al.* Active mitochondria surrounding the pancreatic acinar granule region prevent spreading of inositol trisphosphate-evoked local cytosolic Ca^{2+} signals. *EMBO J.* **18,** 4999–5008 (1999).
17. Kirichok, Y., Krapivinsky, G. & Clapham, D. E. The mitochondrial calcium uniporter is a highly selective ion channel. *Nature* **427,** 360–364 (2004).
18. Nicholls, D. G. Mitochondria and calcium signaling. *Cell Calcium* **38,** 311–317 (2005).
19. Uribe, S., Rangel, P. & Pardo, J. P. Interactions of calcium with yeast mitochondria. *Cell Calcium* **13,** 211–217 (1992).
20. Xiong, Z. H., Ridgley, E. L., Enis, D., Olness, F. & Ruben, L. Selective transfer of calcium from an acidic compartment to the mitochondrion of *Trypanosoma brucei*. Measurements with targeted aequorins. *J. Biol. Chem.* **272,** 31022–31028 (1997).
21. Benaim, G., Bermudez, R. & Urbina, J. A. Ca^{2+} transport in isolated mitochondrial vesicles from *Leishmania braziliensis* promastigotes. *Mol. Biochem. Parasitol.* **39,** 61–68 (1990).
22. de Brito, O. M. & Scorrano, L. Mitofusin 2 tethers endoplasmic reticulum to mitochondria. *Nature* **456,** 605–610 (2008).
23. Hajnóczky, G., Csordás, G., Madesh, M. & Pacher, P. Control of apoptosis by IP_3 and ryanodine receptor driven calcium signals. *Cell Calcium* **28,** 349–363 (2000).
24. Michels, G. *et al.* Regulation of the human cardiac mitochondrial Ca^{2+} uptake by 2 different voltage-gated Ca^{2+} channels. *Circulation* **119,** 2435–2443 (2009).
25. Pinton, P., Rimessi, A., Romagnoli, A., Prandini, A. & Rizzuto, R. Biosensors for the detection of calcium and pH. *Methods Cell Biol.* **80,** 297–325 (2007).
26. Wieckowski, M. R., Giorgi, C., Lebiedzinska, M., Duszynski, J. & Pinton, P. Isolation of mitochondria-associated membranes and mitochondria from animal tissues and cells. *Nature Protocols* **4,** 1582–1590 (2009).

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

Acknowledgements We thank M. Zoratti for help in the analysis of electrophysiological data, G. Merli for carrying out some of the experiments and P. Bernardi, T. Pozzan and L. Cendron for helpful discussions and for the N33-D1cpV expression plasmid (T. Pozzan). This research was supported by grants from the Italian Ministry of Education, University and Research, European Commission (FP7 ''MyoAGE'', no. 223576), National Institutes of Health (Grant #1P01AG025532-01A1), Cariparo Foundation (Padua), the Italian Association for Cancer Research (AIRC) and Telethon-Italy (GPP1005).

Author Contributions D.D.S. performed bioinformatic analysis, Ca^{2+} measurements and morphological analysis of organelles. A.R. performed molecular biology and gene expression analysis. D.D.S. and A.R. contributed equally to the study. E.T. expressed and purified the protein in heterologous systems; I.S. performed and analysed the electrophysiology experiments; R.R. discussed the results and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.R. (rosario.rizzuto@unipd.it).

METHODS

Bioinformatic screening. The MitoCarta database was screened for ubiquitously expressed proteins (expressed in at least 12 out of 14 tissues analysed, obtaining 529 candidates) and proteins containing two or more transmembrane domains were selected using the TMHMM algorithm[27]. Out of 1,098 proteins present in MitoCarta, 89 fulfilled these requirements. Sequences of these candidates were retrieved and aligned with the *Saccharomyces cerevisiae* proteome by a standard Gonnet matrix to exclude homologues. All alignments with a P-value < 0.001 were excluded. Out of 89, 20 remained and were subsequently aligned with kinetoplastids proteomes using the TriTrypDB server. Out of 20, 14 had significant scores. These candidates were manually aligned through different species to look for putative highly conserved channel-forming domains.

RNA extraction and gene expression analyses. For the expression analysis of MCU and MICU1 in mouse tissues, adult male C57BL/6 mice (28–30 g) were used. Skeletal muscles (tibialis anterior), heart, brain, spleen, lung, liver, kidney and visceral fat were excised from three age-matched animals. Total RNA was purified using an SV Total RNA Isolation kit (Promega) following manufacturer instructions. The RNA was quantified and controlled for its quality using a RNA 6000 LabChip kit (Agilent Technologies) in conjunction with an Agilent Bioanalyzer 2001. An equal amount of total RNA from the three animals was pooled together for each organ, complementary DNA was generated with a cDNA synthesis kit (SuperScript VILO, Invitrogen) and analysed by real-time PCR using the SYBR green chemistry (Bio-Rad). The primers were designed and analysed with Primer3. Identity of the amplicons was confirmed by their dissociation profiles and gel analysis. Real-time PCR standard curves were constructed by using serial dilutions of pooled cDNAs of the analysed samples, using at least four dilution points and the efficiency of all primer sets was between 95 and 105%. Real-time PCR primer sequences were as follows: MICU1-fw, 5′-GTCGAACTCTCGGACCATGT-3′; MICU1-rv, 5′-CAAAGTCCC AGGCAGTTTCT-3′. These primers amplify a fragment of 199 base pairs. MCU-fw, 5′-AAAGGAGCCAAAAAGTCACG-3′; MCU-rv, 5′-AACGGCGTGAGTTAC AAACA-3′. These primers amplify a fragment of 200 bp.

For the expression analysis of MCU silencing, total RNA was purified from HeLa cells transfected with scrambled, siRNA-MCU#1 and siRNA-MCU#2 for 48 h following the standard TRIzol protocol. The RNA was quantified, controlled for its quality and retrotranscribed as described earlier. Complementary DNA was analysed by real-time PCR using the SYBR Green chemistry (Bio-Rad). All data were normalized to *GAPDH* expression. The oligonucleotide primers specific for MCU are the same as above. The oligonucleotide primers specific for GAPDH are the following: GADPH-fw, 5′-CACCATCTTCCAGGAGCGAG-3′; GADPH-rv, 5′-CCTTCTC CATGGTGGTGAAGAC-3′. These primers amplify a fragment of 101 bp.

Constructs and siRNA. Mouse MCU (NM_001033259) was amplified from mouse skeletal muscle cDNA by PCR using the following primers.

For cloning into pEGFP-N1: fw, 5′-GAATTCGCCACCATGGCGGCCGCCG CAGGTAG-3′; rv, 5′-GGATCCACTTCCTTTTCTCCGATCTGTCG-3′. The PCR fragment was cloned into EcoRI and BamHI sites in pEGFP-N1 (Clontech).

For cloning into pcDNA3.1: fw, 5′-GGTACCGCCACCATGGCGGCCGCC GCAGGTAG-3′; rv, 5′-GAATTCTCATTCCTTTTCTCCGATCTGTC-3′. The PCR fragment was cloned into KpnI and EcoRI sites in pcDNA3.1 (Invitrogen).

For the cloning of MCU–Flag into pcDNA3.1: fw, 5′-GGTACCGCCACCA TGGCGGCCGCCGCAGGTAG-3′; rv, 5′-GGAATTCTCACTTATCGTCGTC ATCCTTGTAATCTTCCTTTTCTCCGATCTGTC-3′. The PCR fragment was cloned into KpnI and EcoRI sites in pcDNA3.1 (Invitrogen).

For the cloning of MCU in pET-28A(+): fw, 5′-AGGATCCATGGCGGCCG CCGCAGGTAG-3′; rv, 5′-ACTCGAGTCATTCCTTTTCTCCGATCT-3′. The PCR fragment was cloned into BamHI and XhoI sites in pET-28A(+) (Novagen).

For the cloning of MCU deleted of the mitochondrial targeting sequence (amino acids 1–54) ($MCU^{\Delta MTS}$) in pET-28A(+): fw, 5′-CATATGGCTTCCTGGC AGAGCGTGGG-3′; rv, 5′-CTCGAGTCATTCCTTTTCTCCGATCT-3′. The PCR fragment was cloned into NdeI and XhoI sites in pET-28A(+) (Novagen).

For the cloning of MCU in pIVEX 1.3 WG: fw, 5′-ACATATGGCGGCCGC CGCAGGTAGATC-3′; rv, 5′-TCTCGAGTTCCTTTTCTCCGATCTGTC-3′. The PCR fragment was cloned into NdeI and XhoI sites in pIVEX 1.3 WG (Roche).

The generation of the pcDNA3.1-$MCU^{D260Q,E263Q}$-Flag and pEGFP-N1-$MCU^{D260Q,E263Q}$ mutant was performed by mutagenesis PCR using the wild type pcDNA3.1-MCU-Flag and pEGFP-N1-MCU vectors as template and the mutagenesis primer: 5′-CTGGTGGGAGTACTCGTGG**C**A**A**ATCATG**C**A**A**CCCGT CACCTACTTCATCAC-3′.

Mouse MICU1 (NM_144822) was amplified from mouse skeletal muscle cDNA by PCR using the following primers:

For the cloning of MICU1-HA in pcDNA3.1: fw, 5′-CGGATCCGCCACCAT GTTTCGTCTTAACACCCT-3′; rv, 3′-GCTCGAGTCACAGGGAAGCGTAGT CAGGCACATCGTAGGGGTATTTGGGCAGAGCAAAGTCCC-5′. The PCR fragment was cloned into BamHI and XhoI sites in pcDNA3.1 (Invitrogen).

For the cloning of MICU1 in pEGFP-N1: fw, 5′-CCTCGAGATGTTTCGTC TTAACACCCT-3′; rv, 5′-CGGATCCCGTTTGGGCAGAGCAAAGTCCC-3′. The PCR fragment was cloned into XhoI and BamHI sites in pEGFP-N1 (Clontech).

To silence MCU specific siRNA were designed: siRNA-MCU#1: nucleotides 899–917 of the corresponding mRNA (5′-GCCAGAGACAGACAAUACUtt-3′ and 3′-ttCGGUCUCUGUCUGUUAUGA-5′. siRNA-MCU#2: nucleotides 360–378 of the corresponding mRNA (5′-GGGAAUUGACAGAGUUGCUtt-3′and 3′-ttCCCUUAACUGUCUCAACGA-5′). The non-targeting siRNA (scrambled) is the following: 5′-GCCUAAGAACGACAAAUCAtt-3′ and 3′-ttCGGAUUCU UGCUGUUUAGU-5′.

Cell culture and transfection. In all the experiments HeLa cells were used. Cells were grown in Dulbecco's modified Eagle's medium (DMEM) (Euroclone), supplemented with 10% fetal bovine serum (FBS) (Euroclone) and transfected with a standard calcium-phosphate procedure. For aequorin measurements, the cells were seeded 24 h before transfection onto 13 mm glass coverslips and allowed to grow to 50% confluence before transfection. For morphologic analyses cells were seeded 24 h before transfection onto 24-mm glass coverslips and allowed to grow to 50% confluence before transfection, unless otherwise specified.

Aequorin measurements. HeLa cells grown on 13-mm round glass coverslips at 50% confluence were transfected with the cytosolic (cytAEQ) or the low-affinity mitochondrial (mtAEQmut, referred in the text as mtAEQ) probe (as previously described[25]) together with the indicated siRNA or plasmid. pcDNA3.1 was used as control unless otherwise indicated. The coverslip with the cells was incubated with 5 μM coelenterazine for 1–2 h in KRB (Krebs–Ringer modified buffer: 125 mM NaCl, 5 mM KCl, 1 mM Na_3PO_4, 1 mM $MgSO_4$, 5.5 mM glucose, 20 mM HEPES, pH 7.4, 37 °C) supplemented with 1 mM $CaCl_2$, and then transferred to the perfusion chamber. All aequorin measurements were carried out in KRB. Agonists and other drugs were added to the same medium, as specified in the text. The experiments were terminated by lysing the cells with 100 μM digitonin in a hypotonic Ca^{2+}-rich solution (10 mM $CaCl_2$ in H_2O), thus discharging the remaining aequorin pool. The light signal was collected and calibrated into $[Ca^{2+}]$ values by an algorithm based on the Ca^{2+} response curve of aequorin at physiological conditions of pH, $[Mg^{2+}]$ and ionic strength, as previously described[25]. Representative traces are shown in the figures whereas the full data set is included in Supplementary Table 2. In the experiments with permeabilized cells, a buffer mimicking the cytosolic ionic composition, (intracellular buffer) was used: 130 mM KCl, 10 mM NaCl, 2 mM K_2HPO_4, 5 mM succinic acid, 5 mM malic acid, 1 mM $MgCl_2$, 20 mM HEPES, 1 mM pyruvate, 0.5 mM ATP and 0.1 mM ADP (pH 7 at 37 °C). Intracellular buffer was supplemented with either 100 μM EGTA (intracellular buffer/EGTA) or a 2 mM EGTA and 2 mM HEEDTA-buffered $[Ca^{2+}]$ of 2 or 3 μM (intracellular buffer/Ca^{2+}), calculated with the Chelator software[28]. HeLa cells were permeabilized by a 1-min perfusion with 50 μM digitonin (added to intracellular buffer/EGTA) during luminescence measurements. Mitochondrial Ca^{2+} uptake speed was calculated as the first derivative by using the SLOPE Excel function (Microsoft) and smoothed for three time points. The higher value reached during Ca^{2+} addition represents the maximal Ca^{2+} uptake speed. $[Ca^{2+}]_{mt}$ following capacitative Ca^{2+} influx was measured by perfusing HeLa cells with the SERCA blocker cyclopiazonic acid (CPA, 20 μM) in a KRB solution containing no Ca^{2+} and 100 μM EGTA. In this protocol, mitochondrial Ca^{2+} uptake were evoked by adding 2 mM $CaCl_2$ to the medium. All of the results are expressed as means ± s.e.m., and Student's t-test was used for the statistics. All the materials were from Sigma Aldrich unless specified.

Mitochondrial and endoplasmic reticulum morphology analysis. Images for investigating mitochondrial morphology and endoplasmic reticulum/mitochondrial contact sites were taken on a Leica TCS-SP5-II equipped with a PlanApo 100×/1.4 numerical aperture objective. For all images, pinhole was set to 1 Airy unit, pixel size was 75 nm and a Z-stack was acquired for the whole depth of the cell by sampling at 130 nm in the Z plane. An argon laser line (488 nm) was used to excite GFP or AlexaFluor488 and the signal was collected in the 492–537 nm range, whereas RFP (and MitoTracker Red and AlexaFluor546) fluorescence was excited by the 543 nm HeNe laser and its emission was collected in the 555–700 nm range. For each image, PMT gain was slightly adjusted in order to maximize signal and avoid saturation. Images were then all analysed with the Fiji image processing package based on ImageJ. Mitochondrial morphology was calculated with the 3D Object Counter plugin, and colocalization was evaluated with the JaCoP plugin[29]. The colocalization index is represented by Pearson's coefficient calculated following Costes randomization (100 cycles) and automatic threshold calculation[30].

Immunofluorescence. HeLa cells were grown on 24-mm coverslips and transfected with MCU–Flag encoding plasmid when 50% confluent. After 24 h, cells were washed with PBS, fixed in 4% formaldehyde for 10 min and quenched with 50 mM NH_4Cl in PBS. Cells were permeabilized for 10 min with 0.1% Triton X-100 in PBS and blocked in PBS containing 2% BSA and 0.05% Triton X-100

for 1 h. Cells were then incubated with primary antibodies (anti-HSP60 and anti-Flag) for 3 h at room temperature and washed three times with 0.1% Triton X-100 in PBS. The appropriate isotype-matched, AlexaFluor-conjugated secondary antibodies (Invitrogen) were used and coverslips were mounted with ProLong Gold Antifade reagent (Invitrogen). Images were taken on a Leica SP5-II as detailed earlier.

Fractionation and western blotting. HeLa cells (10^8) were homogenized with a Dounce homogenizer (100 strokes) and centrifuged at 600*g* to remove entire cells. The supernatant was then centrifuged at 8,000*g* to pellet the crude mitochondrial fraction. Mitoplasts were obtained by osmotic swelling by incubating the mitochondrial fraction in 20 mM HEPES for 20 min. Proteins were quantified using the BCA Protein Assay Kit (Pierce) following the manufacturer instructions. Proteins (10 μg) were separated by SDS–PAGE, transferred onto nitrocellulose membranes (GE Healthcare) and probed using the following antibodies: anti-VDAC1(Abnova), anti-HSP60 (Santa Cruz) and anti-GFP (Abcam). Isotype-matched, horseradish-peroxidase-conjugated secondary antibodies (Bio-Rad) were used, followed by detection by chemiluminescence (GE Healthcare).

To test the siRNAs efficacy, HeLa cells were grown on 10-cm Petri dishes and transfected when 30% confluent. MCU–Flag encoding plasmid (8 μg) was co-transfected with scrambled, siRNA-MCU#1 or siRNA-MCU#2. After 72 h cells were washed, harvested and lysed in RIPA buffer (150 mM NaCl, 25 mM TRIS, 1% Triton-X100, 0.5% Na-deoxycholate and 0.1% SDS, pH 8) added with proteases and phosphatases inhibitor cocktails (Roche). Proteins were quantified by the bicinchoninic acid assay (BCA) method and 25 μg of each sample were loaded on a Novex NuPage Bis-Tris 4–12% precast gel (Invitrogen), transferred onto nitrocellulose membranes and probed with anti-β-tubulin (Sigma) and anti-Flag (Sigma) antibodies. Isotype-matched, horseradish-peroxidase-conjugated secondary antibodies (Bio-Rad) were used, followed by detection by chemiluminescence (GE Healthcare).

Measurements of mitochondrial $\Delta\Psi$. Mitochondrial $\Delta\Psi$ was measured by loading cells with 20 nM tetramethyl rhodamine methyl ester (TMRM, Invitrogen) for 30 min at 37 °C. Images were taken on a inverted microscope (Zeiss Axiovert 100) equipped with a PlanFluar 40×/1.3 numerical aperture objective, a Photometrics Coolsnap ES and a LED-based illumination system (Cairn OptoLed II). TMRM excitation was performed at 560 nm and emission was collected through a 590–650 nm band-pass filter. Images were taken every 10 s with a fixed 200 ms exposure time. FCCP (carbonyl cyanide p-trifluoromethoxyphenylhydrazone, 10 μM), an uncoupler of oxidative phosphorylation, was added after 12 acquisitions to completely collapse the electrical gradient established by the respiratory chain ($\Delta\Psi$).

Trypan blue quenching experiment. After transfection with mtGFP, MCU–GFP or N33-D1cpV, cells were analysed on spinning disk imaging system composed of an inverted microscope (Zeiss Axiovert 200) equipped with a PlanApo 63×/1.4 numerical aperture objective, a BD-CARVII confocal head, a Photometrics Cascade 512B camera, a piezoelectric *Z* motor (Phisik instruments) and a 300 W Xenon bulb (Sutter Instruments). At each time point, a *Z*-stack was taken (*Z*-step: 500 nm) using standard EGFP filter-set (Semrock GFP-A-Basic-000) and the best focus plane was selected for subsequent analysis. After 1 min permeabilization in intracellular buffer containing 50 μM digitonin, cells were incubated in 1 ml of intracellular buffer, first treated with 4 units of proteinase K and then with 0.005% trypan blue. Best focus images of each time point were background-corrected and mitochondrial mean fluorescence intensity was measured with the MetaMorph software (Molecular devices).

Cell death experiment. Cell sensitivity to apoptotic stimuli was evaluated as previously described[12]. HeLa cells grown on 24-mm round glass coverslips at 30% confluence were transfected with MCU–GFP or GFP alone. The effect on cell fate was evaluated by applying an apoptotic challenge (40 μM C2-ceramide or 200 μM H_2O_2) and comparing the survival of transfected and non-transfected cells. In these experiments, the percentage of GFP-positive cells was calculated before and after applying an apoptotic stimulus (C2-ceramide or H_2O_2). In mock-transfected cells, although the total number of cells is reduced after cell death induction, the apparent transfection efficiency was maintained (that is, transfected and non-transfected cells have the same sensitivity to the apoptotic stimulus and thus die to the same extent). However, when cells are transfected with a construct influencing their sensitivity to apoptosis, this will be reflected by a change in the fraction of fluorescent cells, that is, in the 'apparent' transfection efficiency. Thus, protection from apoptosis results into an apparent increase of transfection, whereas a decrease reflects a higher sensitivity to apoptosis. Data are reported as the mean percentage change in the apparent transfection efficiency after apoptotic challenge compared to vehicle-treated cells. Cells were extensively washed with PBS, stained with 4′,6-diamidino-2-phenylindole (DAPI) and two images per field (blue and green fluorescence) were taken at 20× magnification (mean transfection efficiencies were roughly 30% for both GFP and MCU–GFP). At least 10 fields per coverslip were randomly imaged and counted. Data presented are the sum of at least two different coverslip per experimental condition carried out in three different independent experiments.

Immunogold electron microscopy. HeLa cells were transfected with control (pcDNA3.1) or MCU–Flag plasmids. Cells were fixed in 4% formaldehyde and 0.1% glutaraldehyde in 0.1 M cacodylate buffer 24 h after transfection. Dehydration was carried out by subsequent incubation with increasing ethanol concentrations (20, 40, 60, 80 and 100%). Cells were then infiltrated and embedded in LRW resin and 100-nm slices were obtained. For immunogold staining slices were first blocked in PBS containing 2% BSA and 0.05% Triton X-100 and then incubated for 1 h with anti-Flag (Sigma) antibody. After five washes with PBS, slices were incubated for 1 h with an isotype matched 10-nm gold-particles-conjugated secondary antibody (Sigma). After five washes with PBS and one wash with H_2O, slices were contrasted with uranyl acetate for 15 min and lead citrate for 6 min. Slices observation and imaging were carried out on a Tecnai G^2 TEM.

Protein expression and purification. *E. coli*: MCU and $MCU^{\Delta MTS}$ were cloned into the pET-28A(+) vector, as described above. Competent cells were transformed with the indicated constructs. Induction was performed at 20 °C ($D_{600} = 0.4$) with 0.35 mM isopropyl-β-D-thiogalactoside (IPTG) for 24 h. After sonication in buffer A (50 mM Na-phosphate, pH 7.4, 300 mM NaCl and protease inhibitor cocktail), the membranous fraction was collected by centrifugation (10 min at 12000*g*) and solubilized with *n*-decyl-β-D-maltopyranoside (2.5%) for 3 h at room temperature. The non-solubilized material was removed by centrifugation and the supernatant was loaded on a nickel chromatography column (HIS-Select Nickel affinity gel, Sigma). Column was washed with buffer A and the protein was eluted with imidazole gradient (50–200 mM). Fractions of 500 μl were collected and dialysed against buffer A without NaCl. A concentration of 0.025% of detergent was maintained throughout the purification.

In vitro: MCU was cloned into pIVEX 1.3 WG, as described above. *In vitro* expression was performed by using a RTS100 Wheat Germ CECF Kit (Roche). After expression, the reaction mix was solubilized with 2% Triton X-100 for 90 min at 30 °C under shaking.

Gel electrophoresis: SDS–PAGE was performed using 6 M urea and standard protocols. Each lane was loaded with 30 μl of eluted fractions and 1 μl of the reaction mix.

Electrophysiology. Electrophysiology experiments were carried out as previously described[31,32]. A Warner Instruments (Hamden) electrophysiological planar bilayer apparatus was used. Bilayers of approximately 150–200 pF capacity were prepared using azolectin in decane containing 1% chloroform (Sigma) across a 250 μM hole in a polystyrene cuvette. Azolectin was partially purified by precipitation with cold acetone from a chloroform solution. The inside of the cuvette constituted the *trans* compartment. The standard experimental medium was 100 mM $CaCl_2$ (or Ca-gluconate), 10 mM HEPES/pH 7.2. The lipid membrane was built under symmetric ionic conditions and both chambers contained 3 ml of solution. The contents of both chambers were stirred by magnetic bars when desired. Connections to the electrodes were provided by agar bridges. Purified proteins were added to the *cis* side. Control experiments with empty membrane or with detergents used for the purification showed no activity. All voltages reported are those of the *cis* chamber, zero being assigned by convention to the *trans* (grounded) side. Currents are considered as positive when carried by cations flowing from the *cis* to the *trans* compartment. A BC-525C unit and headstage were used to control parameters and amplify signals. Output was recorded with a 10 kHz bandwidth on videotape using a Medical Systems PCM-2 interface. Data were acquired at 2 kHz, filtered at 500 Hz and analysed offline using the pClamp program set (Axon Instruments). Conductance was determined by averaging the measured amplitudes of single channel events ($n \geq 50$) at various applied voltages.

Statistical analysis of data. Statistical data are presented as mean ± s.e.m., unless otherwise specified. Significance was calculated by Student's *t*-test, and correlation analysis was performed with the SigmaPlot 11.0 software (Systat Software Inc.).

27. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305,** 567–580 (2001).
28. Schoenmakers, T. J. M., Visser, G. J., Flik, G. & Theuvenet, A. P. R. CHELATOR: an improved method for computing metal ion concentrations in physiological solutions. *Biocomputing* **12,** 870–879 (1992).
29. Bolte, S. & Cordelieres, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224,** 213–232 (2006).
30. Costes, S. V. *et al.* Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys. J.* **86,** 3993–4003 (2004).
31. Szabò, I., Soddemann, M., Leanza, L., Zoratti, M. & Gulbins, E. Single-point mutations of a lysine residue change function of Bax and Bcl-x_L expressed in Bax- and Bak-less mouse embryonic fibroblasts: novel insights into the molecular mechanisms of Bax-induced apoptosis. *Cell Death Differ.* **18,** 427–438 (2011).
32. Teardo, E. *et al.* Characterization of a plant glutamate receptor activity. *Cell. Physiol. Biochem.* **26,** 253–262 (2010).

LETTER

doi:10.1038/nature10234

Integrative genomics identifies MCU as an essential component of the mitochondrial calcium uniporter

Joshua M. Baughman[1,2]*, Fabiana Perocchi[1,2]*, Hany S. Girgis[1,2], Molly Plovanich[1,2], Casey A. Belcher-Timme[1,2], Yasemin Sancak[1,2], X. Robert Bao[1,2], Laura Strittmatter[1,2], Olga Goldberger[1,2], Roman L. Bogorad[3], Victor Koteliansky[4] & Vamsi K. Mootha[1,2]

Mitochondria from diverse organisms are capable of transporting large amounts of Ca^{2+} via a ruthenium-red-sensitive, membrane-potential-dependent mechanism called the uniporter[1–4]. Although the uniporter's biophysical properties have been studied extensively, its molecular composition remains elusive. We recently used comparative proteomics to identify MICU1 (also known as CBARA1), an EF-hand-containing protein that serves as a putative regulator of the uniporter[5]. Here, we use whole-genome phylogenetic profiling, genome-wide RNA co-expression analysis and organelle-wide protein coexpression analysis to predict proteins functionally related to *MICU1*. All three methods converge on a novel predicted transmembrane protein, CCDC109A, that we now call 'mitochondrial calcium uniporter' (MCU). MCU forms oligomers in the mitochondrial inner membrane, physically interacts with MICU1, and resides within a large molecular weight complex. Silencing *MCU* in cultured cells or *in vivo* in mouse liver severely abrogates mitochondrial Ca^{2+} uptake, whereas mitochondrial respiration and membrane potential remain fully intact. MCU has two predicted transmembrane helices, which are separated by a highly conserved linker facing the intermembrane space. Acidic residues in this linker are required for its full activity. However, an S259A point mutation retains function but confers resistance to Ru360, the most potent inhibitor of the uniporter. Our genomic, physiological, biochemical and pharmacological data firmly establish MCU as an essential component of the mitochondrial Ca^{2+} uniporter.

To predict proteins that are functionally related to MICU1 (ref. 5) and essential for mitochondrial calcium (Ca^{2+}) uptake, we performed three systematic computational analyses. First, we ranked all ~20,000 mammalian genes on the basis of the similarity of their phylogenetic profile to *MICU1*, where the phylogenetic profile of a gene is defined as the binary vector of presence or absence of its homologues across 500 evolutionarily diverse organisms[6,7]. Second, we scored all ~20,000 mammalian genes for their expression similarity to *MICU1* across 81 mouse cell types and tissues using a genome-wide RNA expression atlas[8]. Third, we scored protein expression similarity between MICU1 and all ~1,100 mitochondrial proteins, based on their pattern of peptide abundance across 14 different mouse tissues[6]. All three computational methods (Fig. 1a–c) spotlight an unstudied protein (previously called CCDC109A, accession number NM_138357.1) that we now call 'mitochondrial calcium uniporter' (MCU). MCU, which has two predicted transmembrane domains, was first discovered in our proteomic analysis as a mitochondrial protein detected in 12 different mouse tissues[6]. *MCU*, which bears no sequence similarity to *MICU1*, is tied as the eighth closest phylogenetic neighbour genome-wide (Fig. 1a), being either co-present or co-absent with *MICU1* in 495 of 500 organisms evaluated (Hamming distance = 5). *MCU* is also the second highest scoring gene in the genome-wide mRNA co-expression analysis (Fig. 1b) and is the top scoring protein amongst all ~1,100 mitochondrial proteins for protein coexpression with MICU1 across 14 mouse tissues (Fig. 1c).

Our prediction of a functional relationship between MICU1 and MCU was further corroborated by evidence of a physical interaction. By transfecting green-fluorescent-protein-tagged MCU (MCU–GFP) into cells stably expressing V5-tagged MICU1 (MICU1–V5) and vice versa, we were able to recover both GFP-tagged proteins following

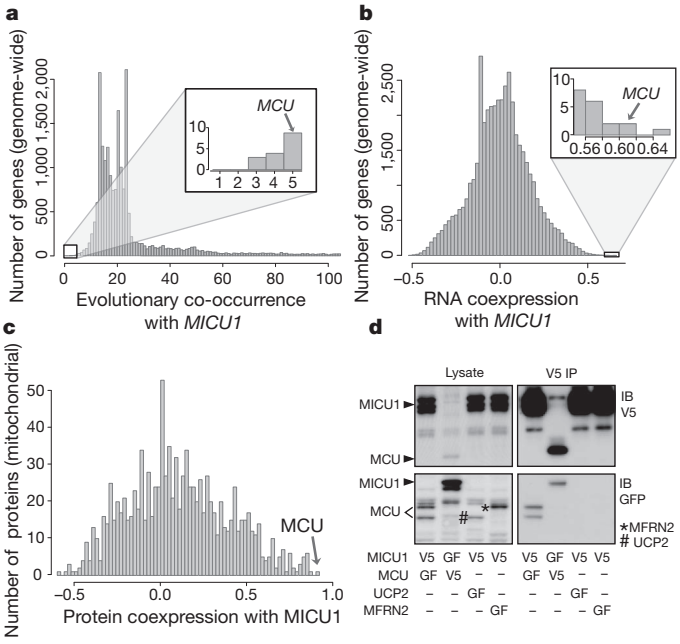


Figure 1 | Integrative genomics predicts MCU to be functionally related to MICU1 **a–c**, Phylogenetic profile neighbours (**a**), RNA co-expression neighbours (**b**) and protein co-expression neighbours (**c**) of MICU1. Hamming distances between phylogenetic profiles were computed genome-wide for all 20,000 mammalian genes across 500 fully sequenced organisms. Genes co-expressed with *MICU1* were computed genome-wide by Pearson correlation using a mouse atlas of 81 tissues. Protein expression correlation with MICU1 was analysed for all mitochondrial proteins across 14 mouse tissues. **d**, Coimmunoprecipitation of MICU1 and MCU. HEK-293 cells stably expressing MICU1–V5 or MCU–V5 were transfected with MICU1–GFP, MCU–GFP, MFRN2–GFP or UCP2–GFP. Cell lysates were incubated with anti-V5 antibody, immunoprecipitates were resolved on SDS–PAGE, and input lysates and immunoprecipitates were blotted with anti-V5 (top) or anti-GFP (bottom) antibodies. Data are representative of three independent experiments. IP, immunoprecipitation; GF, green fluorescent protein; IB, immunoblot.

[1]Departments of Systems Biology and Medicine, Harvard Medical School and Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [2]Broad Institute, Cambridge, Massachusetts 02142, USA. [3]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Alnylam Pharmaceuticals, Inc., Cambridge, Massachusetts 02142, USA.

*These authors contributed equally to this work.

immunoprecipitation with an anti-V5 antibody (Fig. 1d). The interaction was specific as MICU1–V5 was incapable of pulling down two different GFP-tagged inner membrane proteins (UCP2 and MFRN2, also known as SLC25A28). Similar results were also obtained by immunoprecipitating Flag-tagged MCU and probing for endogenous MICU1 (Supplementary Fig. 1).

Collectively, our three complementary genomic analyses combined with our biochemical data (Fig. 1) predict that MCU is functionally related to MICU1 and that it, too, may participate in mitochondrial Ca^{2+} uptake.

We evaluated the impact of silencing *MCU* on mitochondrial Ca^{2+} uptake in intact and permeabilized cells using RNA interference (RNAi). Silencing *MCU* in a HeLa cell line expressing a mitochondria-targeted aequorin (mt-Aeq) reporter[9] attenuates mitochondrial Ca^{2+} uptake (Fig. 2a) proportionate to the strength of knockdown (Fig. 2b). The RNAi-induced phenotype is not off-target because coexpression of a full-length *MCU* cDNA together with a short hairpin RNA (shRNA) that targets the *MCU* 3′ untranslated region (3′UTR) fully rescues Ca^{2+} uptake (Fig. 2a). Moreover, the RNAi effect is not a trivial consequence of interrupting upstream signalling, because histamine mobilization of cytosolic Ca^{2+} remains intact (Supplementary Fig. 2a), and because we obtain similar results when measuring clearance of exogenously added Ca^{2+} by mitochondria in permeabilized HEK-293 (Fig. 2c) and HeLa cells (Supplementary Fig. 3). In HeLa cells, basal and uncoupled respiration were intact (Supplementary Fig. 2b), mitochondrial membrane potential (ψ_m) was not depolarized (Supplementary Fig. 2c), and mitochondrial morphology remained grossly unchanged (data not shown) after silencing of *MCU*. Basal cytosolic Ca^{2+} levels were 53.9 nM ± 16.3 ($n = 30$) in *MCU* knockdown HeLa cells and 70.9 nM ± 11.7 ($n = 14$) in control sh-LACZ cells. Although mitochondrial Ca^{2+} buffering is known to shape cytosolic Ca^{2+} transients in many cell types, its inhibition through silencing of *MCU* did not show a significant impact on cytosolic Ca^{2+} clearance after histamine stimulation (Supplementary Fig. 2a). However, this result could reflect an incomplete silencing of *MCU*-mediated mitochondrial Ca^{2+} uptake in HeLa cells (Fig. 2a, b). Mitochondrial Ca^{2+} uptake has also previously been shown to stimulate ATP production by allosterically activating three tricarboxylic acid (TCA) cycle dehydrogenases[10]. Using a previously reported protocol[11], we found that permeabilized sh-MCU cells exhibit attenuated NAD(P)H elevation in response to exogenously added Ca^{2+} (Supplementary Fig. 2d), demonstrating that silencing *MCU*, like *MICU1* (ref. 5), attenuates Ca^{2+} activation of the TCA cycle.

To complement these cell-based studies, we analysed mitochondrial calcium uptake in mouse liver mitochondria upon *in vivo* silencing of *MCU*. A key advantage of this experimental system is that Ca^{2+} uptake phenotypes are directly attributable to mitochondria, and additional bioenergetic parameters, notably respiratory flux and ψ_m, can be carefully measured under classic respiratory state transitions[12]. As respiratory state transitions are dependent on TCA cycle metabolism, electron transport, H^+ pumping, ATP/ADP exchange, and transport of phosphate, such studies allow us to evaluate the specificity of MCU's role in mitochondrial Ca^{2+} transport. Using previously described siRNA design and delivery technology[13], we screened 46 distinct siRNA duplexes. We selected a duplex characterized by a half-maximal inhibitory concentration of 69 pM that achieved >90% knockdown at a concentration of 5 nM in cultured mouse liver cells (Fig. 2d). We performed a large-scale synthesis of this siRNA duplex and encapsulated it into a lipid-based formulation optimized for liver-specific delivery[14,15]. As a negative control we used a siRNA duplex specific for the firefly luciferase gene (si-LUC). We performed weekly tail vein injections of the siRNA duplexes over a 3 week period and achieved 84 ± 2% ($n = 3$) *in vivo* mRNA knockdown of *MCU* in the liver, which was corroborated by immunoblot analysis (Fig. 2d). Both groups of mice showed normal weight gain (5.6 ± 1.3% of body weight for si-LUC and 5.0 ± 3.0% of body weight for si-MCU, $n = 3$) and the gross appearance of the livers did not differ. Mitochondria isolated from

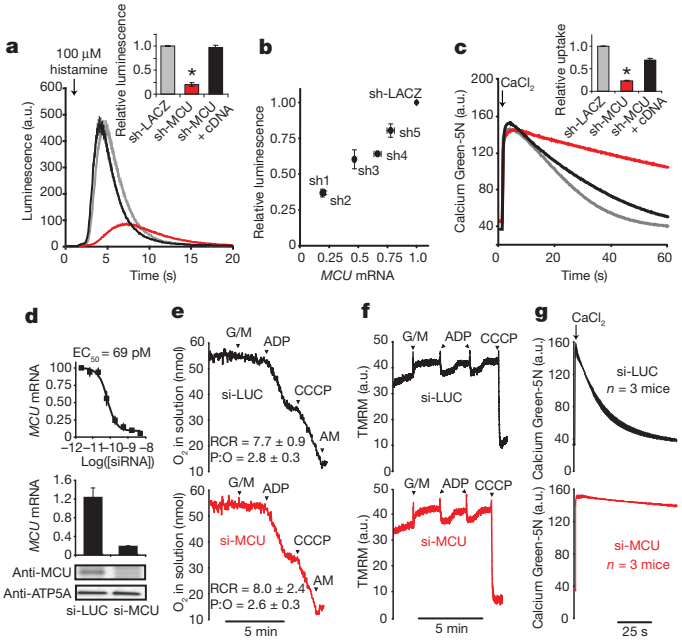


Figure 2 | MCU is required for mitochondrial Ca^{2+} uptake in cultured cells and in purified mouse liver mitochondria. **a**, Representative luminescence measurements of a mitochondrial aequorin Ca^{2+} reporter after histamine stimulation in HeLa cells expressing sh-LACZ, sh-MCU, or a combination of sh-MCU and an RNAi-resistant cDNA for *MCU* (mean ± s.e.m., $n = 10$ traces). Inset shows statistical analysis of the maximal luminescence (mean ± s.d., $n = 10$ traces, $^*P < 0.001$). a.u., arbitrary units. **b**, Relationship between *MCU* mRNA expression and histamine-induced mitochondrial Ca^{2+} uptake (maximal aequorin luminescence) recorded from five independent shRNAs targeting *MCU* and normalized to sh-LACZ (mean ± s.d., $n = 3$). **c**, Representative traces of Ca^{2+} uptake in digitonin-permeabilized sh-MCU HEK-293 cells or sh-LACZ control cells after addition of 50 μM final concentration of $CaCl_2$. Inset reports linear fits of uptake kinetics between 15 and 20 s, normalized to sh-LACZ (mean ± s.d., $n = 3$, $^*P < 0.001$). Ca^{2+} was measured with Calcium Green-5N. **d**, *In vitro* dose–response of a selected siRNA duplex targeting mouse *MCU*. Relative expression of *MCU* mRNA in livers following weekly injections of si-MCU or si-LUC for 3 weeks, normalized to expression in PBS-treated mice. Immunoblot analysis of liver mitochondria isolated from mice treated with si-MCU or control si-LUC using antibodies against MCU and ATP5A1 as a loading control. **e**, Oxygen consumption measurements of isolated mitochondria in a well-stirred cuvette. Glutamate and malate (G/M), ADP, uncoupler (carbonyl cyanide *m*-chlorophenyl hydrazone, CCCP), antimycin (AM) were added at indicated time points. Respiratory control ratio (RCR) and ADP:O ratio (P:O) were computed from three separate mice for each group. **f**, Mitochondrial membrane potential (ψ_m) measured by tetramethyl rhodamine methyl ester (TMRM) in isolated liver mitochondria. **g**, Ca^{2+} uptake kinetics in energized liver mitochondria following the addition of 50 μM final $CaCl_2$. Extra-mitochondrial Ca^{2+} was measured with Calcium Green-5N (mean ± s.e.m., $n = 3$ mice). Traces depicted in **e** and **f** are representative of measurements made from three independent mouse experiments performed on separate days.

these livers were physiologically intact, capable of undergoing robust respiratory state transitions (Fig. 2e, f). The respiratory control ratio (RCR) was comparable in si-LUC (7.7 ± 0.9, $n = 3$) and si-MCU animals (8.0 ± 2.4, $n = 3$). The ADP:O ratios were 2.77 ± 0.3 ($n = 3$) and 2.61 ± 0.3 ($n = 3$) in si-LUC and si-MCU mitochondria, respectively, indicating intact respiratory chain physiology. However, mitochondria from si-MCU mice show a profound and near complete loss in Ca^{2+} uptake in response to extramitochondrial pulses of Ca^{2+} (Fig. 2g and Supplementary Fig. 4).

Collectively, our physiology studies, comprising three different experimental systems (intact cells, permeabilized cells and mitochondria purified from mouse organs), three different cell types (HeLa, HEK-293 and mouse liver), and two different silencing modalities (*in vitro* lentiviral shRNA and *in vivo* siRNA), firmly establish that

MCU is essential for high capacity Ca^{2+} transport into mitochondria, and that the phenotype is not a secondary consequence of alterations in cytosolic Ca^{2+} signalling or of impaired ion or metabolite transport across the mitochondrial inner membrane. To our knowledge, this is the first time that mitochondrial Ca^{2+} uptake has been silenced *in vivo* in an animal.

We next established MCU's precise sub-organellar localization and topology. MCU was first identified as a mitochondrial protein in our previous proteomic survey[6], where it was detected as a low abundance protein in 12 out of 14 mouse tissues, being missed only in heart and kidney. Confirmatory immunoblot analysis suggests it is also present in the heart and the kidney (data not shown), indicating it is universally expressed. As expected, a carboxy terminus GFP-tagged version of MCU localizes exclusively to mitochondria in HeLa cells (Fig. 3a) as well as in HEK-293 cells (data not shown). To biochemically validate MCU's localization, we fractionated HeLa cell lysate and observed significant enrichment of the approximately 35 kDa native MCU in the mitochondrial fraction (Fig. 3b). The observed molecular weight of MCU is lower than the predicted 40 kDa and is consistent with a predicted cleavable amino-terminal mitochondrial targeting sequence[16]. Multiple algorithms predict that MCU contains two transmembrane domains[17,18]. To experimentally confirm mitochondrial inner-membrane integration and to discriminate between transmembrane and peripheral membrane configurations, we performed alkaline carbonate extraction of proteins from intact mitochondria isolated from HEK-293 cells expressing MCU-V5. We find that MCU remains in the insoluble fraction, proving that it is a bona fide transmembrane protein (Fig. 3c). To establish the topology of MCU across the inner membrane, we disrupted the mitochondrial outer membrane by hypotonic swelling and treated the resulting mitoplasts with proteinase K. Proteins such as OXA1 that are exposed to the mitochondrial intermembrane space are digested by proteinase K, whereas COII (an integral inner membrane protein, also known as MT-CO2) and MCU are completely protected (Fig. 3d), despite being substrates for proteinase K (Supplementary Fig. 5). Our data therefore indicate that MCU is localized to the inner membrane with both its N- and C-termini facing the matrix space.

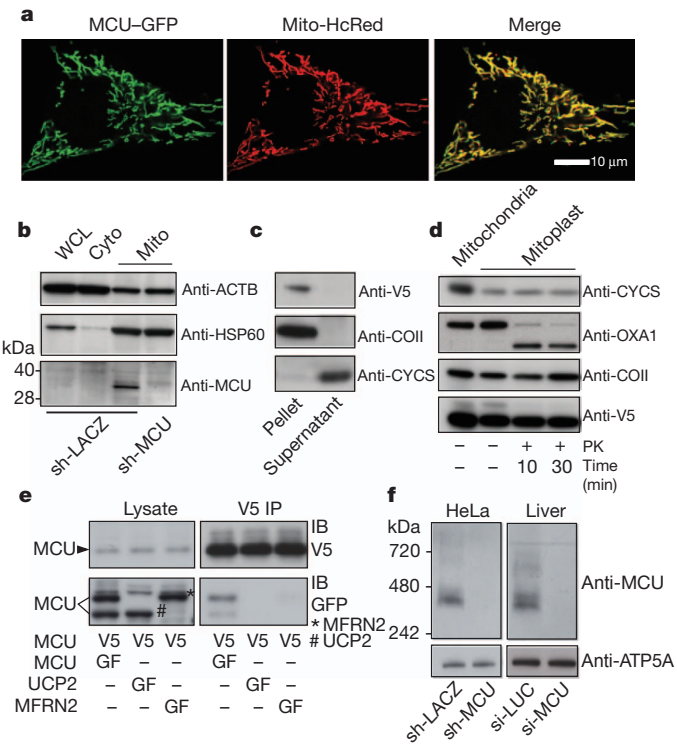


Figure 3 | MCU is oligomeric and resides in the mitochondrial inner membrane as a larger complex. **a**, Confocal imaging of MCU–GFP co-expressed with mitochondria-targeted HcRed (Mito-HcRed) in HeLa cells. **b**, Immunoblot analysis of HeLa whole-cell lysate (WCL), cytosol (Cyto) or crude mitochondrial fractions (Mito), using antibodies against MCU, HSP60 (matrix protein, also known as HSPD1), or ACTB (cytosol). **c**, Immunoblot analysis of soluble (supernatant) and insoluble (pellet) fractions following alkaline carbonate extraction of mitochondrial fractions from HEK-293 cells expressing MCU–V5. Immunoblot analysis was performed using antibodies against V5, COII (integral inner membrane protein) and CYCS (soluble intermembrane space protein). **d**, Immunoblot analysis after proteinase K (PK) treatment of MCU–V5-expressing HEK-293 mitoplasts for indicated times. **e**, Anti-V5 immunoprecipitations performed as in Fig. 1d using lysates from HEK-293 cells stably expressing MCU–V5 and transiently transfected with MCU–GFP, UCP2–GFP, or MFRN2–GFP. **f**, Blue native PAGE analysis of mitochondrial fractions from HeLa cells (stably expressing sh-LACZ or sh-MCU, left panel) or from livers of mice (si-LUC or si-MCU, right panel) and immunoblotted for MCU. ATP5A1 is used as a loading control.

If MCU is an integral component of the uniporter, we might expect that it oligomerizes and operates within a larger complex. To test for MCU oligomerization, we performed pull-down experiments and found that V5-tagged MCU is capable of immunoprecipitating GFP-tagged MCU (Fig. 3e), indicating that MCU can form multimers. To control for nonspecific interactions we used two inner mitochondrial membrane spanning transporters as controls, neither of which immunoprecipitated with MCU–V5 (Fig. 3e). We performed blue native gel separation of digitonin-solubilized mitochondria purified from HeLa cells and found that MCU migrates at an apparent molecular weight of ~450 kDa. This larger complex disappears following silencing with sh-MCU (Fig. 3f, left panel). We obtained similar results using mitochondria from livers of si-LUC and si-MCU mice (Fig. 3f, right panel). Collectively, these studies indicate that MCU oligomerizes in the mitochondrial inner membrane as part of a larger molecular weight complex.

Our topology analysis (Fig. 3c, d) combined with computational predictions of membrane spanning domains indicate that MCU's two transmembrane helices, TM1 and TM2, are linked by a short stretch of amino acids that face the intermembrane space and contain what we now term a 'DIME' motif (Fig. 4a). We sought to determine whether conserved amino acids near and within the DIME motif are required for MCU-mediated Ca^{2+} uptake. We created a series of four alanine mutants at conserved residues (E257A, S259A, D261A, E264A) and expressed them on an sh-MCU background to evaluate their ability to rescue Ca^{2+} transport (Fig. 4b). One mutant, S259A, was capable of providing significant rescue of the sh-MCU phenotype (Fig. 4b). The remaining three mutants (E257A, D261A, E264A), however, failed to restore mitochondrial calcium uptake despite comparable expression to the S259A mutant (Supplementary Fig. 6a), signifying that these highly conserved acidic residues are critical for Ca^{2+} transport.

Ruthenium red and a related compound, Ru360, are the most potent inhibitors of the mitochondrial Ca^{2+} uniporter, yet their target remains unknown[19,20]. These cell-impermeant inhibitors have been useful pharmacological tools in studies of isolated mitochondria[21,22] and are believed to act on the outer face of the inner membrane. We next sought to determine whether the functional S259A mutant might influence Ru360 sensitivity. Ru360 strongly abolishes mitochondrial Ca^{2+} uptake in permeabilized HEK-293 cells (Fig. 4c). Transient overexpression of *MCU* leads to a mild resistance to Ru360 compared to control (Fig. 4c). However, transient expression of the S259A mutant confers marked resistance to Ru360 in both HEK-293 and HeLa cells (Fig. 4c, Supplementary Fig. 6b, c, d, e). These experiments clearly indicate a functional role for the linker region in both activity and Ru360 sensitivity of MCU.

Our genomic, physiological, biochemical and pharmacological data firmly establish that MCU is an oligomeric, mitochondrial inner membrane protein that is essential for calcium uniporter activity. In our current studies, overexpression of *MCU* alone did not give rise to a

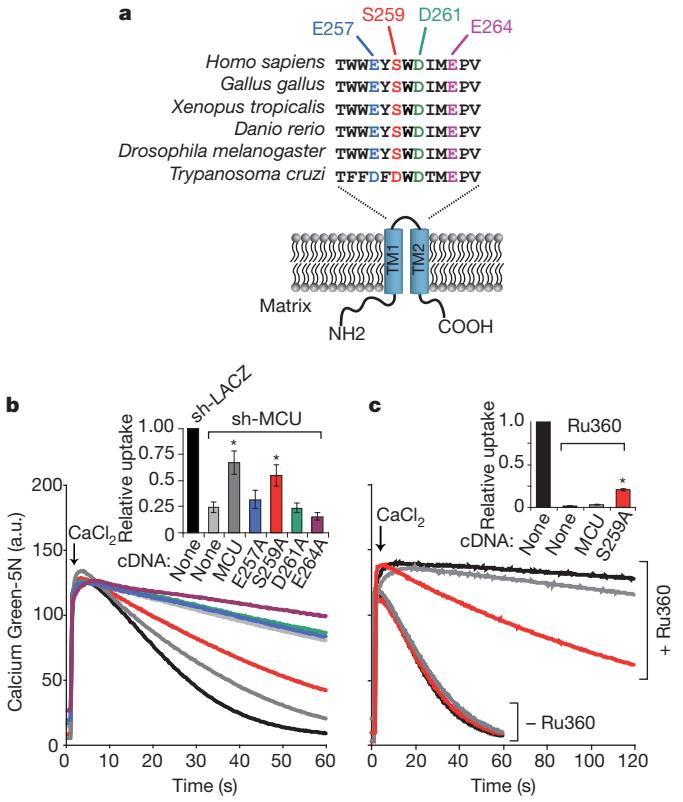


Figure 4 | Impact of point mutations on MCU activity and its sensitivity to Ru360. **a**, Schematic of MCU topology across the mitochondrial inner membrane and a multiple sequence alignment of the linker sequence containing a DIME motif. TM1 and TM2 are two transmembrane domains. **b**, Ca^{2+} uptake in permeabilized sh-MCU HEK-293 cells transiently expressing MCU mutants. Inset reports linear fits of uptake kinetics between 15 and 25 s, normalized to sh-LACZ (mean ± s.d., $n = 3$, $^*P < 0.01$) **c**, Ca^{2+} uptake in HEK-293 cells transiently expressing wild-type MCU or the S259A mutant, in the presence or absence of 0.5 μM Ru360. Inset reports linear fits of uptake kinetics between 30 and 60 s for Ru360-treated cells, and between 15 and 25 s for untreated cells. Uptake rates are normalized to untreated HEK-293 cells (mean ± s.d., $n = 3$, $^*P < 0.01$).

marked gain of Ca^{2+} uptake in HeLa cells (Fig. 2a), indicating that additional components or chaperones may be limiting *in vivo*. Although its precise mechanism remains to be elucidated, MCU fulfils multiple criteria for being an integral component of the uniporter, where it could serve as the pore-forming subunit or as a closely associated partner protein. First, loss of *MCU* leads to a profound defect in mitochondrial Ca^{2+} uptake in three distinct systems: intact cells, permeabilized cells and mouse liver (Fig. 2). Second, it has two bona fide transmembrane domains and oligomerizes to form a larger molecular weight complex in the mitochondrial inner membrane (Fig. 3e, f). Third, mutations at evolutionarily conserved acidic residues attenuate its activity (Fig. 4b). Fourth, a point mutation in MCU confers resistance against Ru360, the most potent inhibitor of the uniporter (Fig. 4c and Supplementary Fig. 6). The current study further supports our previously reported phylogenomic signature of the mitochondrial Ca^{2+} uniporter[5], and together these studies identify MICU1 and MCU as an evolutionarily conserved duo that are critical for mitochondrial Ca^{2+} uptake. Whereas MICU1 may serve as a regulator, MCU is likely to represent a core component of the uniporter. Future studies will be necessary to establish the exact mechanism by which these proteins function together to orchestrate the mitochondrial Ca^{2+} cycle.

METHODS SUMMARY

Candidate human genes required for mitochondrial Ca^{2+} uptake were prioritized on the basis of the co-occurrence of their homologues with *MICU1* across 500 evolutionarily diverse organisms[6,7] and their RNA and protein co-expression across mouse cell types and tissues[6,8]. Stable knockdown of *MCU* in HeLa and HEK-293 cells was achieved using lentiviral shRNA-expressing constructs available from the Broad Institute RNAi Consortium[23]. *In vivo* silencing of *MCU* in mouse liver was achieved via weekly tail vein injections of selected siRNAs in lipid-based formulations[13–15,24]. cDNA rescue studies in *MCU* knockdown cells were carried out by overexpression of wild-type or mutant versions of full-length MCU cDNAs together with a shRNA that targets the *MCU* 3′UTR. Agonist-induced rises in mitochondrial Ca^{2+} in intact HeLa cells were measured by luminescence of a mitochondria-targeted aequorin reporter[9]. Extra-mitochondrial calcium in permeabilized HeLa cells, HEK-293 and isolated liver mitochondria was measured using Calcium Green-5N (ref. 5). Single-cell studies of cytosolic calcium were performed as previously described[25]. Mitochondrial respiration, membrane potential and NAD(P)H were measured via established protocols[11,26]. Crude mitochondria and mitoplasts were prepared from cultured HEK-293 cells expressing a C terminus V5-tagged version of MCU[5]. Protein topology was assessed by alkaline carbonate extraction from crude mitochondria and proteinase K digestion of mitoplasts[27]. Protein interaction studies were performed by immunoprecipitation with anti-V5 and anti-Flag antibodies. Native gel electrophoresis was performed as previously described[28]. Unless otherwise indicated, data are summarized as mean ± standard deviation (s.d.), and *P*-values were computed from *t*-tests.

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

Received 9 November 2010; accepted 2 June 2011.
Published online 19 June 2011.

1. DeLuca, H. F. & Engstrom, G. W. Calcium uptake by rat kidney mitochondria. *Proc. Natl Acad. Sci. USA* **47,** 1744–1750 (1961).
2. Vasington, F. D. & Murphy, J. V. Ca^{++} ion uptake by rat kidney mitochondria and its dependence on respiration and phosphorylation. *J. Biol. Chem.* **237,** 2670–2677 (1962).
3. Carafoli, E. & Lehninger, A. L. A survey of the interaction of calcium ions with mitochondria from different tissues and species. *Biochem. J.* **122,** 681–690 (1971).
4. Gunter, K. & Gunter, T. E. Transport of calcium by mitochondria. *J. Bioenerg. Biomembr.* **26,** 471–485 (1994).
5. Perocchi, F. *et al. MICU1* encodes a mitochondrial EF hand protein required for Ca^{2+} uptake. *Nature* **467,** 291–296 (2010).
6. Pagliarini, D. J. *et al.* A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134,** 112–123 (2008).
7. Pellegrini, M., Marcotte, E. M., Thompson, M. J., Eisenberg, D. & Yeates, T. O. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc. Natl Acad. Sci. USA* **96,** 4285–4288 (1999).
8. Lattin, J. E. *et al.* Expression analysis of G Protein-Coupled Receptors in mouse macrophages. *Immunome Res.* **4,** 5 (2008).
9. Rizzuto, R., Simpson, A. W., Brini, M. & Pozzan, T. Rapid changes of mitochondrial Ca^{2+} revealed by specifically targeted recombinant aequorin. *Nature* **358,** 325–327 (1992).
10. Denton, R. M. & McCormack, J. G. The role of calcium in the regulation of mitochondrial metabolism. *Biochem. Soc. Trans.* **8,** 266–268 (1980).
11. Territo, P. R., Mootha, V. K., French, S. A. & Balaban, R. S. Ca^{2+} activation of heart mitochondrial oxidative phosphorylation: role of the F_0/F_1-ATPase. *Am. J. Physiol. Cell Physiol.* **278,** C423–C435 (2000).
12. Chance, B. & Williams, G. R. Respiratory enzymes in oxidative phosphorylation. I. Kinetics of oxygen utilization. *J. Biol. Chem.* **217,** 383–393 (1955).
13. Musunuru, K. *et al.* From noncoding variant to phenotype via *SORT1* at the 1p13 cholesterol locus. *Nature* **466,** 714–719 (2010).
14. Akinc, A. *et al.* A combinatorial library of lipid-like materials for delivery of RNAi therapeutics. *Nature Biotechnol.* **26,** 561–569 (2008).
15. Akinc, A. *et al.* Targeted delivery of RNAi therapeutics with endogenous and exogenous ligand-based mechanisms. *Mol. Ther.* **18,** 1357–1364 (2010).
16. Claros, M. G. & Vincens, P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur. J. Biochem.* **241,** 779–786 (1996).
17. Bernsel, A., Viklund, H., Hennerdal, A. & Elofsson, A. TOPCONS: consensus prediction of membrane protein topology. *Nucleic Acids Res.* **37,** W465–W468 (2009).
18. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305,** 567–580 (2001).
19. Moore, C. L. Specific inhibition of mitochondrial Ca^{++} transport by ruthenium red. *Biochem. Biophys. Res. Commun.* **42,** 298–305 (1971).
20. Ying, W. L., Emerson, J., Clarke, M. J. & Sanadi, D. R. Inhibition of mitochondrial calcium ion transport by an oxo-bridged dinuclear ruthenium ammine complex. *Biochemistry* **30,** 4949–4952 (1991).
21. Bernardi, P. Mitochondrial transport of cations: channels, exchangers, and permeability transition. *Physiol. Rev.* **79,** 1127–1155 (1999).
22. Kirichok, Y., Krapivinsky, G. & Clapham, D. E. The mitochondrial calcium uniporter is a highly selective ion channel. *Nature* **427,** 360–364 (2004).

23. Moffat, J. *et al.* A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124,** 1283–1298 (2006).
24. Semple, S. C. *et al.* Rational design of cationic lipids for siRNA delivery. *Nature Biotechnol.* **28,** 172–176 (2010).
25. Rebres, R. A. *et al.* Synergistic Ca^{2+} responses by $G\alpha_i$- and $G\alpha_q$-coupled G-protein-coupled receptors require a single PLCβ isoform that is sensitive to both $G\beta\gamma$ and $G\alpha_q$. *J. Biol. Chem.* **286,** 942–951 (2011).
26. Gohil, V. M. *et al.* Nutrient-sensitized screening for drugs that shift energy metabolism from mitochondrial respiration to glycolysis. *Nature Biotechnol.* **28,** 249–255 (2010).
27. Ryan, M. T., Voos, W. & Pfanner, N. Assaying protein import into mitochondria. *Methods Cell Biol.* **65,** 189–215 (2001).
28. Schagger, H. & von Jagow, G. Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.* **199,** 223–231 (1991).

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

Acknowledgements We thank R. Nilsson, J. Engreitz and S. Calvo for bioinformatics assistance; D. Root and S. Silver for assistance in lentiviral RNAi; B. R. Bettencourt, K. Charisse, S. Kuchimanchi and L. Speciner for siRNA design, synthesis and formulation; M. Blower, J. Avruch and R. Ward for advice; and members of the Mootha laboratory for valuable feedback. J.M.B. and L.S. were supported by graduate student fellowships from the National Science Foundation. This work was supported by grants from the National Institutes of Health (GM0077465, DK080261) awarded to V.K.M.

Author Contributions J.M.B., F.P. and V.K.M. conceived of the project and its design. J.M.B., F.P., H.S.G., M.P., O.G., L.S., C.A.B.-T., X.R.B., Y.S. and R.L.B. performed experiments and data analysis. V.K. aided in experimental design. V.K.M., J.M.B., F.P. and M.P. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to V.K.M. (vamsi@hms.harvard.edu).

METHODS

Phylogenetic profiling and co-expression analyses. We performed genome-wide phylogenetic profiling using a previously reported phylogenetic matrix across 500 sequenced genomes using the Hamming distance as our similarity metric[6]. For genome-wide mRNA co-expression analysis, we used a publicly available genome-wide tissue gene expression atlas (NCBI GSE10246)[8]. We used the published gene expression intensities normalized to the median for each chip, and then computed the Pearson correlation to *MICU1* expression. For protein co-expression analysis, we used our MitoCarta protein atlas of mitochondrial protein expression across 14 tissues[6]. We summarized the abundance of a protein as the number of detected peptides per protein per tissue, and computed the Pearson correlation to MICU1 expression.

Cell culture. HEK-293 cells, Hepa-1c1c7 and HeLa cells were received from the ATCC. HeLa cells expressing a mitochondrial matrix targeted aequorin (mt-Aeq) were purchased from Aequotech (AT-002-H). All cells were grown in Dulbecco's modified Eagle medium (DMEM) high glucose medium (Invitrogen) with 10% FBS (Sigma) at 37 °C and 5% CO_2. mt-Aeq cells were maintained in 100 μg ml^{-1} geneticin (Gibco).

Plasmids and reagents. Lentiviral vectors for expressing shRNA (pLKO.1) or V5-tagged cDNAs (pLEX983) were obtained from the Broad Institute's RNAi Consortium (TRC). To silence *MCU*, we used the following hairpins from the TRC: sh1 (TRCN0000137529, 5′-CCAGCAACTATACACCACACT-3′), sh2 (TRCN0000133861, 5′-GCAAGGAGTTTCTTTCTCTTT-3′), sh3 (TRCN0000133984, 5′-CAATCAACTCAAGGATGCAAT-3′), sh4 (TRCN0000135430, 5′-GCCATG GCAATGTATGCATAT-3′), sh5 (TRCN0000138929, 5′-GATCGCTTCCTGG CAGAATTT-3′). For transfections, cDNAs were inserted into pcDNA vectors with C-terminal V5/His6× or GFP tags (Invitrogen pDEST40 and pDEST47, respectively) using Gateway LR (Invitrogen). A full-length *MCU* human cDNA (NM_138357.1) without a stop codon was cloned into the pDONR223 Gateway vector and subsequently cloned into pLEX983, pDEST40 and pDEST47 vectors. MCU mutants E257A, S259A, D261A and E264A mutations were created using Stratagene QuikChange mutagenesis. We used TaqMan assays (ABI) to quantify the *MCU* transcript (*CCDC109A*).

Lentivirus production and infection. Procedures and reagents for virus production and infection were adapted from the Broad RNAi Consortium protocols and were described previously[23]. Selection was begun 24 h post-infection with 2 μg ml^{-1} puromycin or 5 μg ml^{-1} blasticidin.

Rescue of sh-*MCU* and overexpression of MCU mutants. HeLa cells stably overexpressing shRNAs targeting either *LACZ* or the 3′UTR of *MCU* (TRCN0000133861) were transduced with either pLEX983-MCU wild-type or mutant versions and selected with 5 μg ml^{-1} blasticidin for at least 1 week before assaying. Wild-type HEK-293 or HEK-293 cells stably overexpressing sh-MCU (TRCN0000133861) were transfected with wild-type or mutant MCU cDNAs using Fugene6 reagent from Roche. To ensure high transfection efficiency HEK-293 were transfected twice 3 days apart and Ca^{2+} uptake was assayed 7 days after the first transfection.

Synthesis and *in vitro* screening of siRNAs targeting *MCU*. siRNAs with the lowest predicted off-target potentials (46) and 100% homology with mouse sequence NM_001033259.3 were selected for synthesis and screening. Single-strand RNAs were produced at Alnylam Pharmaceuticals as previously described[13]. MCU siRNA duplex with lowest EC_{50} and lowest predicted off-target potential was selected. The siRNA sense strand is 5′-GAcuGAGAGA cccAuuAcAdTsdT-3′, antisense is 5′-UGuAAUGGGUCUCUcAGUCdTsdT-3′; control luciferase siRNA sense is 5′-cuuAcGcuGAGuAcuucGAdTsdT-3′, and antisense is 5′-UCGAAGuACUcAGCGuAAGdTsdT-3′. Chemical modifications were introduced to stabilize siRNA *in vivo*, reduce off-target potential of sense strand and minimize immune response (2′-OMe modified nucleotides are in lower case, and phosphorothioate linkages are indicated by 's'). Hepa-1c1c7 cells (seeded 15,000 cells per well in 96-well plates) were transfected with siRNA using the Lipofectamine RNAiMAX reagent according to the manufacturer's protocols. *MCU* mRNA levels were quantified in cell lysates 18–24 h post-transfection using a branched-DNA assay (QuantiGene Reagent System, Panomics), according to the manufacturer's protocols. *MCU* mRNA levels were normalized to *GAPDH* mRNA.

***In vivo* silencing of *MCU*.** All procedures used in animal studies were approved by the Institutional Animal Care and Use Committee. C57BL/6 mice (Charles River laboratories) received either PBS or siRNA in lipid-based formulations (as previously described[14,15,24]) via tail vein injection weekly. After overnight fasting the animals were euthanized by CO_2 inhalation and liver tissues were collected into ice-cold PBS for mitochondria isolation. A piece of liver tissue was snap-frozen in liquid nitrogen for measurements of mRNA and protein levels.

Luminescence-based measurement of mitochondrial Ca^{2+}. Assays were performed using either mt-Aeq HeLa cells that stably express mitochondrial aequorin (Fig. 2b) or wild-type HeLa cells that were transduced with an equal titre of lentivirus to express mitochondrial-targeted aequorin[9] 3 days before assaying (Fig. 2a). 12–18 h before assaying, 50,000 cells were seeded per well in a 96-well plate. Light emission after histamine stimulation was measured as reported previously[5]. After histamine stimulation, cells were treated with 1% Triton X-100 and 10 mM $CaCl_2$ (final concentrations) and the light emission was recorded. To normalize, the maximal luminescence signal after histamine stimulation was divided by the total luminescence recorded per well in the entire experiment.

Measurements of cytosolic Ca^{2+}. Calibrated measurements of cytosolic Ca^{2+} were made via single-cell imaging of cells loaded with 500 nM Fura-2, calibrated as previously described[25]. Population measurement of cytosolic Ca^{2+} in response to 100 μM histamine was performed essentially as described in the Alliance for Cell Signalling protocols (PP00000210) except for using 4 μM Fluo-4 AM-ester and 200,000 cells per well in suspension. Minimum calibration was achieved with 5 μM ionomycin and 5 mM EGTA. Maximum calibration was achieved with 1% Triton X-100 and 20 mM $CaCl_2$ final concentrations. The K_d of Fluo-4 was assumed to be 350 nM.

Assays of ψ_m, oxygen consumption and NAD(P)H in intact and permeabilized cells. For measurements of membrane potential, HeLa cells (50,000 per treatment) were seeded in eight wells of a 96-well plate and incubated with 3.5 μM JC-1 for 20 min. The uncoupler CCCP (5 μM) was added as a positive control. The wells were washed at least four times with phenol-red-free DMEM before measuring. For assaying, red fluorescence (excitation 531 nm, emission 590 nm) was measured followed by green fluorescence (excitation 488 nm, emission 535 nm) in a Perkin-Elmer Envision fluorescence plate reader. Background fluorescence values recorded from wells without cells were subtracted from all fluorescence values before calculating red/green fluorescence ratios. Intact cell respiration was measured using Seahorse XF-24 Extracellular Flux Analyzer, as previously described[26]. To investigate Ca^{2+} stimulation of mitochondrial NAD(P)H in permeabilized HEK-293 cells, we adapted a previously described protocol[11].

Measurement of mitochondrial Ca^{2+} uptake in permeabilized cells. Extramitochondrial free Ca^{2+} was monitored in digitonin-permeabilized cells as previously described[5]. Calcium Green-5N fluorescence (excitation 506 nm, emission 531 nm) was monitored every 0.2 s at room temperature using a Perkin-Elmer Envision plate reader after injection of $CaCl_2$ (~50 μM final concentration). For Ru360 (Calbiochem #557440) treatments, selected concentrations of Ru360 or H_2O (vehicle) were added to the digitonized cells and incubated 2 min before $CaCl_2$ injection. For calculation of relative Ca^{2+} uptake, linear fits were computed for the times indicated.

Mitochondrial isolation, mitoplast preparation, and topology analysis. Cell lysates, crude mitochondria and mitoplasts were prepared from cultured HEK-293 cells as previously described[5]. Immunoblotting of the four fractions (5 μg) was performed with commercially available antibodies: anti-MCU (Sigma #HPA016480), anti-HSP60 (Abcam ab3080), anti-CYCS (MitoSciences #MSA06), anti-OXA1 (BD Biosciences #611980), anti-COII (MitoSciences #MS405) and anti-V5 (Invitrogen, R96025). Alkaline carbonate extraction from crude mitochondria and proteinase K digestion of mitoplasts were performed as described previously[27].

Measurement of respiration, ψ_m and Ca^{2+} uptake in isolated mouse mitochondria. Mitochondria were isolated from freshly collected mouse livers as previously described[26] and resuspended into a buffer containing 220 mM mannitol, 75 mM sucrose, 10 mM HEPES-KOH pH 7.4, 1 mM EDTA and 0.5% BSA and stored on ice until further use. Respiration and ψ_m were measured optically in a well-stirred cuvette as previously described[26]. For Ca^{2+} uptake measurements, 400 μg of isolated mitochondria were resuspended in 500 μl of potassium chloride (KCl) media (125 mM KCl, 2 mM K_2HPO_4, 1 mM $MgCl_2$, 20 mM HEPES-KOH pH 7.4) containing 1 μM Calcium Green-5N and 5 mM glutamate and 5 mM malate. Fluorescence was monitored at room temperature using a Perkin-Elmer LS-50B fluorescence spectrometer equipped with a stirring device. Respiratory control ratios (RCRs) were computed as the ratio of the ADP-stimulated respiration to the respiratory substrates-based respiration. The ADP:O ratio was computed after correcting for the oxygen leak into the cuvette. Values represent the mean ± s.d. of three, independent biological replicate mouse injections performed on three different days.

Confocal microscopy. HeLa cells cultured in Lab-Tek II chambered slides were co-transfected with plasmids expressing MCU–GFP and Mito-HcRed (Clontech #632434). Twenty-four hours post-transfection, cells were washed three times with PBS and imaged using a Leica TCS SP5 confocal microscope.

Blue native studies. Membrane protein complexes were separated by electrophoresis as previously described[28]. Briefly, mitochondria were isolated from ~20 million HeLa cells stably expressing sh-LACZ or sh-MCU according to the MitoSciences protocol (catalogue no. MS851). Total protein was quantified by bicinchoninic assay and 5 μg of mitochondria were solubilized in 1% digitonin

(w/v) on ice. Electrophoresis was performed using the NativePAGE Novex Bis-Tris Gel System from Invitrogen. Protein was transferred to a polyvinylidene fluoride (PVDF) membrane and immunoblotted with a commercially available antibody to MCU (Sigma #HPA016480). Membranes were stripped and re-probed with anti-ATP5A1 (MitoSciences MS507) as a loading control.

Immunoprecipitation studies. One confluent 10-cm dish of HEK-293 cells stably overexpressing either MICU1–V5 or MCU–V5 was washed twice with ice-cold PBS approximately 36 h post-transfection with either MCU–GFP or MICU1–GFP. Cells were lysed for 30 min with 1 ml of lysis buffer containing 1% *n*-dodecyl β-D-maltoside (DDM), 150 mM NaCl, and 50 mM Tris pH 7.4, and the insoluble fraction was removed by centrifugation. The soluble fraction was incubated with Dynabeads Protein G (Invitrogen 100-04D) coated with anti-V5 antibody (Invitrogen #R960-25). Immunoprecipitations were performed according to the manufacturer's guidelines. Eluents were immunoblotted with anti-V5 antibody to confirm binding and with anti-GFP antibody to probe for interactions (Novus catalogue number NB600-308). For anti-Flag immunoprecipitation studies in Supplementary Fig. 1, 2 million HEK-293 cells were plated in 10-cm plates. Cells were transfected with 2 μg of indicated plasmids using Fugene6 1 day later. After 2 days, cells were solubilized in 1 ml of lysis buffer (50 mM HEPES KOH, pH 7.4, 10 mM sodium pyrophosphate, 2 mM EDTA, 1% Triton X-100), centrifuged for 10 min at 15,000*g* and cleared lysates were incubated with anti-Flag M2 affinity gel (Sigma A2220) for 2 h at 4 °C. The immunoprecipitates were washed once with 1 ml of lysis buffer and three times with lysis buffer including 200 mM NaCl. After the last wash, the immunoprecipitates were boiled in 40 μl of SDS sample buffer. The immunoprecipitates (15 μl) were used for immunodetection using anti-Flag (Sigma F1804) and anti-MICU1 (Abcam).

ARTICLE

doi:10.1038/nature10236

Structure of the human histamine H_1 receptor complex with doxepin

Tatsuro Shimamura[1,2,3]*, Mitsunori Shiroishi[1,2,4]*, Simone Weyand[1,5,6], Hirokazu Tsujimoto[1,2], Graeme Winter[6], Vsevolod Katritch[7], Ruben Abagyan[7], Vadim Cherezov[3], Wei Liu[3], Gye Won Han[3], Takuya Kobayashi[1,2], Raymond C. Stevens[3] & So Iwata[1,2,5,6,8]

The biogenic amine histamine is an important pharmacological mediator involved in pathophysiological processes such as allergies and inflammations. Histamine H_1 receptor (H_1R) antagonists are very effective drugs alleviating the symptoms of allergic reactions. Here we show the crystal structure of the H_1R complex with doxepin, a first-generation H_1R antagonist. Doxepin sits deep in the ligand-binding pocket and directly interacts with Trp 428$^{6.48}$, a highly conserved key residue in G-protein-coupled-receptor activation. This well-conserved pocket with mostly hydrophobic nature contributes to the low selectivity of the first-generation compounds. The pocket is associated with an anion-binding region occupied by a phosphate ion. Docking of various second-generation H_1R antagonists reveals that the unique carboxyl group present in this class of compounds interacts with Lys 191$^{5.39}$ and/or Lys 179^{ECL2}, both of which form part of the anion-binding region. This region is not conserved in other aminergic receptors, demonstrating how minor differences in receptors lead to pronounced selectivity differences with small molecules. Our study sheds light on the molecular basis of H_1R antagonist specificity against H_1R.

Histamine is a biogenic amine and an important mediator in various physiological and pathophysiological conditions such as arousal state, allergy and inflammation[1–3]. Histamine exerts its effects through the activation of four distinct histamine receptors (H_1, H_2, H_3 and H_4) that belong to the G-protein-coupled-receptor (GPCR) superfamily. The H_1R, originally cloned from bovine H_1R[4], is now known to be expressed in various human tissues including airway, intestinal and vascular smooth muscle and brain[2]. In type I hypersensitivity allergic reactions, H_1R is activated by histamine released from mast cells, which are stimulated by various antigens[5]. Many studies have been performed to develop H_1R antagonists, known generally as antihistamines. Many of these compounds inhibit the action of histamine on H_1R to alleviate the symptoms of allergic reactions, making H_1R one of the most validated drug targets judging from the number of drugs approved[6]. H_1R shows constitutive activity, and H_1R antagonists generally act as inverse agonists for H_1R[7,8]. Development of H_1R antagonists has progressed through two generations. First-generation drugs such as pyrilamine and doxepin (Supplementary Fig. 1) are effective H_1R antagonists. These compounds are, however, known to show considerable side effects such as sedation, dry mouth and arrhythmia, because of penetration across the blood–brain barrier and low receptor selectivity. These H_1R antagonists can bind not only to H_1R but also to other aminergic GPCRs, monoamine transporters and cardiac ion channels. Second-generation drugs such as cetirizine and olopatadine (Supplementary Fig. 1) are less sedating and in general have fewer side effects. The improved pharmacology of the second-generation zwitterionic drugs can be attributed to a new carboxylic moiety, in combination with the protonated amine, which significantly reduces brain permeability, although residual central nervous system effects are still reported[9]. The introduction of the carboxyl moiety also improves the H_1R selectivity of these compounds, but certain second-generation H_1R antagonists—such as terfenadine—still show cardiotoxicity because of their interaction with cardiac potassium channels[10,11].

A first-generation H_1R antagonist, doxepin, can cause many types of side effects due to its antagonistic effects on histamine H_2 (ref. 12), serotonin 5-HT_2, α_1-adrenergic, and muscarinic acetylcholine receptors[13] in addition to the inhibition of the reuptake of serotonin and noradrenaline[14]. Although even raw homology models of GPCRs may facilitate discovery of novel ligands[15,16], reliable receptor structures are essential to improve the reliability of the predictions and understand the structural basis of subtype specificity. Recently determined GPCR structures have enabled structure-based approaches to modelling ligand interactions in the binding pocket[17–23] and are already yielding novel chemotypes predicted by virtual screening of large chemical libraries[24,25]. Here we report the 3.1 Å resolution structure of the H_1R–T4-lysozyme fusion protein (H_1R–T4L) complex with doxepin. The crystal structure reveals the atomic details of doxepin binding and its inverse agonistic activity. The H_1R crystal structure and models of second-generation H_1R antagonists will be highly beneficial for guiding rational design of ligands that do not penetrate the blood–brain barrier while maintaining H_1 selectivity.

Overall architecture of H_1R

In the H_1R construct, T4L[26] was inserted into the third cytoplasmic loop (intracellular loop 3 (ICL3)) (Gln 222–Gly 404) and 19 residues were truncated from the amino-terminal region (Met 1–Lys 19) (see Methods). H_1R–T4L showed similar binding affinities for H_1R antagonists and for histamine as the wild-type H_1R expressed in yeast cells (Supplementary Table 1) and in COS-7 cells[27]. The structure of the H_1R–T4L crystals obtained in the lipidic cubic phase (see Methods)

[1]Human Receptor Crystallography Project, ERATO, Japan Science and Technology Agency, Yoshidakonoe-cho, Sakyo-ku, Kyoto 606-8501, Japan. [2]Department of Cell Biology, Graduate School of Medicine, Kyoto University, Yoshidakonoe-cho, Sakyo-Ku, Kyoto 606-8501, Japan. [3]Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA. [4]Graduate School of Pharmaceutical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan. [5]Division of Molecular Biosciences, Membrane Protein Crystallography Group, Imperial College, London SW7 2AZ, UK. [6]Diamond Light Source, Harwell Science and Innovation Campus, Chilton, Didcot, Oxfordshire OX11 0DE, UK. [7]Skaggs School of Pharmacy and Pharmaceutical Sciences and San Diego Supercomputer Center, University of California, San Diego, La Jolla, California 92093, USA. [8]Systems and Structural Biology Center, RIKEN, 1-7-22 Suehiro-cho Tsurumi-ku, Yokohama 230-0045 Japan.

*These authors contributed equally to this work.

was determined in complex with the H_1R antagonist doxepin at 3.1 Å resolution (Supplementary Table 2).

H_1R is structurally most similar to the aminergic receptors (Fig. 1): β_2-adrenergic (β_2-AR)[18], β_1-adrenergic (β_1-AR)[19] and dopamine D3 (D3R)[23] receptors, while having larger deviations from the more phylogenetically distant rhodopsin[17,21], A_{2A} adenosine receptor (A_{2A}AR)[20] and chemokine receptor CXCR4 (ref. 22) (Supplementary Table 3). H_1R also shares common motifs with other GPCRs including D(E)RY in helix III, CWxP in helix VI and NPxxY in helix VII, as well as a disulphide bond connecting extracellular loop 2 (ECL2) with the extracellular end of helix III (Cys $100^{3.25}$ to Cys 180; superscripts indicate residue numbers as per the Ballesteros–Weinstein scheme[28]) but lacks the palmitoylation site at the end of helix VIII found in many other GPCRs[29].

Previous GPCR structures revealed that not only the residues in the transmembrane segments but also those in the loops are critical for ligand specificity[17–23]. ECL2 connecting helices IV and V is attached to helix III through a disulphide bond between Cys 180 in ECL2 and Cys $100^{3.25}$ in helix III. Seven residues (Phe 168–Val 174) before the disulphide are not included in the structure, as they did not have interpretable densities. A section of ECL2, between the disulphide bridge and the extracellular end of helix V, is particularly important because it is located at the entrance to the ligand-binding pocket. This section of ECL2 contains 7 amino acids in H_1R, as compared to 5 in β_2-AR, 4 in D3R, and 8 in A_{2A}AR. The extra length of this ECL2 section is apparently accommodated by the increased distance between the extracellular ends of helices III and V by ~1.5 Å and ~3.1 Å when compared to β_2-AR and D3R, respectively (Fig. 1b, c). This creates more space within the ligand-binding pocket, which can now accommodate the larger second-generation H_1R antagonists, as discussed later.

Some unique features are also observed in the transmembrane segments. A conserved Pro $161^{4.59}$-induced kink in helix IV forms a tight i+3 helical turn, instead of i+4 as in β_2-AR and D3R (Fig. 2a). This tighter turn allows accommodation of a bulky Trp side chain at position 4.56, which seems to be essential for ligand specificity of aminergic GPCRs because this position is occupied by Ser in β_2-AR and D3R, and mutations of this Trp in guinea-pig H_1R to Ala, Met and Phe reduce affinity against the antagonist pyrilamine[30].

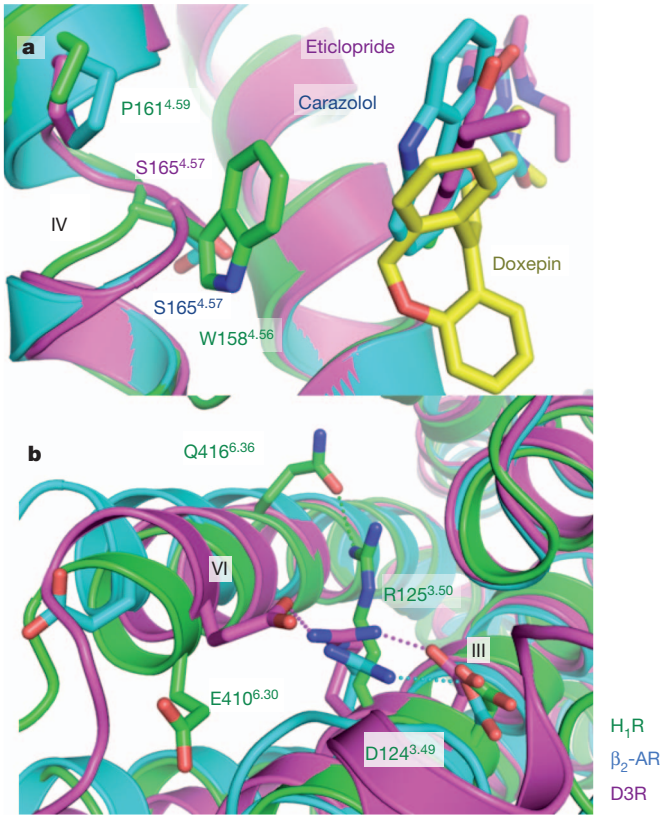


Figure 2 | Comparison of the structures of H_1R, β_2-AR and D3R. **a**, Proline-induced kink in helix IV (H_1R, green; β_2-AR, cyan; D3R, magenta). The side chain of Trp $158^{4.56}$ and Pro $161^{4.59}$ of H_1R and the equivalent residues of β_2-AR (Ser $165^{4.57}$ and Pro $168^{4.60}$) and of D3R (Ser $165^{4.57}$ and Pro $167^{4.59}$) are also shown. **b**, Variations in the D(E)RY motif structures of H_1R, β_2-AR and D3R are coloured in green, cyan and magenta, respectively. Side chains of Asp $124^{3.49}$, Arg $125^{3.50}$, Glu $410^{6.30}$ of H_1R and the equivalents in β_2-AR and D3R are represented as stick models. For H_1R, Gln $416^{6.36}$, which forms a hydrogen bond with Arg $125^{3.50}$, is also shown. Possible hydrogen bonds are indicated by dotted lines.

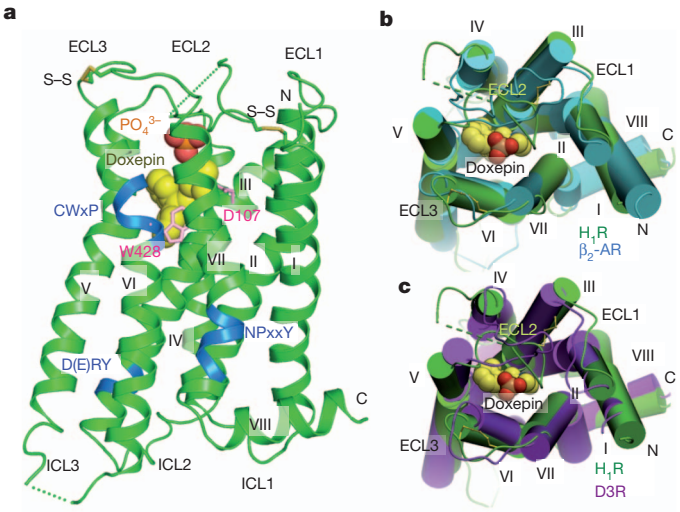


Figure 1 | Structure of H_1R complex with doxepin. **a**, Ribbon representation of the H_1R structure. Doxepin is shown as yellow spheres whereas the phosphate ion is shown as spheres with the carbon and oxygen atoms coloured orange and red, respectively. Disulphide bonds are shown as yellow sticks, and Trp 428 and Asp 107 are shown as pink sticks. Three conserved motifs D(E)R$^{3.50}$Y, CWxP$^{6.50}$ and NP$^{7.50}$xxY are highlighted in blue. **b**, Superimposition of the H_1R (green) and β_2-AR (cyan) structures. **c**, Same as **b** but with the D3R structure coloured violet.

The 'ionic lock', a salt bridge between Arg$^{3.50}$ in the conserved D(E)R$^{3.50}$Y motif and Asp/Glu$^{6.30}$, which is suggested to stabilize the inactive conformation, was observed in rhodopsin structures[17,21] and D3R[23], but was broken in all the other GPCRs[18–20,22]. In H_1R, Arg $125^{3.50}$ of the D(E)R$^{3.50}$Y motif does not form a salt bridge either with Glu $410^{6.30}$ or with Asp $124^{3.49}$. Instead, the side chain of Arg $125^{3.50}$ adopts in a new conformer relative to previous structures, forming a hydrogen bond to Gln $416^{6.36}$ in helix VI (Fig. 2b). Different structures of the ionic lock regions of the receptors could be caused by modifications of ICL3. Otherwise, they might be related to the different levels of constitutive activities of the receptors.

Doxepin isomers and conformers

The doxepin used in this study contains a mixture of E and Z isomers, and each isomer can take two distinct rotational conformers of the dibenzo[b,e]oxepin ring, resulting in four distinct conformers (conformers 1–4; Supplementary Fig. 2). Two conformers, one E isomer (conformer 1) and one Z isomer (conformer 4), fit the electron density better than the other two (Supplementary Fig. 3). This result is also consistent with the R_{free} and the averaged *B*-factor values for each conformer (Supplementary Table 4). A 1:1 mixture of the E and Z isomers was used in the refinement. The two conformers are indistinguishable at this resolution and have nearly identical interactions

with the binding pocket; consequently, in the following sections we are referring to the E isomer, unless noted otherwise.

Ligand–binding pocket

Doxepin binds in a pocket mainly defined by the side chains of helices III, V and VI (Fig. 3a, b). Asp $107^{3.32}$, a strictly conserved residue in aminergic receptors (Supplementary Table 5), forms an anchor salt bridge with the amine moiety of the ligand. This interaction has been reported to be essential for the binding of H_1R antagonists as well as agonists in mutational studies[31–33]. This amine moiety is connected via a flexible carbon chain to the tricyclic dibenzo[b,e]oxepin ring in a hydrophobic pocket comprised of the side chains of helices III, V and VI. The tricyclic ring of doxepin sits much deeper (by ~5 Å) in the binding pocket than the ligands in the other non-rhodopsin GPCR structures (Fig. 3c). The ligand is surrounded mainly by highly conserved residues among aminergic receptors, including Ile $115^{3.40}$, Phe $424^{6.44}$, Trp $428^{6.48}$ and Phe $432^{6.52}$, whereas the non-conserved residues Trp $158^{4.56}$ and Asn $198^{5.46}$ in the pocket make only minor hydrophobic interactions with doxepin (Fig. 3a, b). The importance of a large side chain at position 6.52 has been suggested for the binding of pyrilamine[30,33]. Thr $112^{3.37}$ can form a hydrogen bond to the oxygen atom of the E isomer (but not the Z isomer) of doxepin, as shown in Fig. 3a, b. A suboptimal geometry and the bifurcated nature of this hydrogen bond indicate that it does not contribute significantly to binding affinity, which is supported by lack of this hydrogen-bond interaction in the H_1R–olopatadine complex, as described later. This well-conserved pocket and its mostly hydrophobic nature should contribute to low selectivity of doxepin and other first-generation H_1R-antagonists[13,32]. Moreover, because of its deep binding position, doxepin does not interact with ECL2, whose highly variable primary and tertiary structures are known to contribute to binding specificity of GPCR ligands[34].

A novel feature of the H_1R–doxepin complex is the existence of an anion-binding site at the entrance to the ligand-binding pocket (Fig. 3d). We modelled a phosphate ion, which is present at a high concentration in the crystallization buffer (300 mM ammonium phosphate), into the observed strong density in the site. This model is supported by the fact that a phosphate ion affects the binding of some ligands and the stability of H_1R (Supplementary Tables 1 and 6). The phosphate ion is coordinated by Lys 179^{ECL2}, Lys $191^{5.39}$, Tyr $431^{6.51}$ and His $450^{7.35}$; all of which, except for Tyr $431^{6.51}$, are unique to H_1R (Supplementary Table 5). This encasement of the ligand in the pocket combined with an ionic interaction between the phosphate ion and the tertiary amine of doxepin (N–O distance 4.8 Å) indicates that a phosphate ion may serve as a positive modulator of ligand binding. This hypothesis has been validated by comparing thermostability (Supplementary Table 6) and ligand affinity (Supplementary Table 1) in buffers with and without phosphate. Thermostability of the receptor is increased in the presence of phosphate for all ligands except for cetirizine, which can prevent the phosphate binding, according to the modelling study discussed later. The phosphate effect is observed at a concentration as low as 1.5 mM,

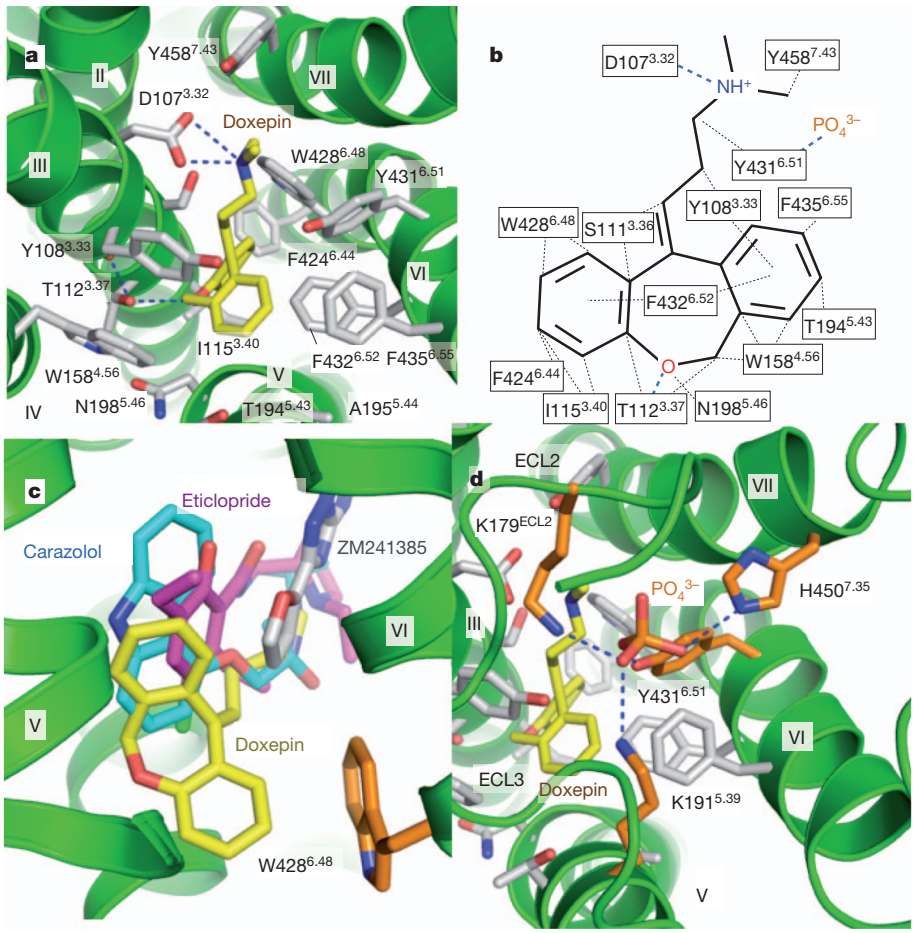


Figure 3 | Binding interactions of doxepin. **a**, Doxepin is shown as sticks with yellow carbon atoms, whereas the contact residues within 4 Å are shown with grey carbon atoms. Nitrogen and oxygen atoms are coloured blue and red, respectively. Hydrogen bonds/salt bridges are indicated as blue dotted lines. **b**, Doxepin binding interactions. Hydrophobic interactions are shown in black dotted lines. **c**, Ligand-binding positions in non-rhodopsin GPCRs. Carbon atoms of doxepin (H_1R) are shown in yellow, carazolol (β_2-AR) in cyan, eticlopride (D3R) in magenta and ZM241385 ($A_{2A}AR$) in grey. **d**, Structure of the anion-binding region with a phosphate ion.

indicating its physiological relevance. The affinity of histamine and pyrilamine to the receptors also increased in the presence of phosphate.

H_1 selectivity of H_1R antagonists

Supplementary Fig. 1 lists the first and second generations of H_1R antagonists. It has been shown that second-generation H_1R antagonists are much more specific to H_1R and show much lower affinity to the other aminergic receptors[32,35]. H_1R-antagonist specificity has been previously analysed using H_1R-homology models based on the bacteriorhodopsin or bovine rhodopsin crystal structure in combination with the H_1R antagonist pharmacophore model and mutational studies[30,36,37]. These studies have successfully determined some residues important for selectivity, including Lys $191^{5.39}$; however, contributions of the ECL residues have not been examined because these loops could not be modelled accurately based on the bacteriorhodopsin or bovine rhodopsin structure. Our H_1R structure with the extracellular loops should significantly improve understanding of H_1R-antagonist selectivity. Using flexible ligand-receptor docking[38,39] in the ICM molecular modelling package[40] (also see Methods), we have studied H_1R selectivity for representative second-generation zwitterionic H_1R-antagonists: olopatadine, acrivastine, *R*-cetirizine (levocetirizine) and fexofenadine (Fig. 4). Olopatadine (Fig. 4a) is a close doxepin analogue with a methyl–carboxyl substitution in one of its benzene rings. Its binding mode closely resembles doxepin, while the carboxyl group extends out of the pocket towards the extracellular space and interacts with Lys $191^{5.39}$ and Tyr $108^{3.33}$ without displacing the phosphate ion. These additional interactions can explain a reduced effect of the mutation of the conserved Asp $107^{3.32}$ to Ala on olopatadine binding (14-fold for olopatadine as compared to 280-fold for doxepin)[32,41]. The orientation of the carboxyl moiety in the ECL region dictates that the oxygen atom of the dibenzo[b,e]oxepin ring is in a position where it cannot form a hydrogen bond with Thr $112^{3.37}$. Although the marketed drug is only the Z isomer, both olopatadine Z and E isomers show similar H_1R affinities[41].

Acrivastin (Fig. 4b) has a different chemical scaffold with a carboxyl group in its pyridine ring. Its longer carbon chain positions the carboxyl group higher in the ECL region, where it can form salt bridges to both Lys $191^{5.39}$ and Lys 179^{ECL2} amine moieties. *R*-cetirizine (Fig. 4c) has its carboxylic moiety attached directly to a piperazine amino group. The conformational modelling indicates that the carboxyl moiety can reach towards the ECL region, forming salt bridges to Lys $191^{5.39}$ and Lys 179^{ECL2}. Finally, fexofenadine (*R* isomer; Fig. 4d) has the most extended carboxyl-containing substituent,

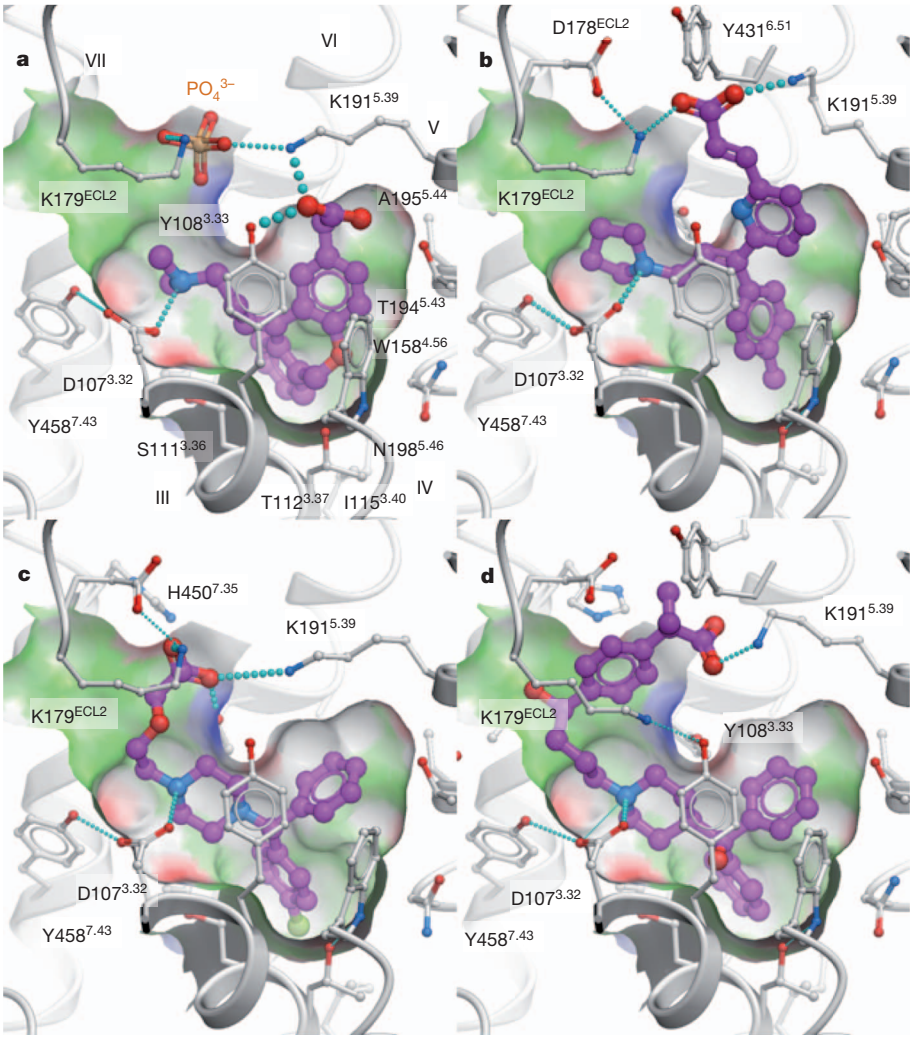


Figure 4 | Interactions of second-generation selective H_1R antagonists with the H_1R ligand-binding pocket. **a–d**, Conformation of each complex was predicted by global optimization of the ligand in the all-atom flexible H_1R model[38–40] based on the H_1R–doxepin complex structure. Carbon atoms for Z-olopatadine co-bound with the phosphate ion (**a**), acrivastine (**b**), *R*-cetirizine (levocetirizine) (**c**) and fexofenadine (**d**) are coloured violet. Nitrogen and oxygen atoms are coloured blue and red, respectively. Ligand contact residues of H_1R are shown with grey carbon atoms; parts of helices III, IV and ECL2 are not shown for clarity. Hydrogen bonds are shown in cyan.

which reaches outside of the binding cavity and forms a salt bridge to Lys 191$^{5.39}$.

Modelling of second-generation H_1R-antagonist binding to H_1R indicates that no significant protein backbone rearrangements are required to accommodate these diverse ligands. The enhanced H_1R selectivity of these compounds[32,35] can be explained by the specific interaction of the carboxyl group with Lys residues in the ECL region, unique to H_1R. The result also shows a good agreement with earlier modelling and site-directed mutagenesis studies. Lys 191$^{5.39}$ is known to be important for increasing affinity for some of these ligands[30,42,43], whereas the involvement of Lys 179^{ECL2} was suggested in the modelling study of 8R-lisuride into the ligand-binding pocket[44]. Our modelling results also indicate that olopatadine is the only second-generation compound studied here for which the carboxyl moiety does not interfere with phosphate binding. The results are also supported by the fact that the presence of the phosphate ion increased the thermal stability of the H_1R–doxepin or H_1R–olopatadine complexes, whereas it did not affect the stability of the H_1R–cetirizine complex (Supplementary Table 6).

Mechanism of H_1R inactivation

H_1R antagonists act as highly effective inverse agonists of H_1R, which reduce basal activity of the receptor and therefore are expected to interfere with the key molecular switches involved in the GPCR activation mechanism. One of the switches is represented by Trp$^{6.48}$ of the conserved CWxP$^{6.50}$ motif, which helps to stabilize rhodopsin in its inactive dark state through a direct interaction with retinal. The recently published structure of the active-state A_{2A}AR[45] also showed that Trp$^{6.48}$ participates in the activation-related conformational changes, where a small ligand-induced shift of Trp$^{6.48}$ was observed in concert with the large movement of the intracellular part of helix VI. In other receptors, the role of Trp$^{6.48}$ is less obvious; for instance, it lacks direct ligand interactions with either inverse agonists or full agonists of β_2-AR[46]. It is interesting to note that in the H_1R structure, like in inactive rhodopsin, the H_1R-antagonist doxepin does make extensive hydrophobic interactions with the Trp 428$^{6.48}$ rings, which is unique among the known non-rhodopsin GPCR structures and could stabilize the hydrophobic packing around helix VI (Fig. 3c). Another important ligand-induced switch described in β_2-AR is activation-related contraction of the extracellular ligand-binding pocket[37]. Because the natural agonist histamine is much smaller than bulky H_1R antagonists, some contraction of the binding pocket is likely to accompany ligand-induced H_1R activation. Bulky compounds, capable of blocking both activation-related contraction of the pocket and the Trp 428$^{6.48}$ switch would be very efficient in locking H_1R in an inactive conformation, which is likely to explain the as much as 78% reduction of H_1R basal activity by some H_1R antagonists[8].

METHODS SUMMARY

H_1R–T4L was expressed in yeast *Pichia pastoris*. Ligand-binding assays were performed as described in Methods. *P. pastoris* membranes were solubilized using 1% (w/v) n-dodecyl-B-D-maltopyranoside and 0.2% (w/v) cholesteryl hemisuccinate, and purified by immobilized metal ion affinity chromatography (IMAC). After IMAC, the C-terminal GFP was cleaved by Tobacco Etch virus (TEV) protease. Then the sample mixture was passed through IMAC to remove the cleaved His-tagged GFP and TEV protease. Receptor crystallization was performed by the lipidic cubic phase (LCP) method. The protein-LCP mixture contained 40% (w/w) receptor solution, 54% (w/w) monoolein, and 6% (w/w) cholesterol. Crystals were grown in 40–50 nl protein-laden LCP boluses overlaid by 0.8 µl of precipitant solution (26–30% (v/v) PEG400, 300 mM ammonium phosphate, 10 mM $MgCl_2$, 100 mM Na-citrate pH 4.5 and 1 mM doxepin) at 20 °C. Crystals were harvested directly from LCP matrix and flash frozen in liquid nitrogen. X-ray diffraction data were collected at 100 K with a beam size of 10×10 µm on the microfocus beamline I24 at the Diamond Light Source (UK). Data collection, processing, structure solution and refinement are described in Methods.

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

Received 15 March; accepted 1 June 2011.
Published online 22 June 2011.

1. Schwartz, J. C., Arrang, J. M., Garbarg, M., Pollard, H. & Ruat, M. Histaminergic transmission in the mammalian brain. *Physiol. Rev.* **71,** 1–51 (1991).
2. Hill, S. J. Distribution, properties, and functional characteristics of three classes of histamine receptor. *Pharmacol. Rev.* **42,** 45–83 (1990).
3. Hill, S. J. *et al.* International Union of Pharmacology. XIII. Classification of histamine receptors. *Pharmacol. Rev.* **49,** 253–278 (1997).
4. Yamashita, M. *et al.* Expression cloning of a cDNA encoding the bovine histamine H1 receptor. *Proc. Natl Acad. Sci. USA* **88,** 11515–11519 (1991).
5. Simons, F. E. Advances in H_1-antihistamines. *N. Engl. J. Med.* **351,** 2203–2217 (2004).
6. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nature Rev. Drug Discov.* **5,** 993–996 (2006).
7. Bakker, R. A., Wieland, K., Timmerman, H. & Leurs, R. Constitutive activity of the histamine H_1 receptor reveals inverse agonism of histamine H_1 receptor antagonists. *Eur. J. Pharmacol.* **387,** R5–R7 (2000).
8. Bakker, R. A., Schoonus, S. B., Smit, M. J., Timmerman, H. & Leurs, R. Histamine H_1-receptor activation of nuclear factor-κB: roles for Gβγ- and $G\alpha_{q/11}$-subunits in constitutive and agonist-mediated signaling. *Mol. Pharmacol.* **60,** 1133–1142 (2001).
9. Tashiro, M. *et al.* Dose dependency of brain histamine H_1 receptor occupancy following oral administration of cetirizine hydrochloride measured using PET with [^{11}C]doxepin. *Hum. Psychopharmacol.* **24,** 540–548 (2009).
10. Woosley, R. L., Chen, Y., Freiman, J. P. & Gillis, R. A. Mechanism of the cardiotoxic actions of terfenadine. *J. Am. Med. Assoc.* **269,** 1532–1536 (1993).
11. Yap, Y. G. & Camm, A. J. Potential cardiac toxicity of H_1-antihistamines. *Clin. Allergy Immunol.* **17,** 389–419 (2002).
12. Okamura, N. *et al.* Functional neuroimaging of cognition impaired by a classical antihistamine, *d*-chlorpheniramine. *Br. J. Pharmacol.* **129,** 115–123 (2000).
13. Cusack, B., Nelson, A. & Richelson, E. Binding of antidepressants to human brain receptors: focus on newer generation compounds. *Psychopharmacology (Berl.)* **114,** 559–565 (1994).
14. Sarker, S. *et al.* The high-affinity binding site for tricyclic antidepressants resides in the outer vestibule of the serotonin transporter. *Mol. Pharmacol.* **78,** 1026–1035 (2010).
15. Klabunde, T. & Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *ChemBioChem* **3,** 928–944 (2002).
16. de Graaf, C., Rognan, D. & Customizing, G. Protein-coupled receptor models for structure-based virtual screening. *Curr. Pharm. Des.* **15,** 4026–4048 (2009).
17. Palczewski, K. *et al.* Crystal structure of rhodopsin: a G protein-coupled receptor. *Science* **289,** 739–745 (2000).
18. Cherezov, V. *et al.* High-resolution crystal structure of an engineered human β2-adrenergic G protein-coupled receptor. *Science* **318,** 1258–1265 (2007).
19. Warne, T. *et al.* Structure of a β_1-adrenergic G-protein-coupled receptor. *Nature* **454,** 486–491 (2008).
20. Jaakola, V. P. *et al.* The 2.6 angstrom crystal structure of a human A_{2A} adenosine receptor bound to an antagonist. *Science* **322,** 1211–1217 (2008).
21. Murakami, M. & Kouyama, T. Crystal structure of squid rhodopsin. *Nature* **453,** 363–367 (2008).
22. Wu, B. *et al.* Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **330,** 1066–1071 (2010).
23. Chien, E. Y. *et al.* Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **330,** 1091–1095 (2010).
24. Kolb, P. *et al.* Structure-based discovery of β_2-adrenergic receptor ligands. *Proc. Natl Acad. Sci. USA* **106,** 6843–6848 (2009).
25. Katritch, V. *et al.* Structure-based discovery of novel chemotypes for adenosine A_{2A} receptor antagonists. *J. Med. Chem.* **53,** 1799–1809 (2010).
26. Rosenbaum, D. M. *et al.* GPCR engineering yields high-resolution structural insights into β2-adrenergic receptor function. *Science* **318,** 1266–1273 (2007).
27. Ratnala, V. R. *et al.* Large-scale overproduction, functional purification and ligand affinities of the His-tagged human histamine H1 receptor. *Eur. J. Biochem.* **271,** 2636–2646 (2004).
28. Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure–function relations in G protein-coupled receptors. *Methods Neurosci.* **25,** 366–428 (1995).
29. Qanbar, R. & Bouvier, M. Role of palmitoylation/depalmitoylation reactions in G-protein-coupled receptor function. *Pharmacol. Ther.* **97,** 1–33 (2003).
30. Wieland, K. *et al.* Mutational analysis of the antagonist-binding site of the histamine H_1 receptor. *J. Biol. Chem.* **274,** 29994–30000 (1999).
31. Ohta, K. *et al.* Site-directed mutagenesis of the histamine H_1 receptor: roles of aspartic acid^{107}, asparagine^{198} and threonine^{194}. *Biochem. Biophys. Res. Commun.* **203,** 1096–1101 (1994).
32. Nonaka, H. *et al.* Unique binding pocket for KW-4679 in the histamine H_1 receptor. *Eur. J. Pharmacol.* **345,** 111–117 (1998).
33. Bruysters, M. *et al.* Mutational analysis of the histamine H_1-receptor binding pocket of histaprodifens. *Eur. J. Pharmacol.* **487,** 55–63 (2004).
34. Peeters, M. C., van Westen, G. J., Li, Q. & Ijzerman, A. P. Importance of the extracellular loops in G protein-coupled receptors for ligand recognition and receptor activation. *Trends Pharmacol. Sci.* **32,** 35–42 (2011).
35. Gillard, M. *et al.* H_1 antagonists: receptor affinity versus selectivity. *Inflamm. Res.* **52** (Suppl. 1), S49–S50 (2003).

36. Kiss, R., Kovari, Z. & Keseru, G. M. Homology modelling and binding site mapping of the human histamine H1 receptor. *Eur. J. Med. Chem.* **39,** 959–967 (2004).
37. Jongejan, A. & Leurs, R. Delineation of receptor–ligand interactions at the human histamine H1 receptor by a combined approach of site-directed mutagenesis and computational techniques—or—how to bind the H1 receptor. *Arch. Pharm. (Weinheim)* **338,** 248–259 (2005).
38. Totrov, M. & Abagyan, R. Flexible protein–ligand docking by global energy optimization in internal coordinates. *Proteins* (Suppl. 1) **29,** 215–220 (1997).
39. Katritch, V. *et al.* Analysis of full and partial agonists binding to β2-adrenergic receptor suggests a role of transmembrane helix V in agonist-specific conformational changes. *J. Mol. Recognit.* **22,** 307–318 (2009).
40. Katritch, V., Kufareva, I. & Abagyan, R. Structure based prediction of subtype-selectivity for adenosine receptor antagonists. *Neuropharmacology* **60,** 108–115 (2011).
41. Matsumoto, Y., Funahashi, J., Mori, K., Hayashi, K. & Yano, H. The noncompetitive antagonism of histamine H1 receptors expressed in Chinese hamster ovary cells by olopatadine hydrochloride: its potency and molecular mechanism. *Pharmacology* **81,** 266–274 (2008).
42. Gillard, M., Van Der Perren, C., Moguilevsky, N., Massingham, R. & Chatelain, P. Binding characteristics of cetirizine and levocetirizine to human H_1 histamine receptors: contribution of Lys^{191} and Thr^{194}. *Mol. Pharmacol.* **61,** 391–399 (2002).
43. Leurs, R., Smit, M. J., Meeder, R., Ter Laak, A. M. & Timmerman, H. $Lysine^{200}$ located in the fifth transmembrane domain of the histamine H_1 receptor interacts with histamine but not with all H_1 agonists. *Biochem. Biophys. Res. Commun.* **214,** 110–117 (1995).
44. Bakker, R. A. *et al.* 8R-lisuride is a potent stereospecific histamine H_1-receptor partial agonist. *Mol. Pharmacol.* **65,** 538–549 (2004).
45. Xu, F. *et al.* Structure of an agonist-bound human A_{2A} adenosine receptor. *Science* **332,** 322–327 (2011).
46. Rasmussen, S. G. *et al.* Structure of a nanobody-stabilized active state of the β_2 adrenoceptor. *Nature* **469,** 175–180 (2011).

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

Acknowledgements This work was supported by the ERATO Human Receptor Crystallography Project from the Japan Science and Technology Agency and by the Targeted Proteins Research Program of MEXT (S.I.), Japan; the NIH Common Fund grant P50 GM073197 for technology development (R.C.S.) and NIH PSI:Biology grant U54 GM094618 (R.C.S., V.C., V.K. and R.A.); R.A. was also partly funded by NIH R01 GM071872. The work was also partly funded by the Biotechnology and Biological Sciences Research Council (BBSRC) BB/G023425/1 (S.I.), Grant-in-Aid for challenging Exploratory Research (T.S.), the Mochida Memorial Foundation for Medical and Pharmaceutical Research (T.S. and T.K.), Takeda Scientific Foundation (M.S.) and the Sumitomo Foundation (T.K.). A part of the work was performed in the Membrane Protein Laboratory funded by the Wellcome Trust (grant 062164/ Z/00/Z) at the Diamond Light Source Limited and at The Scripps Research Institute. We thank D. Axford, R. Owen and G. Evans for help with data collection at I24 of the Diamond Light Source Limited, H. Wu for help with the preparation of Supplementary Fig. 1 and Q. Xu for help on validation of data processing and A. Walker for assistance with manuscript preparation. The authors acknowledge Y. Zheng (The Ohio State University) and M. Caffrey, Trinity College (Dublin, Ireland), for the loan of the in meso robot (built with support from the NIH (GM075915), the National Science Foundation (IIS0308078), and Science Foundation Ireland (02-IN1-B266)). S.I. is thankful for the help of L. E. Johnson, a co-founder of the Diamond-MPL Project and R. Tanaka, the technical coordinator of the ERATO Human Receptor Crystallography Project.

Author Contributions T.S. purified and crystallized the receptor in LCP, optimized crystallization conditions, grew crystals for data collection, solved and refined the structure, and prepared the manuscript. M.S. designed, characterized and screened the constructs, purified the receptor, and prepared the manuscript. S.W. and S.I. collected the data and processed diffraction data with G.W.H. H.T. expressed the receptor, prepared the membrane, and performed the ligand-binding assay. V.K. and R.A. performed flexible ligand-receptor docking, and prepared the manuscript. V.C. assisted with the crystallization in LCP and prepared the manuscript. W.L. performed the thermal stability assay and assisted with the crystallization in LCP. G.W.H. refined the structure and assisted with preparing the manuscript. T.K. designed the receptor production strategy and assisted with preparing the manuscript. R.C.S. and S.I. were responsible for the overall project strategy and management and wrote the manuscript.

Author Information The coordinates and the structure factors have been deposited in the Protein Data Bank under the accession code 3RZE. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.C.S. (stevens@scripps.edu), S.I. (so_iwata@me.com) or T.K. (t-coba@mfour.med.kyoto-u.ac.jp).

METHODS

Construction of the H_1R expression vectors for *Pichia pastoris*. The coding sequence of the full-length human histamine H_1 receptor (H_1R-fl), in which N-linked glycosylation sites (Asn 5 and Asn 18) were mutated to glutamines, was synthesized with optimization of codon usage for *P. pastoris* (TAKARA Bio), and cloned into the pPIC9K expression vector (Invitrogen). The H_1R–T4L construct with an N-terminal 19-residue deletion and insertion of cysteine-less (C54T, C97A) T4 lysozyme into the third intracellular loop was generated by the yeast homologous recombination technique in *Saccharomyces cerevisiae* with the SmaI linearized plasmid pDDGFP2 (ref. 47) and three PCR products with ~30 bp overlapping sequences. The three fragments were individually generated by standard PCR techniques with the indicated primers. The generated plasmid integrating H_1R–T4L followed by TEV cleavage sequence (ENLYFQG), yeast enhanced GFP and octa-histidine tag (H_1R–T4L–GFP) was isolated from *S. cerevisiae*. Coding regions of the H_1R–T4L–GFP fusions were amplified by PCR using a forward primer containing a BamHI site (5′-CTAGAACTAGTGGATCCACCATG-3′) and a reverse primer containing an EcoRI site (5′-GCTTGATATCGAATTCCTGCAGTTAATG-3′). The PCR products were digested with BamHI and EcoRI, and subcloned into the pPIC9K vector.

Expression and membrane preparation. The PmeI linearized pPIC9K expression vector integrating H_1R-fl–GFP or H_1R–T4L–GFP was then transformed into the *P. pastoris* SMD1163 strain by electroporation (2,000 V, 25 μF, and 600 Ω) using a Gene Pulser I (Bio-Rad). Clone selection was performed on the YPD-agar plate containing 0.1 mg ml^{-1} geneticine. A single colony of *P. pastoris* transformant was inoculated into BMGY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 1.34% (w/v) yeast nitrogen base without amino acids, 0.00004% (w/v) biotin, 1% (w/v) glycerol, 0.1 M phosphate buffer at pH 6.0) at 30 °C with shaking at 250 r.p.m. until an OD_{600} of 2–6 was reached. The cells were harvested by centrifugation. To induce expression, the cell pellet was resuspended to an OD_{600} of 1.0 in BMMY medium (1% (w/v) yeast extract, 2% (w/v) peptone, 1.34% (w/v) yeast nitrogen base without amino acids, 0.00004% (w/v) biotin, 0.5% (v/v) methanol, 0.1 M phosphate buffer at pH 6.0) containing 2.5% (v/v) DMSO at 30 °C. Cells were harvested within 20 to 24 h after induction, and stored at −80 °C. Yeast cells were disrupted with 0.5 mm glass beads in a buffer containing 50 mM HEPES, pH 7.5, 120 mM NaCl, 5%(v/v) glycerol, 2 mM EDTA and EDTA-free protein inhibitor cocktail (Roche). Undisrupted cells and cell debris were separated by centrifugation at 3,000*g*, and yeast membranes were collected by ultracentrifugation at 100,000*g* for 30 min at 4 °C. Washing of the membranes was performed by repeating dounce homogenation and centrifugation in a high salt buffer containing 10 mM HEPES, pH 7.5, 1 M NaCl, 10 mM $MgCl_2$, 20 mM KCl and EDTA-free protease inhibitor cocktail. Prepared membranes were resuspended in a buffer containing 50 mM HEPES pH 7.5, 120 mM NaCl, 20% (v/v) glycerol and EDTA-free protease inhibitor cocktail, and snap-frozen in liquid nitrogen and stored at −80 °C until use. Membrane proteins were quantified using the bicinchoninic acid method (Pierce).

Purification of H_1R–T4L. Membrane suspension containing H_1R–T4L–GFP was thawed and incubated on ice for 30 min in the presence of 5 mM doxepin, 10 mg ml^{-1} iodoacetamide, and EDTA-free protease inhibitor cocktail (Roche). The membrane suspension was poured into the buffer containing 20 mM HEPES pH 7.5, 500 mM NaCl, 1% (w/v) n-dodecyl-β-D-maltopyranoside (DDM, Anatrace), 0.2% (w/v) cholesteryl hemisuccinate (CHS, Sigma), 20% (v/v) glycerol and 2–3 mg ml^{-1} membrane, and stirred gently at 4 °C for 1–2 h. The unsolubilized material was separated by centrifugation at 100,000*g* for 30 min. The supernatant was incubated with TALON IMAC resin (Clontech) overnight. The resin was washed with twenty column volumes of 20 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 0.025% (w/v) DDM, 0.005% (w/v) CHS, 100 μM doxepin and 20 mM imidazole. The protein was eluted with 4 column volumes of 20 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 0.025% (w/v) DDM, 0.005% (w/v) CHS, 500 μM doxepin and 200 mM imidazole. The eluted fractions were concentrated to 2.5 ml with a 100 kDa molecular weight cut-off AmiconUltra (Millipore). Imidazole was removed using PD-10 column (GE healthcare). Then the protein was loaded onto the Ni-Sepharose high performance resin (GE healthcare) (1.5 ml resin for ~10 mg of protein). The resin was washed with 20 column volumes of 20 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 0.025% (w/v) DDM, 0.005% (w/v) CHS, 500 μM doxepin and 20 mM imidazole. The sample was eluted with 3 column volumes of 20 mM HEPES pH 7.5, 500 mM NaCl, 10% glycerol, 0.025% (w/v) DDM, 0.005% (w/v) CHS, 1 mM doxepin and 500 mM imidazole. Imidazole was removed using PD-10 column (GE healthcare). The protein was processed overnight with His-tagged TEV protease (expressed and purified in house). TEV protease and the cleaved His-tagged GFP were removed by passing the sample through the Ni-Sepharose high performance resin. The receptor was concentrated to 30–40 mg ml^{-1} with a 100 kDa molecular weight cut-off Vivaspin concentrator (Vivascience). Protein purity and monodispersity were tested by SDS–PAGE and by size-exclusion chromatography using Superdex 200 (GE healthcare).

Lipidic cubic phase crystallization. Lipidic cubic phase (LCP) crystallization trials were performed using an *in meso* crystallization robot as previously described[48]. Ninety-six-well glass sandwich plates were filled with 40–50 nl protein-laden LCP boluses overlaid by 0.8 μl of precipitant solution in each well and sealed with a glass coverslip. The protein-LCP mixture contained 40% (w/w) receptor solution, 54% (w/w) monoolein, and 6% (w/w) cholesterol. Crystallization set-ups were performed at room temperature (20–22 °C). Plates were incubated and imaged at 20 °C using an automated incubator/imager (RockImager 1000, Formulatrix). Crystals were obtained in 26–30% (v/v) PEG400, 300 mM ammonium phosphate, 10 mM $MgCl_2$, 100 mM Na-citrate pH 4.5 and 1 mM doxepin (Sigma) (Supplementary Fig. 4). Crystals were harvested directly from LCP matrix using MiTeGen micromounts and were flash-frozen in liquid nitrogen without additional cryoprotectant.

Data collection and refinement. X-ray diffraction data were collected at 100 K with a wavelength of 0.97780 Å and with a beamsize of 10 × 10 μm on the microfocus beamline I24 at the Diamond Light Source (UK) with a Pilatus 6M detector. Each loop was subjected to a grid scanning[49] in order to locate the crystals, which are invisible in the LCP once they are mounted. The exact locations and dimensions of the chosen crystals were determined by further grid scanning with a smaller search area. Data collection was carried out by collecting several overlapping wedges of data from adjacent positions within a single crystal. The data were processed initially with xia2[50] using Mosflm[51] and Scala[52] with the merging statistics used to determine an optimum subset of measurements to merge. The final data set consisted of data from five of the eight positions recorded, giving a total of 75 degrees of data. These data were then remerged with Scala to give the final data set summarized in Supplementary Table 2. The space group was determined to be *I*422 with one molecule in the asymmetric unit. Diffraction data were slightly anisotropic, extending to 2.9 Å in the c^* direction and to 3.1 Å in the a^* and b^* directions. The structure factors up to 3.1 Å resolution were anisotropically scaled by PHASER[53] and then used for the subsequent molecular replacement and refinement. The structure was determined by molecular replacement with the program PHASER[53] using two independent search models (polyalanine of the 7 transmembrane α-helices, and T4L) from the β_2-AR (PDB code 2RH1) structure. We chose β_2-AR as a model structure because it has the highest homology of transmembrane helices with H_1R (41.7%) among the human GPCR structures. For the initial map calculation after molecular replacement, however, we used a β_2-AR model without side chains, loops, ligand, lipids and any solvents; therefore, the final H_1R structure is not biased to the β_2-AR structure. This is supported by low R_{work} and R_{free} values (Supplementary Table 2). All refinements were performed with REFMAC5[54] and autoBUSTER[55] followed by manual examination and rebuilding of the refined coordinates in the program Coot[56]. The non-lysozyme portion contains higher *B* factors (116 Å^2) owing to fewer contacts as compared to T4 lysozyme (36 Å^2). Calculation of the surface area buried by crystal contacts also explains this. For the non-lysozyme portion, only 8% (1,225 Å^2) of 15,689 Å^2 solvent-accessible surface area is buried by crystal contacts. In contrast, for the T4 lysozyme portion, 32% (2,733 Å^2) of the solvent accessible area (8,648 Å^2) is buried by crystal interactions. Supplementary Fig. 5 also shows that there are strong interactions between T4 lysozyme domains, but relatively fewer between non-lysozyme domains throughout the crystal packing. Although the average *B* factor of the non-lysozyme domain is high as compared to T4 lysozyme, electron densities were clear for unambiguous model building (Supplementary Figs 3 and 5). The eight H_1R N-terminal residues (Thr 20–Leu 27), two C-terminal residues (Arg 486–Ser 487), and seven residues (Phe 168–Val 174) in the second extracellular loop (ECL2) are not included in the structure, as they did not have interpretable densities.

Strong and spherical electron densities (about 4 sigma) were found in the anion-binding region in the $F_o - F_c$ omit map. We excluded the presence of a water molecule in this region owing to strong residual positive $F_o - F_c$ densities when we modelled it as a water molecule. The coordination geometry in the highly electropositive environment surrounded by His 450$^{7.35}$, Lys 179^{ECL2} and Lys 191$^{5.39}$ implied that either a phosphate or sulphate ion could be modelled. Because ammonium phosphate was added to our crystallization buffer, we modelled it as a phosphate ion. The average *B* factors of the phosphate ion and the interacting atoms are 177 Å^2 and 154 Å^2, respectively.

Ligand-binding assays. For the saturation binding experiment, yeast membrane suspensions containing H_1R-fl–GFP (20 μg) or H_1R–T4L–GFP (5 μg) were incubated with increasing concentrations of [^{3}H] pyrilamine (from 0.15 to 40 nM) in a total assay volume of 200 μl for 1 h at 25 °C. To investigate the effect of phosphate on the ligand binding, assays were performed in PBS buffer pH 7.4 (138 mM NaCl, 8.1 mM Na_2HPO_4, 27 mM KCl, 1.8 mM KH_2PO_4) or in the HEPES buffer containing 20 mM HEPES pH 7.5 and 150 mM NaCl. Nonspecific binding was

determined in the presence of 1,000-times excess unlabelled pyrilamine. Membranes were trapped on Whatman GF/B filters pre-soaked in 0.3% polyethylenimine, and unbound radioligands were washed with 9 ml of the PBS or HEPES buffers. The retained radioactivity was measured on an LCS-5100 liquid scintillation counter (ALOKA) in a Clearzol I scintillation liquid (Nakarai, Japan). Data were analysed by nonlinear curve-fitting with a rectangular hyperbola function using the Prism 4.0 software (GraphPad) to determine dissociation constant (K_d).

For competition binding assays, yeast membrane suspensions containing H_1R-fl–GFP or H_1R–T4L–GFP were incubated with 4 nM or 20 nM [^{3}H]pyrilamine in the PBS buffer or the HEPES buffer in the presence of 10 nM to 100 mM histamine hydrochloride or 0.001 nM to 1 μM doxepin, or 0.01 nM to 10 μM cetirizine, pyrilamine, olopatadine and fexofenadine. Data were analysed by nonlinear curve fitting with a sigmoidal function using the Prism 4.0 to determine the half maximal inhibitory concentrations (IC_{50}). All data shown were calculated based on more than three independent experiments. Inhibition constant K_i was calculated based on the equation $K_i = IC_{50}/(1+L/K_d)$, where L is the concentration of [^{3}H]pyrilamine with the dissociation constant K_d.

Thermal stability assay. *N*-[4-(7-diethylamino-4-methyl-3-coumarinyl) phenyl]maleimide (CPM) dye was purchased from Invitrogen and dissolved in DMSO (Sigma) at 4 mg ml^{-1} as the stock solution for future use. The stock solution was kept at −80 °C and was diluted 1:40 in dye dilution solution (10 mM buffer, 500 mM NaCl, 10% glycerol, 0.025% DDM and 0.005% CHS) before use. The thermal denaturation assay was performed with a total volume of 200 μl sample in a quartz fluorometer cuvette (Starna Cells). H_1R (4 μg) was diluted in the appropriate buffer solution to a final volume of 200 μl. Five microlitres of the diluted dye was added to the protein solution and it was incubated for 30 min at 4 °C. The mixed solution was transferred to the cuvette and the data were collected by a Cary Eclipse spectrofluorometer (Varian) with a temperature ramping rate at 1 °C min^{-1}. The excitation wavelength was 387 nm and the emission wavelength was 463 nm. All assays were performed over a temperature range starting from 20 °C to 80 °C. The stability data were processed with GraphPad Prism program (Graphpad Sofware). To determine the melting temperature (T_m), a Bolzmann sigmoidal equation was used to fit to the data.

Flexible ligand-receptor docking. Docking of ligands was performed using the all-atom flexible receptor docking algorithm in the ICM-Pro molecular modelling package[57] as described previously[39,58]. The initial H_1R model was generated in ICM by building hydrogen atoms for the crystal structure of H_1R. Internal coordinate (torsion) movements were allowed in the side chains of the binding pocket, defined as residues within 8 Å distance of doxepin in the H_1R–doxepin complex. Other side chains and the backbone of the protein were kept as in the crystal structure. An initial conformation for each of the ligands was generated by Cartesian optimization of the ligand model in MMFF force field. Docking was performed by placing the ligand in a random position within 5 Å from the binding pocket and global optimization of the complex conformational energy. The global energy of the complex was calculated as a sum of van der Waals, electrostatic, hydrogen-bonding and torsion stress terms. Stochastic global energy optimization of the complex was performed using the ICM Monte Carlo procedure with minimization[59]. To facilitate side-chain rotamer switches in flexible H_1R models, the first 10^6 steps of the Monte Carlo procedure used 'soft' van der Waals potentials and high Monte Carlo temperature, followed by another 10^6 steps with 'exact' van der Waals method and gradually decreasing temperature. A harmonic 'distance restraint' has been applied between the amino group of the ligand and the carboxyl of the Asp 107 side chain in the initial 10^6 steps to facilitate formation of the known salt-bridge interaction between these two groups. At least ten independent runs of the docking procedure were performed for each H_1R ligand. The docking results were considered 'consistent' when at least 80% of the individual runs resulted in conformations clustered within a root mean squared deviation of <0.5 Å to the overall best energy pose of the ligand.

47. Newstead, S., Kim, H., von Heijne, G., Iwata, S. & Drew, D. High-throughput fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **104,** 13936–13941 (2007).
48. Cherezov, V., Peddi, A., Muthusubramaniam, L., Zheng, Y. F. & Caffrey, M. A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases. *Acta Crystallogr. D* **60,** 1795–1807 (2004).
49. Aishima, J. *et al.* High-speed crystal detection and characterization using a fast-readout detector. *Acta Crystallogr. D* **66,** 1032–1035 (2010).
50. Winter, G. *xia2*: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* **43,** 186–190 (2010).
51. Leslie, A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 ESF-EACMB Newslet. Protein Crystallogr.* No. 26 (1992).
52. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62,** 72–82 (2006).
53. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40,** 658–674 (2007).
54. Skubák, P., Murshudov, G. N. & Pannu, N. S. Direct incorporation of experimental phase information in model refinement. *Acta Crystallogr. D* **60,** 2196–2201 (2004).
55. 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).
56. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and Development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
57. Abagyan, R., Orry, A., Raush, E. & Totrov, M. ICM Manual v.3.0 (MolSoft LLC, 2011).
58. Totrov, M. & Abagyan, R. Derivation of sensitive discrimination potential for virtual ligand screening. *Proceedings of the third annual international conference on computational molecular biology* 312–317 (1999).
59. Abagyan, R. & Totrov, M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **235,** 983–1002 (1994).