nature

ON THE WING

'Morphing' gives the swift exquisite flight control

The university of the future

The traditional model of the US research university — based on the pre-eminence of the single-discipline department — needs to be stretched and challenged.

The American research university is a remarkable institution, long a source of admiration and wonder. The idyllic, wooded campuses, the diversity and energy of the student populations, and, most of all, the sheer volume of public and private resources available to run them, have made them the envy of the world.

Seen from the inside, however, everything is not quite so rosy. Setting aside the habitual complexity of medical schools, which have separate healthcare and finance issues, the structure of these institutions is straightforward and consistent. The bedrock of each university is a system of discipline-specific departments. The strength of these departments determines the success and prestige of the institution as a whole.

This structure raises a few obvious questions. One is the relevance of the department-based structure to the way scientific research is done. Many argue that in a host of areas — ranging from computational biology and materials science to pharmacology and climate science — much of the most important research is now interdisciplinary in nature. And there is a sense that, notwithstanding years of efforts to adapt to this change by encouraging interdisciplinary collaboration, the department-based structure of the university is essentially at odds with such collaboration.

A second set of issues surrounds the almost static nature of the departmental system. In a country where most things are highly fluid, the fields covered by departments, as well as the pecking order between them, have remained largely unchanged for many years. As people and money have flowed, particularly over the past twenty years, to the south and the southwest, the strongest US universities and departments remain embedded in the northeast and in California. League tables drawn up by the National Academy of Sciences and others show little movement in this pecking order, even over several decades.

Another, perhaps more contentious, issue concerns the relevance of the modern research university to the community it serves. The established model, whatever else its strengths and weaknesses, reflects the desire of the middle classes for undergraduate training that prepares their offspring for a stable career. But how does it serve a society in which people may have to retrain and recreate their careers throughout their adult lives?

These questions are being asked throughout American academia, but nowhere more searchingly than at Arizona State University (ASU), a huge public university that is expanding to meet the needs of the United States' fastest-growing major city (see page 968). Michael Crow, its president, is executing an ambitious plan to replace the traditional model with one in which both influence and research excellence are concentrated not in departments, but in large, broadly based interdisciplinary centres with clear commercial or societal goals.

Whatever its outcome, this experiment will not of itself uproot the traditional university system. Incremental change, notably the establishment of stronger multidisciplinary entities such as Bio-X at Stanford University in California, and several new centres at Harvard, may have a greater bearing on the overall development of the system.

But ASU's effort already tells us plenty about the likely direction of the research university in the up-and-coming regions of America. The university of the future will be inclusive of broad swaths of the population, actively engaged in issues that concern them, relatively open to commercial influence, and fundamentally interdisciplinary in its approach to both teaching and research. ∎

> "There is a sense that the department-based structure of universities is essentially at odds with collaboration."

Taking the first step

China will join efforts to cut carbon emissions, but should not be expected to lead them.

With a new coal-fired plant coming online about once a week, China is on course to overtake the United States as the world's leading carbon dioxide emitter this year or next, an official at the International Energy Agency said last week (see page 954).

This means, of course, that unless China and other major nations such as India and Brazil join a global effort to mitigate the effects of climate change, tough targets and sophisticated carbon markets across the developed world will eventually pale into insignificance.

But the onus is still on the developed nations, who created the problem, to lead that effort.

China is now working on a national plan that will lay out the measures it intends to take to deal with climate change. Despite all the new power stations and the usual difficulty in discerning China's real intentions, there is one good reason to accept that its efforts are sincere: it is in China's self-interest to confront the problem. Assessments by both China and the Intergovernmental Panel on Climate Change have made clear that the nation is likely to be hit hard by climate change, which will increase pressure on water supplies and cut agricultural yields.

In the United States, momentum for domestic action on carbon emissions is growing. As it does so, the issue of how to ensure that China and other developing countries are on board — the issue that prevented the US Senate from ratifying the Kyoto Protocol in the

first place — has resurfaced. Again, those who favour doing nothing are citing possible inaction in developing countries as an excuse for their own refusal to exercise leadership on the issue.

As before, attempts to force China and India into accepting caps on their carbon emissions by refusing to take action without them are doomed to fail. Developing countries have not played a significant role in raising levels of greenhouse gases in the atmosphere in the past, and their per-capita emissions remain far lower than those of the wealthier nations.

Practical considerations, as well as moral ones, preclude the rapid introduction of emissions caps and associated carbon-trading schemes in developing countries. Cap-and-trade schemes are only going to be effective in reasonably advanced economies, where accurate monitoring and reporting of emissions is feasible.

One starting point may be to push developing countries to sign up to more realistic goals than a blanket emissions cap, which as well as addressing the problem of climate change will also deliver associated benefits of high national priority. For example, building-efficiency standards can simultaneously tackle problems of energy security, materials shortages and air pollution in China. Bilateral agreements between rich and poor nations to share clean-energy technologies could help to ensure that these goals are met.

The Chinese government has taken some tentative steps towards managing its energy consumption. It has pledged, for example, to cut its energy intensity — the amount of energy used per unit of economic growth — by 20% between 2005 and 2010 (it isn't, so far, on course to meet this target). It has also put some fairly stringent vehicle-efficiency standards in place. Consistent pressure to step up efforts to meet such commitments could add up to significant reductions in emissions.

Carefully designed incentive mechanisms that encourage developing countries to engage in a global carbon market could also play a role. The Kyoto Protocol established the Clean Development Mechanism (CDM) in 2003. This allows industrialized nations to reduce the costs of complying with Kyoto, essentially by paying individuals in developing countries to adopt lower-polluting technologies than they otherwise would.

Currently, the CDM is rather limited in scope. However, as interest in a global carbon market grows, mechanisms of this type hold out some promise for delivering real global emissions reductions. The major developing nations are increasingly dependent on global economic integration and cannot afford to become outliers in any new global energy strategy that emerges.

In the short term, the best thing that wealthy countries can do to motivate action in the developing world is to take action themselves. Congress should note that a demonstration of good faith will be needed before the United States regains the credibility to engage seriously with countries such as China on this issue. ∎

Conflicted contractors

Government agencies should act to ensure the neutrality of research contractors.

Private contractors play a discreet but increasingly important role in the research operations of government agencies. These companies are generously paid to do everything from surveying endangered species to collecting and reviewing toxicology studies of potentially dangerous chemicals. The information is then used by research agencies or regulators to make critical decisions on public health or the environment.

The relationship can be mutually beneficial: the contractors make money, and the agencies can meet their research needs without assembling a large army of government employees. But many contractors also have private clients — and these are sometimes the same companies that their public-sector clients are trying to regulate.

On page 958 of this issue, we report on an example of the problems this can create. Last year, the US National Institute of Environmental Health Sciences (NIEHS) hired a private firm, Sciences International of Alexandria, Virginia, to help it review a chemical called bisphenol A, which may affect the endocrine system. The contractor, it turns out, also had private clients that produced bisphenol A. Environmentalists cried foul, and the NIEHS has been forced to withdraw the contract. This week, it announced that it would re-review all 20 chemicals that Sciences International had examined.

Under the original contract, a panel of external scientists was due to have the final word on the chemical's safety — and there is no evidence that the contractor was trying to manipulate them. But the potential for a conflict of interest still seems to exist. In this case, although the NIEHS required individual scientists on its advisory panel to sign statements about possible conflicts, it didn't place the same requirement on the contractor.

The NIEHS is hardly alone in this. Individual researchers are often screened quite rigorously for potential conflicts before they participate in government reviews, but contractors are rarely asked to disclose possible conflicts. Many agencies have standard warnings about conflicts in their contracts, but few have comprehensive rules about how companies should segregate their public and private work. There is also little active oversight of declarations that the contractors make, or fail to make.

In the United States, as elsewhere, reliance on contractors has been steadily rising. The time has come for agencies to re-examine how they use contractors, and how they screen for potential conflicts. Agencies should create explicit conflict-of-interest guidelines similar to those already in place for individual scientists. Contractors should be required to divulge potential problems and should be given explicit instructions on how to segregate government and private clients. Agencies should also conduct periodic audits to ensure that contractors are delivering high-quality information.

The vast majority of contractors are honest brokers, and many employ former government officials who are fully aware of what their public-sector clients need and expect. But the system seems to be open to abuse. By taking the right steps now, regulatory agencies can prevent wasteful episodes such as the one at the NIEHS. More importantly, they can ensure that they are receiving the data they need to perform their missions. ∎

RESEARCH HIGHLIGHTS

King of the swingers

Biol. Lett. doi:10.1098/rsbl.2007.0049 (2007)

Sumatran orangutans like to take it easy: new research has shown that the way they move from tree to tree by swaying branches saves them energy.

Susannah Thorpe of the University of Birmingham, UK, and her colleagues studied video footage of orangutans (*Pongo abelii*, pictured) moving through the forest canopy. To reach the branches of a neighbouring tree, the orangutans shifted their weight to set the branch they were clinging to swinging. Thorpe's team calculated that this maneouvre consumes half as much energy as the orangutan would take to jump the gap, and an order of magnitude less energy than climbing down, walking across to the next tree and climbing up again.

A. SHAH

CANCER BIOLOGY
Stop or grow

Nature Cell Biol. doi:10.1038/ncb1567 (2007)

A gene that has confused cancer biologists through seeming to have two opposing roles — causing cells either to proliferate or to permanently stop dividing — works in a dose-dependent way. The finding, which applies to the *Ras* gene, comes from Lewis Chodosh and his colleagues at the University of Pennsylvania in Philadelphia.

The researchers made a transgenic mouse in which they could control the activity level of *Ras* in mammary cells. At high levels of activation, the mammary cells stopped dividing. But at low levels, the cells multiplied into a precancerous state. If these low *Ras* levels persisisted for many days, tumours developed.

The team proposes that *Ras*-induced tumorigenesis occurs when mutations arise that block the pathway that would usually stop the cells' growth.

CHEMISTRY
Quantum brew

J. Am. Chem. Soc. doi:10.1021/ja063766t (2007)

The much-advertised health benefits of green tea rely on an esoteric quantum-mechanical process, say Àngels González-Lafont and her co-workers at the Autonomous University of Barcelona in Spain. They have used quantum-chemical calculations to deduce how catechol-containing compounds, the antioxidant flavonoids in tea, mop up damaging free radicals in the body.

Catechols defuse free 'peroxy' radicals by giving to them a hydrogen atom. The researchers say this switch happens by quantum-mechanical tunnelling — the hydrogen atoms hop through an energy barrier rather than over it, allowing the reaction to proceed quickly. Their calculations suggest that this occurs because the catechol forms a tightly bound complex with the radical, making the energy barrier narrow.

PHYSICS
Pull yourself together

Phys. Rev. Lett. **98**, 156103 (2007)

Wouldn't it be wonderful if flat-pack furniture came with one simple instruction: just add water? That's how the miniature boxes designed by Charlotte Py of the City of Paris Industrial Physics and Chemistry Higher Educational Institution (ESPCI) and her co-workers are assembled.

The starting material is a thin polymer sheet, cut into geometric shapes measuring a few millimetres across. A water droplet placed at each shape's centre pulls the film into a three-dimensional object by capillary forces as the water evaporates. Triangles are transformed into pyramids, crosses into cubes, and six-pointed flower shapes into approximate spheres (pictured right).

ATMOSPHERIC SCIENCE
Winds of change

Geophys. Res. Lett. **34**, L08702 (2007)

Debate over the effect of global warming on hurricanes needs to factor in new data on wind shear.

Wind shear — the difference in wind speed between the top and bottom of the atmospheric layer in which tropical storms form — can tear apart storms before they intensify. Now, Gabriel Vecchi of the National Oceanic and Atmospheric Administration in Princeton, New Jersey, and Brian Soden of the University of Miami, Florida, have analysed data from 18 climate models to assess trends in wind shear up to 2100, assuming that global temperatures increase by a few degrees.

Wind shear in the tropical Atlantic Ocean is projected to increase, inhibiting hurricanes; that in the western Pacific Ocean is expected to decrease.

AM. PHYS. SOC.

SOLID-STATE PHYSICS
Getting warmer

Appl. Phys. Lett. **90**, 142511 (2007)

A theoretical analysis of silicon's ability to superconduct has suggested a way to raise by an order of magnitude the temperature at which this property is seen.

Last year, researchers showed that they could turn cubic silicon — the form of silicon used in electronics — into a superconductor by blasting boron atoms into its structure. However, this property only

appeared when the material's temperature was within 0.3 degrees of absolute zero.

Emmanuel Bourgeois and Xavier Blase of the University of Lyon, France, calculate that silicon doped with aluminium instead of boron atoms could remain superconducting to nearly 3 degrees above absolute zero, because electrons in this system would be more strongly coupled to the material's vibrations.

FERTILITY
Scrambled eggs

Dev. Biol. doi:10.1016/j.ydbio.2007.03.006 (2007)

Recent work suggesting that bone-marrow stem cells can migrate to mouse ovaries and generate new eggs there has been controversial because it challenges the long-standing idea that mammals are born with a limited stock of eggs.

Now researchers led by David Keefe of the University of South Florida in Tampa and his colleague Lin Liu, who also holds a post at Sun Yat-Sen University in Guangzhou, China, say they can find no evidence for generation of new eggs in women. They searched fruitlessly for markers of stem cells or of meiotic cell division in ovarian cells biopsied from 12 women aged between 28 and 53. It remains possible that new eggs might be generated at younger ages.

COSMOLOGY
Wanted: gravity waves

Astrophys. J. **659,** 918–930 (2007)

For gravitational waves, LIGO is the long arm of the law. Its 4-kilometre-long detectors, at two sites in the United States, are Earth's best bet for catching the elusive ripples.

So far, the detectors have yet to find any hard evidence for gravitational waves. But the latest results from LIGO — the Laser Interferometer Gravitational-Wave Observatory — do set a new limit on the strength of the gravitational-wave background. This background is the sum of gravitational waves produced by distant, cataclysmic events, such as stellar explosions and merging black holes, and perhaps even by the Universe's own beginning. The new limit, 13 times lower than before, tightens constraints on cosmological models.

PLANT BIOLOGY
### Border control

Science **316,** 421–425 (2007)

Researchers have discovered a mechanism by which plant roots limit the endodermis — which contains cells that are important for nutrient uptake — to a single layer of cells (pictured below).

Specialization of root cells into endodermis requires a protein called SHORTROOT (SHR). SHR is mobile but travels only one cell layer away from its site of production in the centre of the root. What stops it?

Philip Benfey of Duke University in Durham, North Carolina, and his colleagues show that a protein called SCARECROW (SCR), expressed in the endodermis, tethers SHR in the cells' nuclei and thus prevents SHR movement beyond this layer. Formation of the complex boosts SCR production, further ensuring that SHR does not escape.

The team also found that analogous proteins in rice behave in a similar way.
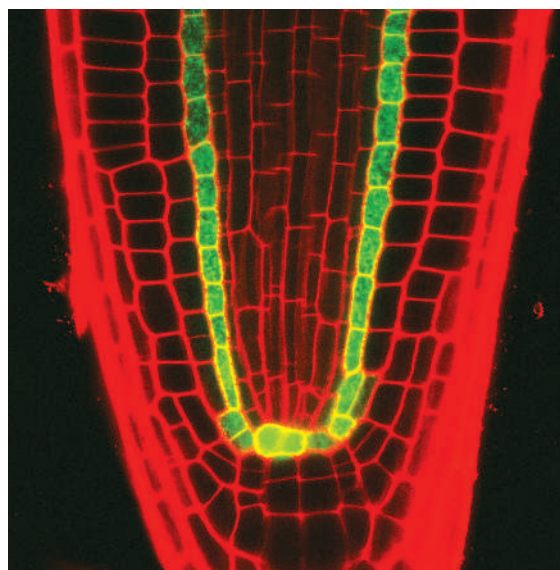
H. CUI

IMMUNOLOGY
HIV plays hide and seek

J. Cell Biol. doi:10.1083/jcb.200609050 (2007)

Researchers in the United Kingdom have uncovered new details of how HIV-1 replicates in immune cells.

In lymphocytes, viruses assemble at the cell's external plasma membrane. But in macrophages, and possibly dendritic cells also, assembly is often seen in compartments within the cell, which previous work had suggested were endosomes.

Mark Marsh and his colleagues of the Medical Research Council Laboratory for Molecular Cell Biology in London, building on recent results that show HIV is produced at the plasma membrane even in macrophages, now report that the intracellular virus assembly sites are plasma-membrane domains connected to the cell surface. The function of these domains is unknown, but the way the virus hides in cells may have implications for vaccine design.

JOURNAL CLUB

David M. Wilkinson
Liverpool John Moores
University, UK

An ecologist enjoys a smelly experiment on a neglected link in the food web.

I have long been fascinated by an idea from the 1970s about rotting food. Daniel Janzen, now at the University of Pennsylvania in Philadelphia, suggested then that many of the noxious chemicals secreted by microbes in decaying food are produced to fend off large animals, allowing the microbes to keep the resource for themselves.

It's an intuitively appealing hypothesis. Our own experience is to be repulsed by putrid food, and several studies have shown that birds prefer fresh over rotted fruit. Most recently, a careful study in the seas off the southeaten United States provided further support for Janzen's idea (D. E. Burkepile *et al*. *Ecology* **87,** 2821–2831; 2006).

In what must have been a gloriously smelly experiment, the researchers baited crab traps with dead fish, either rotten or fresh. The microbe-laden carrion was four times less likely to be consumed by scavengers than the fresh fish.

This provides clear evidence that microbes compete for food with larger animals, something that has been largely overlooked in the huge ecological literature on food webs and feeding relationships. But it doesn't tell us how the chemicals evolved.

Last year, I published with colleagues a theoretical analysis of the evolutionary implications of Janzen's idea (T. N. Sherratt *et al*. *Ecol. Modell*. **192,** 618–626; 2006). Our model suggested that the chemicals cannot have evolved solely to protect against large animals, because the temptation for microbes to 'cheat' by free-riding on toxin production by others undermines the system.

The experiments done by Burkepile *et al*. show that the effect is real, but perhaps these chemicals first evolved for other reasons, such as inter-microbe competition?

NEWS

China struggles to square growth and emissions

China this week postponed its scheduled release of a national plan to deal with climate change. But observers say the move is not so much a delay in dealing with the issue as an indication that the Chinese government is taking climate change more seriously — and that this is as much to protect its own needs as it is a response to international pressures.

China's booming economy — it grew by 11% in the first quarter of 2007 — has long worried those trying to control the rise of greenhouse gases in the atmosphere. And estimates made public last week back up that concern: they suggest that China will exceed the United States as the biggest emitter of greenhouse gases as early as the end of this year — not 2010, as energy experts had previously forecast.

With a new coal power plant being built there roughly every four days, China's emissions are undoubtedly soaring. The question is by how much. A report released by Greenpeace on 24 April says that under a "business as usual" scenario, they will triple by 2050. But compared with the United States, China's per capita emissions are still quite low: in 2003, China emitted 3.2 tonnes of carbon dioxide per person, whereas the United States emitted 20 tonnes per person.

The Chinese government has consistently argued that it cannot put caps on greenhouse-gas emissions while the country is still developing. But the United States has refused to join any international regulatory framework on climate change, such as the Kyoto Protocol, unless limits for developing countries are also included.

China's climate-change plan, which had been set for release on 24 April, has now been put off until next month. But urgent action seems essential. A national assessment released in January paints a grim picture for the country's future. It notes that over the past few years there has been a 12% drop in water supply to northern China. Without significant measures to combat global warming, the report predicts that the productivity of China's three staple crops — rice, wheat and maize (corn) — will fall 5–10% by 2030, and will be close to 40% by the second half of the century.

Such specific predictions are spurring the government to act after many years of inaction,

says Yang Ailun, a Beijing-based environmental expert with Greenpeace. "Before, China always relied on the same arguments about the historical responsibility of developed countries," she says. "That makes sense. But now China is greatly concerned about its vulnerability and social stability."

China is already committed to reining in increases in energy use, and such efforts are closely linked to the country's ability to limit emissions. It plans to close down 50 gigawatts' worth of its least-efficient power plants by 2010. That would mean shutting down roughly 1,000 small plants. "If the small, old plants are there, there is no way to decrease energy consumption," says Yang Fuqiang, a Beijing-based energy expert with the Energy Foundation.

These efficiency initiatives are part of the country's Five Year Plan for National Economic Development for 2006–2010, which calls for a reduction in the amount of energy spent to produce a unit of gross domestic product by 20% within five years. "The targets are tough, but they are possible," says the Energy Foundation's Yang.

Energy-efficiency targets have, however, become a sore point. The first year of this plan fell far short, achieving only a 1.23% drop according to government figures. And so far, signs for 2007 are not good, says James Brock, an independent energy adviser in Beijing.

Brock has advised both the UK and US chambers of commerce on Chinese energy issues. He says he has witnessed the tension in meetings at which officials from the National Development and Reform Commission, which recommended the 20% target, suggested to the powerful State Council that the target can't be met and should be recalculated. "They got read the riot act," says Brock.

Part of the problem is the difficulty in measuring emissions precisely. These are generally estimated on the basis of the types and amounts of energy consumed. But remote-sensing studies of nitrogen dioxide concentrations — which are also released during the burning of fossil fuels — suggest that estimates based on energy consumption could be too low (H. Akimoto *et al. Atmos. Environ.* **40,** 7663–7667; 2006, and A. Richter *et al. Nature* **437,** 129–132; 2005).

Hajime Akimoto of the Frontier Research

> **"Now China is greatly concerned about its vulnerability and social stability."**

Predictions for China's emissions of carbon dioxide could be underestimated by 15–20%.

Center for Global Change in Yokohama, Japan, says the satellite data show a 50% increase in nitrogen dioxide concentrations between 1996 and 2002. This compares with a 15% increase calculated by the International Energy Agency based on energy consumption over that period.

Indeed, experts say the uncertainty is so great that it is difficult to know how China's carbon emissions compare with those of the United States. Gregg Marland, an environmental scientist at Oak Ridge National Laboratory in Tennessee, says that crucial figures needed to estimate China's energy consumption are often missing. Changes in remaining stocks of coal go

FOSSIL FOREST
Extinct species found in coal mine.
www.nature.com/news

H. FALCON-LANG

Primate work faces German veto

German scientists have condemned a parliamentary decision to stop primate experiments at the University of Bremen, calling it unacceptable political interference with the freedom to conduct research.

Animal-rights activists in the state of Bremen have been campaigning for years against the macaque experiments of Andreas Kreiter, a neuroscientist at the university's brain research centre. Now, in the run-up to regional elections on 13 May, the activists have finally won cross-party political support.

Bremen's parliament last month unanimously asked the state government not to reapprove Kreiter's experiments from the end of 2008, when his licence becomes due for renewal. Meanwhile, a commission of scientists and representatives of animal-protection groups is to re-evaluate the scientific value of Kreiter's research, and assess whether the experiments could be replaced by non-invasive methods. Both issues have been addressed before by granting agencies and by the local authorities and ethics committees, Kreiter says.

Neither the state parliament nor the state government can order the university to close down the centre. But Kreiter fears that political pressure will force the local authorities not to approve further experiments.

"This decision was motivated by the election campaign," says Matthias Kleiner, president of the DFG, Germany's main funding agency for university research. "It's an attempt to interfere with the constitutionally guaranteed freedom to research. We hope that the approval authorities in charge will decide solely on the basis of applicable law."

Primate researchers in Germany are used to violent protests. Kreiter's research, on cognitive processes such as attention, involves placing electrodes in monkeys' brains. He and his family were put under police protection in 1997 after they received threatening letters. Attacks have since ceased, but in March animal campaigners submitted to Bremen's parliament 15,000

> **"It seems weird that the same people who have fought me to the finish will now judge my work."**
> — Andreas Kreiter

signatures from people who opposed the experiments.

Considering the ethical pros and cons of animal experiments is always crucial, says Alexander Thiele, a brain researcher at Newcastle University, UK. But he says that the level of political interference with Kreiter's project, which has been approved both legally and ethically, is "scandalous".

Thiele had been offered a professorship at the Humboldt University in Berlin, which he rejected when local authorities last week denied approval for his planned experiments on visual attention in macaques. He says that such blatant political intervention at a local level is less common in Britain because experiments there are

approved by the Home Office.

The state science ministry in Bremen, which oversees the university, is also unhappy with the parliamentary decision. "This is very disturbing," says Holger Bienhold, a ministry official in charge of the natural sciences. "It's hardly conceivable to stop a professor from doing legally and ethically approved research."

Kreiter has won around €3 million (US$4 million) in grant money from various sources, including the DFG, the federal science ministry, and the European Union.

Kreiter is determined to continue his research in Bremen, and has started to prepare for possible lengthy lawsuits. The commission is due to release its review of his work is due in June. "It seems weird to me that the same people who have fought me to the finish will now judge my work," he says. "They will most likely misuse every bit of information they can get."

But the experiments are cruel to animals, says Wolfgang Apel, president of the German Animal Welfare Association and organizer of the petition against Kreiter's work. The committee, he says, will suggest how, and not just whether, the experiments will be phased out.

Alternatives to some types of invasive experiments on primates do exist. For example, functional magnetic resonance imaging can effectively measure activity in large groups of brain cells. But to measure the activity of individual neurons, which is crucial to study cognitive processes such as visual attention, electrodes must be directly inserted into a monkey's brain. ■

Quirin Schiermeier

U. PERREY/DPA/PA

---

unreported, for example, or, for a single year, a figure will inexplicably drop to zero. Marland says that, as a result, any estimates of China's emissions could be off by 15–20%.

He adds that the considerable amount of carbon dioxide produced in making cement, if included in the figures, could also tip the balance. China produces 45% of the world's cement. "China could be exceeding the United States right now," Marland says.

Further details of China's carbon dioxide emissions may have to wait until next year, when NASA plans to launch the Orbiting Carbon Observatory. The satellite is designed to provide high-resolution data on worldwide carbon dioxide emissions. ■

David Cyranoski
See Editorial, page 949

N. BEHRING/PANOS

Elusive flowering signal pruned of mystery at last

Last week, two papers in *Science* reported the discovery of florigen, a long-sought compound with the power to make flowers bloom. But if the celebration of its discovery seems a little muted, it is because many researchers have heard this claim before. And this time, the reports come as an old one is retracted amid charges of data manipulation.

The discovery of florigen was heralded in 2005, when another *Science* paper[1] claimed that it was the RNA produced by a gene called *FLOWERING LOCUS T*, or *FT*. But now the authors of that paper have retracted their findings, and in its stead come two papers that say florigen is not *FT* RNA, but the protein produced by the *FT* gene.

The finding could solve a riddle that has been around since 1865, when German botanist Julius von Sachs observed that illuminating a single leaf on a darkened morning-glory plant was enough to prompt the plant to bloom. That suggested that a signal travelled from the leaf to the site of flower initiation. Some 70 years later, the signal was christened florigen by the Russian plant physiologist Mikhail Chailakhyan.

The hunt was on, and physiologists spent decades testing compounds extracted from flowering plants, only to fail to find the compound responsible. Over time, the florigen concept fell out of fashion, giving way to a hypothesis that the flowering signal was not a dedicated compound but rather a complex mix of nutrient and hormonal signals. "For a long time, florigen was the f-word," says Joe Colasanti, a plant biologist at the University of Guelph in Ontario, Canada. "You didn't want to bring it up."

But the recent work shows that researchers were looking in the wrong places over all those years, says Jan Zeevaart, an emeritus plant biologist at Michigan State University in East Lansing. Most people expected florigen to be a small chemical compound. "They weren't looking for proteins," he says.

Then, in August 2005, two papers reported that although the *FT* gene produces RNA in the leaf, the encoded protein acts in the tip of the shoot, where flowers form. The simplest explanation was that a product of *FT* — either the RNA or the protein — somehow travelled from the leaf to the shoot tip. Within a month, a team led by Ove Nilsson of the Swedish University of Agricultural Sciences in Umeå announced that this product was *FT* RNA. Although that didn't rule out the possibility that the FT protein also travelled from leaf to shoot, it suggested that *FT* RNA was at least a component of florigen. The discovery was significant enough to make *Science*'s list of 'Breakthroughs of the Year'.

But now Nilsson has retracted the paper and has accused Tao Huang, the paper's first author, of manipulating data. Nilsson says Huang selectively excluded some data points and statistically overweighted others. Huang, who left Nilsson's lab for a faculty position at Xiamen University in China after the paper was published, maintains that excluding the data was justifiable. He says he circulated the data — with excluded data points marked — to

Brain's speech site is revisited and revised

Analysis of two damaged brains, preserved in a museum since the nineteenth century, could force neuroscientists to rethink the area where language resides in the brain.

In 1861, the French surgeon and anatomist Paul Broca described two patients who had lost the ability to speak. One patient, Lelong, could produce only five words, and the second, Leborgne, could utter only one sound — *"tan"*. After their deaths, Broca examined their brains and noticed that both had damage to a region in the frontal area on the left side. Broca's area, as it became known, is now thought to be the brain's speech-processing centre.

Broca kept the patients' brains for posterity, preserving them in alcohol and placing them in a Paris museum. And that's where Nina Dronkers, of the VA Northern California Health Care System in Martinez, and her colleagues picked them up, in order to reinspect the damage using magnetic resonance imaging.

Leborgne's brain had been scanned twice before, but not Lelong's. And neither had been compared with modern interpretations of Broca's area. After the team put the two brains through a scanner, they came up with a surprising finding: in both patients, the damaged

Paul Broca: discovered a region of the brain responsible for language.

area was much larger than the region that is now considered to be Broca's area.

"We were noticing that what people were calling Broca's area encompassed large areas of the frontal lobe," says Dronkers. The scans show that neither of the old brains had damage that affected the whole region now known as Broca's area. But damage also stretched far into other regions beyond this spot.

Broca realized this at the time, says Dronkers, and noted that the areas of damage were different in the two patients. But his conception of the area involved in speech processing has become simplified by others over time, the authors argue. They published their findings online earlier this

Root cause: the signal for flowering in the thale cress *Arabidopsis* has been found to be a protein.

his lab colleagues before publication, but no one objected to the exclusion. Huang has not agreed to the retraction, calling it premature.

Nilsson's retraction was published at the same time that *Science* released the two new papers[2,3]. Both papers — one on the thale cress *Arabidopsis* and the other on rice — report that although *FT* gene expression is restricted to the leaves, the protein can travel to the tip of the shoot. And both papers fail to find evidence for movement of *FT* RNA.

The timing of the papers, coupled with what several researchers have described as an unusually short, 40-day review, has led some to speculate that the papers were pushed to publication more quickly to coincide with the retraction. But Katrina Kelner, *Science*'s deputy editor for life sciences, says the review period was not abbreviated. "It was sensible to have them come out at the same time for maximum clarity of the literature," she says. "We coordinated them, but the review process of those two papers was in no way abnormal."

The new work comes with its own share of caveats. Both groups rely on commonly used but indirect measures of protein movement, and some researchers have pointed out that key controls are lacking.

Overall, however, many experts say the new papers are convincing. "None of these is really the killer experiment," says Detlef Weigel of the Max Planck Institute for Developmental Biology in Tübingen, Germany. "But I would say the overwhelming evidence is that the protein moves." Zeevaart goes even further. "The problem is solved," he says.

But with a history as chequered as florigen's, not everyone is ready to close the book. "It's always good to be cautious," says Colasanti, "especially in this field." ■
Heidi Ledford

1. Huang, T. *et al. Science* **309,** 1694–1696 (2005).
2. Tamaki, S. *et al. Science* doi:10.1126/science.114753 (2007).
3. Corbesier, L. *et al. Science* doi:10.1126/science.114752 (2007).

month in the journal *Brain* (N. F. Dronkers, O. Plaisant, M. T. Iba-Zizen and E. A. Cabanis *Brain* doi:10.1093/brain/awm042; 2007).

This misplaced focus could lead to problems when diagnosing people with language impairments, says Dronkers. By assuming that only one small area of the brain is responsible for language, clinicians might overlook other regions involved in speech production. In other words, focusing too heavily on Broca's area could be missing the point, Dronkers argues.

Others agree. "There's a tendency for researchers to see activation in somewhere like Broca's area and to say 'oh well, we're tapping into a language area'," says Joseph Devlin, a neuroscientist at the University of Oxford, UK, who images language networks in the brain.

Newer imaging techniques may also help researchers to discover what Broca was unable to see. Dronkers and Devlin are both working on the use of alternative imaging techniques to investigate other regions of the brain that may be important in language processing but which are not detected by magnetic resonance imaging, such as the tracts of white matter that connect areas of grey matter. ■
Kerri Smith

The brain of Lelong, one of Broca's patients, about to be scanned.

SIDELINES

ZOO NEWS

Puppy love
Researchers at Seoul National University in South Korea will this year mate Snuppy, the world's first cloned dog (right), with Bona, the world's second (and first female) clone, to check their reproductive abilities.

REUTERS/SEOUL NATL UNIV.

NUMBER CRUNCH

US$421,200 was the amount paid at auction last week for the skeleton of a mammoth nicknamed 'The President' — a record for such an artefact.

11 other items in the same sale of palaeontological curiosities, which was held at Christie's in Paris, France, were also sold for world-record prices.

US$1.53 million is the total amount of cash splashed out at the auction, mostly by private collectors.

ON THE RECORD

"This proves it's possible for humans to change the weather on the world's highest plateau."
Yu Zhongshui, an official at China's Tibet meteorological station, on the successful effort to create snowfall over the city of Nagqu by seeding clouds with silver iodide particles.

OVERHYPED

Kryptonite
The name of Superman's nemesis has been given to the newly discovered mineral sodium lithium boron silicate hydroxide, because it happens to have a very similar name to the formulation for kryptonite quoted in the film *Superman Returns*. The real-life version, however, is not green, does not come from outer space and can't kill superheroes.

Sources: AFP, Associated Press, Daily Telegraph, Natural History Museum

Regulators pull contract for chemical review

When US government regulators hired the firm Sciences International to help review a potentially dangerous chemical, they almost certainly didn't intend to ignite a political firestorm.

But the company's links to the chemical industry have enraged environmentalists and forced the National Institute of Environmental Health Sciences (NIEHS) to pull the contract. Last week, in the face of growing pressure from both Congress and the public, the agency announced that it would re-review the 20 chemicals that the company had been involved in studying.

Scientists, environmental groups and even private contractors say that the situation highlights a weakness in federal environmental and health regulation. Private companies hired to do risk assessments are rarely required to disclose conflicts of interest. And often, the companies have relationships with the industries that the regulators are charged with overseeing. "It is definitely a problem," says Sandra Schubert, director of government affairs for the Environmental Working Group, the watchdog based in Washington DC that criticized the work.

The review involves bisphenol A, a chemical commonly found in plastics, including baby bottles and microwave containers. The chemical disrupts the endocrine system in mice, and some researchers think that it might cause similar effects in humans. Both Japanese and European regulators say that more research is needed to determine whether the chemical is harmful.

In July 2005, the NIEHS hired Sciences International of Alexandria, Virginia, as part of a panel to summarize hundreds of studies on bisphenol A. The firm had done the same sort of work many times before for the agency. But this time, the Environmental Working Group caught wind of the study and learned that Sciences International had two clients — Dow Chemical and BASF — that produce bisphenol A. In late February, the advocacy group wrote to the agency demanding a review of Sciences International's role in the study. After several weeks of mounting pressure, the NIEHS announced that it was scrapping its contract and reviewing the work already done on bisphenol A and the other chemicals.

Several scientists involved in the review told *Nature* they think Sciences International had behaved ethically and impartially. "I really don't

> **"The panel should be shut down."**

think that they were trying to influence the process," says Simon Hayward, a cancer biologist at Vanderbilt University in Nashville, Tennessee.

But others contend that the company was deliberately stacking the review literature in favour of industry. "All those reports are suspect," says Frederick vom Saal, an endocrinologist at the University of Missouri at Columbia, who was not involved in the review. "The panel should be shut down."

Independent scientists on the bisphenol-A panel were required to sign conflict-of-interest forms, but Sciences International was not. The

Online resources threaten livelihood of libraries

The closure of five of the 26 regional libraries of the US Environmental Protection Agency (EPA) last year sparked international protest. Congressional hearings were held, and a government investigation was launched. In February this year, the president of the American Library Association told Congress that the closures have restricted access to information in at least 31 states.

The EPA is not alone — last year, the Department of Energy closed its headquarters library in Washington DC. And now, NASA is considering downsizing its network of libraries — including the one at its leading science-research centre, the Goddard Space Flight Center in Greenbelt, Maryland.

Tight budgets could bring even more whittling of libraries in the United States. "When budgets are threatened, agencies tend to say let's put the library on the chopping block," says Tara Olivero, assistant director at the American Library Association in Washington DC.

Libraries are already working to reinvent themselves in a digital world in which online access is fast reducing the need for rows of books and stacks of journals. But a full transition to electronic resources might not save money as agencies hope.

Officials at the EPA, whose libraries provide a wide range of information about environmental protection and management,

initially cited a proposed budget cut of US$2 million — an 80% drop from the previous year — as the main reason for downsizing its library network. Yet critics point out that internal EPA studies have suggested that having a librarian saves between $3 and $7

in professional staff time for every dollar invested. Since the closures, EPA librarians have struggled to maintain the same level of cost effectiveness. On requesting a publication from another library, they are sometimes told that the item is not available and that no one

Replacing books with electronic resources may not save money.

CORBIS

SCRATCHING DIAMOND JUST GOT EASIER
Ultra-hard material made in the lab without high pressures.
www.nature.com/news

R. B. KANER & J. B. LEVINE/UCLA

A review on the safety of bisphenol A, found in baby bottles, is under scrutiny for conflicts of interest.

CORBIS

Martin, a firm that has been sued for leaking perchlorates. Syracuse has an internal panel to review potential conflicts of interest, says spokeswoman Lisa Mondello. In this case, she says, the company collaborated with a division of Lockheed that doesn't handle perchlorates. "It's really not connected," she says.

"The larger the company, the more it becomes an issue," says James Lamb, a toxicologist with the Weinberg Group, a consultancy firm based in Washington DC that does regulatory work for private industry. Lamb says that some companies, like Sciences International, try to segregate industrial and government work to limit conflicts, but few if any regulators have rules about disclosing or managing such issues. "Conflicts of interest have never really been completely ignored," he says. "But I really don't know what the legal process has been."

In fact, the relationships are a legal grey area that has troubled regulators in the past. In 2004, the US Environmental Protection Agency and the NIEHS came under fire for accepting US$2 million from the American Chemistry Council — an industry lobby group — to conduct a children's health study (see *Nature* **432,** 6; 2004). The agencies eventually withdrew from the arrangement, and a subsequent investigation by the US Government Accountability Office found that stronger guidelines were needed. Both agencies say that they have since rewritten their guidelines about directly taking funds from industry. ∎

Geoff Brumfiel
See Editorial, page 950.

company did volunteer to report any potential conflicts of interest, according to Sciences International's president Herman Gibb. But he says that there was no conflict because the staff working on bisphenol A were not working for private clients. "We didn't do anything wrong," he says. "This is unfair with a capital 'U'."

Sciences International's website says that about half of its business comes from private companies and half from the government, and it is hardly alone in that regard. A brief investigation by *Nature* turned up at least half a dozen other firms that work simultaneously for federal regulators and industry — sometimes in the same area.

For instance, in 2005, Syracuse Research Corporation, a non-profit company based in New York, was hired by the US Centers for Disease Control and Prevention to help review perchlorates, which are reactive salts that can damage the thyroid. At the same time, the company's defence division was partnered with Lockheed

---

knows where it is, says Bernadine Abbott Hoduski, a former EPA librarian.

The agency is working to convert all EPA documents into an electronic format, and its spokeswoman Jessica Emond says that the project "has not incurred additional cost". But critics argue that for agencies considering downsizing their physical libraries, going electronic will almost certainly require more, not less, money.

For a start, not everything can be digitized, and dedicated staff are needed to assess what material should be put in digital form. Content must also be produced using technologies that will be usable decades in the future. "In many ways you need a higher level of expertise to produce and maintain an electronic service than a physical one," says Anne Kinney, a

NASA scientist and chair of a group that recently assessed how the print materials at the Goddard library could be reduced.

The demand for library services other than shelf space has shown no sign of tapering off and, if anything, has increased in recent years. To ensure that electronic resources don't result in costs simply being transferred to individual researchers, those services need sustained funding, librarians say.

At the Naval Research Laboratory library in Washington DC, chief librarian James King says that he has seen a dramatic falling-off in the number of people walking through the door. But "space is not synonymous with service", he says. Use of the naval library's online databases has doubled in the past five years.

"There is a risk of people seeing libraries only as warehouses," King says. He and others argue that in a time of information overload, librarians have an ever-more-valuable role in designing web interfaces that facilitate browsing

"Space is not synonymous with service."

and focused searches. They also create and operate databases specific to the needs of the research communities they serve, and have intricate knowledge of electronic resources across agency- and university-library networks. And in a world where new journals are continually coming online, librarians say they are best placed to negotiate with publishers to obtain the cheapest site-licence

contracts and to monitor the changing needs of a specific community of users.

For her part, Kinney envisions a different future for the science library. Instead of silent halls with towering racks of books, smaller meeting places could double up as information centres, where researchers can plug in their laptops, hold discussions, and talk to librarians about how to navigate the myriad resources online.

NASA has not yet made any decisions about closing any libraries, but the issue is likely to remain on the table: "The NASA budget isn't getting any bigger," notes Robin Dixon, chief librarian of Goddard. The challenge will be to cut back on physical resources without cutting back on service. ∎

Lucy Odling-Smee

Rival genetics projects build bridges

Projects that explore the world's biodiversity might seem like natural allies. But last week a workshop was held in North Carolina expressly to mend fences between two such initiatives that have different approaches.

Both projects use genetics to catalogue the relationships between organisms. The Assembling the Tree of Life project, funded by the US National Science Foundation, explores the evolutionary relationships between different classes of organisms. And the Consortium for the Barcode of Life initiative, funded largely by private foundations, uses short DNA signatures to discriminate between species.

Some systematists, such as those in the Tree of Life project, think that the upstart barcoding approach ignores crucial information that helps to define organisms, such as their habitats and their relationships with other species. As a result, the two groups tend to regard each other with more scepticism than friendliness.

The workshop, which was held on 19–20 April and attended by about 40 scientists from both communities, was intended "to build bridges", says co-organizer Joel Cracraft, an ornithologist at the American Museum of Natural History in New York.

For some, the occasion was a milestone. "This really is a

> **"The joint workshop is a signal that a lot of the hostility to genetic barcoding has died down."**

signal that a lot of the hostility about barcoding has died down and disappeared," says David Schindel, executive secretary of the Barcode of Life project.

But others remain more reserved. "I still have some scepticism about barcoding, depending on what its goals are," says Cracraft.

Proponents of barcoding have encountered particularly tough resistance to the idea that the technology will help the world to catalogue its fast-disappearing biodiversity, by quickly identifying new species based on DNA differences from known species. Systematists scoff at this idea and say that because species have varying levels of genetic diversity, setting a uniform cut-off for how much diversity distinguishes a new species is tough.

But barcoders have now published research showing that their approach can work in some contexts (K. C. R. Kerr *et al. Molecular Ecology Notes* doi:10.1111/j.1471-8286.2007.01670.x; 2007). And systematists have decided to try to find common focus with barcoders, rather than turning up their nose at the entire enterprise.

The meeting's organizers hope that these connections will ultimately lead to collaborations. But for now, they say, just getting the scientists to talk to each other is a good first step. ■

Erika Check

Crop biodiversity project receives seed money

Funding has been secured for a major biodiversity project that could help safeguard seeds and genetic data for up to 165,000 varieties of 21 food crops.

The project, carried out by the Global Crop Diversity Trust and the United Nations Foundation, will replenish stores of seeds in poorly maintained seed banks throughout the world and ensure that back-up stocks are established. It will also create a database that gives plant breeders access to the genetic blueprints underlying the different strains.

Designed to ensure that plant breeders retain as much flexibility as possible in creating improved crop strains, the project will be paid for by a $30-million grant from the Bill & Melinda Gates Foundation and $7.5 million from the government of Norway.

India puts Italian AGILE satellite into orbit

Italy's satellite AGILE was launched by India's PSLV rocket on 23 April, in India's first commercial launch solely for a foreign spacecraft.

AGILE, which had been delayed for two years as Italy's space leaders debated the launch, will provide the first measurements of high-energy gamma rays since NASA's Compton Gamma-Ray Observatory fell back to Earth in 2000. A NASA programme called GLAST will make similar measurements following its launch later this year.

The PSLV is due to launch India's first spacecraft mission to the Moon, Chandrayaan-1, in 2008.

M. LAKSHMAN/AP

AGILE manoeuvre: India's PSLV rocket takes off.

Cassini offers solution to the riddle of the rings

Fresh light has been shed on why William Herschel was able to sketch the rings around Uranus in the late eighteenth century (inset). Other astronomers were unable to confirm his sighting with more powerful telescopes until the Keck telescope on Hawaii confirmed the rings' existence in 1977 (right).

Stuart Eves of Surrey Satellite Technology, a UK satellite company, told the Royal Astronomical Society's National Astronomy Meeting in Preston that the Cassini mission has shown that Saturn's rings are darkening and expanding. If the same phenomenon occurred around Uranus, he says, that would explain why its rings have become progressively harder to see. Herschel's sketch shows the rings in the right orientation, and his description that they are reddish has also been confirmed.
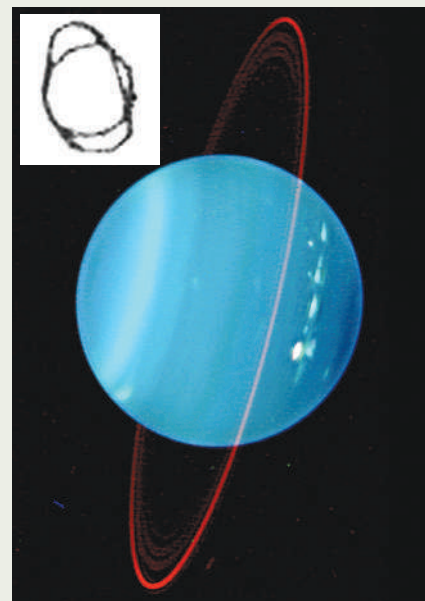
L. SROMOVSKY/UNIV. WISCONSIN-MADISON/W. M. KECK OBSERV.

Contaminated-blood inquiry begins in Britain

An independent public inquiry has started in Manchester, UK, into how thousands of haemophiliacs were infected with HIV and hepatitis C in the 1970s and 1980s by being given contaminated blood products. It follows related scandals in France, Japan, Germany and the United States, where similar proportions of haemophiliacs were infected.

The inquiry, led by Labour peer Peter Archer, will examine the timetable of decisions to introduce heat-inactivated clotting factors and products from the United States that were known to be high-risk because of the US practice of paying donors. The focus of HIV infection will be the narrow window between the discovery of the virus as the cause of AIDS in 1983 and the introduction of safe blood products in 1985. Haemophiliacs have long demanded an inquiry, which could open the way for compensation.

Germany relaxes rules on postdoc contracts

A controversial rule disallowing scientists in Germany from working on short-term contracts for longer than 12 years — or 15 years in the case of clinical research — has been retracted.

The rule was introduced in 2001 as part of a package of measures intended to improve the lot of young German researchers by limiting the duration of insecure postdoc work (see *Nature* 415, 257–258; 2002). But it was criticized from the beginning for introducing unnecessary restrictions — and for discriminating against women, who are likely to take time off to start a family during their 20s and 30s.

A replacement law that allows grant money to be used to pay postdoctoral researchers without time restriction came into force last week.

Canada launches open-access medical journal

An open-access medical journal has been set up by editors who left Canada's best-known medical publication, the *Canadian Medical Association Journal* (*CMAJ*), last year in a disagreement over its editorial independence.

The journal, *Open Medicine*, will be published by a non-profit organization and will not accept advertising from makers of drugs or medical devices. It plans to sustain itself mostly by voluntary contributions for the first year and then start charging authors to publish.

"We are launching *Open Medicine* as an open-access journal because we believe the current, closed-access model is inconsistent with the values of medicine," says co-editor Anita Palepu. Palepu, seven colleagues and 16 board members left the *CMAJ* after two of its editors were dismissed for publishing a politically sensitive news story in the journal.

Correction
In our News Feature "Is French science in decline..." (*Nature* **446,** 854; 2007), the axis on the graph should have read: "Gross domestic expenditure on R&D as a percentage of GDP". And the sentence "The challenge for French research in which politicians can make the biggest difference is the decline in government spending" should have referred to 'science' not 'government' spending.

BUSINESS

The Stagecoach fleet now uses the Envirox fuel additive in all its vehicles.

D. STARES/ALAMY

Nanotech growing pains

A licensing dispute over fuel additives could spell trouble for one of Britain's nanotechnology stars, as **Katharine Sanderson** reports.

A few months back, the outlook was bright for Oxonica. Last August, the nanotechnology firm signed a £6-million (US$12-million) deal to sell its nanoparticle-based fuel additive to the Turkish oil firm Petrol Ofisi. But now, that deal is in trouble — and Oxonica is locked in a court battle over the intellectual property behind the additive, called Envirox.

On 23 February, the company asked the High Court in London to rule that it doesn't need to keep paying royalties for the product to Neuftec, a company registered in the Caribbean island of Dominica that first licensed the fuel-additive technology to Oxonica in 2001.

Oxonica says that it now uses a different technology to make its additive, and that its product therefore no longer falls under Neuftec's patent. The case is characteristic, analysts say, of the intellectual-property arguments that are likely to sear through the nanotechnology industry as more ideas make the transition from research to commercial products.

"When it comes to intellectual property, companies will need to get together," says Michael Holman of Lux Research, a New York consultancy firm that specializes in nanotechnology. "A lot of areas have very dense and overlapping patent spaces." If such legal battles are to be avoided, Holman says, companies that hold patents on ideas will have to cooperate closely and carefully with those, such as Oxonica, that specialize in their practical application.

Envirox was originally based on an idea developed by two businessmen, Ronen Hazarika and Bryan Morgan, who were looking for ways to improve fuel efficiency and were conducting their own, crude trials in their families' cars. In 2000, they patented an additive consisting of nanoparticles of cerium oxide, an oxidation catalyst, that had its surface treated to help it disperse in diesel. Morgan and Hazarika then set up Neuftec in Dominica, and gave it control of the patent. They also approached Oxonica with their idea. Oxonica licensed the patent from Neuftec, hired Hazarika in July 2001, and developed the product as Envirox. A third company, Australian-based Advanced Nanotechnology, supplied Oxonica with the cerium oxide particles used in the additive.

Oxonica, which was started at the University of Oxford, UK, in 1999, says that Envirox improved fuel efficiency by up to 11% in initial trials. The product soon became popular. In 2004, Stagecoach, one of Britain's largest bus operators, said that it would use it in all of its buses. Last week, the bus firm confirmed that it was still doing so.

But Hazarika left Oxonica in September 2005 for undisclosed reasons. A year later, Oxonica signed a deal with Petrol Ofisi to use Envirox in its diesel fuel. But by then, Oxonica had introduced a second additive, also based on cerium oxide, from a different supplier. Oxonica claims that this alternative additive does not fall under Neuftec's patent — but Neuftec disagrees.

Hazarika, who now works for a Singapore company called Energenics, says that Oxonica refused to divulge the composition of its new additive. "We were constrained by confidentiality from our supplier — we made that clear to Neuftec," says Kevin Matthews, chief executive of the Oxford firm. Neuftec gave Oxonica an ultimatum: reveal the details of the product by 23 February, or face legal action. "We wanted to understand why they did it — what led to the second product being developed and what the composition was," says Hazarika.

The day the deadline passed, Oxonica did not give Neuftec the details, but instead took the matter to court. "We filed to protect ourselves," says Matthews. "We asked the court to take a judgment as to whether the product was licensed or not." On 11 April, Neuftec filed a defence and counter-claim against Oxonica at the High Court.

Hazarika also claims that Oxonica is selling its new additive on the basis of test results from the first product — something that Matthews denies. "Of course we have got new test results," he says. The difference between the products concerns the surface treatment used to stabilize the cerium oxide nanoparticles in the fuel, Matthews says. He adds that Oxonica would like to sell the two in combination — although it won't use Neuftec's formulation while the dispute plays out in court.

Trial troubles

Meanwhile, Oxonica's Turkish deal seems to be in trouble. On 28 March, Oxonica issued a statement saying that although initial field tests with Petrol Ofisi in Turkey had shown a fuel saving of 1–1.5%, a second trial was "inconclusive and it has not been possible to identify whether there has been any improvement in fuel economy". These results were a "hit in the shins" for Oxonica, says David Gittins, a chemical engineer who assesses new technologies for Imerys Performance Minerals near St Austell, UK. The company's shares dropped after the announcement (see chart).

In an extra twist, Advanced Nanotechnology is now working with Hazarika's new company, Energenics, to try to sell the fuel additive in North America. Even if it loses its battle with Oxonica, Neuftec will still have its patents — although Hazarika is unsure whether the same additive will be used again. "The product is tainted by what is going on in Turkey," he says.

The case isn't the first of its kind and certainly won't be the last, as nanotechnology products and processes flood the market. "Oxonica need to be able to get these intellectual-property issues sorted out to get the value they need from the technology," says Holman. "It's a lesson that companies need to pay attention to." ∎

OXONICA STOCK

Price (British pence)

170

130

90

Nov '06 — Apr '07

PIMP MY ANTIBODY

Antibody therapies have had more than their fair share of crashes. But designers are at work on faster, fancier new models, finds **Erika Check**.

DAVID SIMONDS
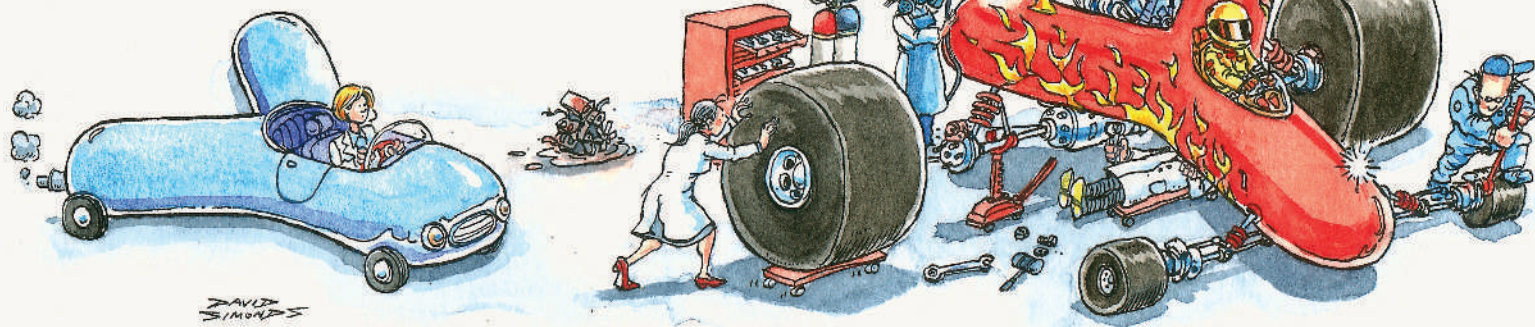
You know the scene. The mechanics are in the body shop tinkering with their toys, revving things up. They're reworking big ends, stripping machines down to the basics, welding the back half of one to the front half of another in search of something new. They're after something unexpected, something flashy, something no-one else has but everyone will want. It has to be streamlined; it has to be powerful; it has to be fast. And because these mechanics are customizing the design of medically important antibodies — not just pimping up their dragsters and low-riders — it also has to be very reliable and extremely safe.

The first therapeutic antibody, a mouse-derived molecule introduced in 1986, succeeded in blocking the normal immune attack on transplanted organs. But this was a jalopy by today's standards, and such molecules have already been through two major redesigns. In the first, scientists developed 'humanized' antibodies: safe, efficient molecules that more closely mimic antibodies made by the human body. In the second, companies worked out how to mass produce them, making antibodies serious contenders in the drug market.

Now antibodies are being souped up for a third time, and this time the result is set to be more radical. During the next decade, researchers and drug companies hope to introduce a new convoy of antibody drugs that is safer to patients and more lethal to pathogens or destructive cells. Many members of this new breed are fragments of antibodies that can reach targets that whole antibodies cannot. Some are proving useful for treating diseases once thought to be way beyond antibodies' range.

The latest spruce-up is partly motivated by rapid recent progress in biology. Scientists now understand more about how antibodies interact with their target molecules and how they recruit the body's own defences to help eliminate them. This knowledge can be converted into more effective therapies.
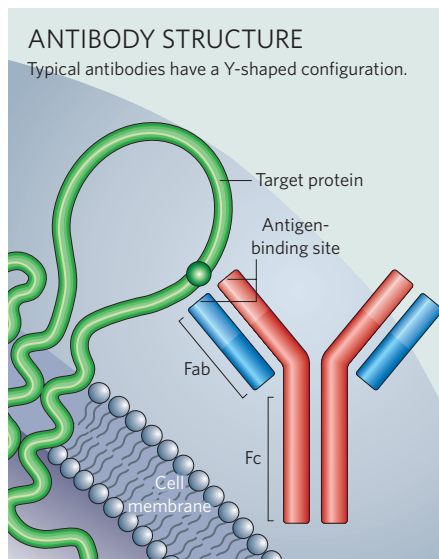
The redesign is also inspired by some very public disasters that have occurred during the past few years. In 2005, Biogen Idec, based in Cambridge, Massachusetts, and Elan, in Dublin, Ireland, were forced to temporarily remove Tysabri — an antibody intended to treat multiple sclerosis — from the US market after it was linked to a rare infection that can be fatal. And last year, six volunteers ended up in intensive care after taking an experimental antibody drug made by the now-insolvent German company TeGenero[1].

These failures have prompted caution in the design, trial and sale of antibody drugs. Biogen Idec and Elan relaunched Tysabri last July under a tightly controlled programme so that it can track and respond to any suspected side effects. And the scientists behind TeGenero's approach are now looking to treat patients' cells outside the body, which they hope will sidestep the problems caused when their antibody was delivered straight into the bloodstream. "There is definitely a consensus that if you are giving antibodies to patients you should be jolly cautious," says Richard Begent, a doctor and researcher at Royal Free and University College Medical School in London, UK, who helped investigate what went wrong in the TeGenero trial.

Revving up research

In the early 1990s, drug companies were dubious about antibodies because some of the earliest candidates didn't work, and because they are large molecules that are expensive and difficult to make. It took the blockbuster success of drugs such as rituximab — the first antibody approved to treat cancer in 1997, developed by Genentech and IDEC Pharmaceuticals — to coax major drug companies to take a second look.

Antibody therapies have proved successful in part because they work against some diseases for which treatment options are scarce, such as cancer. They also make money. "There's nothing like success in the marketplace to get pharma to look at something really hard," says Simon Moroney, co-founder and chief executive of the fifteen-year-old protein-design company MorphoSys, of Martinsried in Munich, Germany. Today, 21 antibodies are approved for sale, and the market is projected to be worth US$22 billion by the end of this year. Genentech, based

ANTIBODY STRUCTURE
Typical antibodies have a Y-shaped configuration.

Target protein

Antigen-binding site

Fab

Fc

Cell membrane

in South San Francisco, California, has grown from a small, cash-strapped start-up to a multibillion-dollar biotechnology giant on the strength of antibodies. Most major pharmaceutical firms now have partnerships or in-house discovery units devoted to antibody therapies, and much of the basic research is being done by them and in biotech companies.

It may seem audacious for scientists to try to improve on our own antibodies, which are elegantly constructed to recognize and eliminate harmful cells and organisms from the body. Natural circulating antibodies have a structure that resembles a Y (see 'Antibody structure'). The portion of the antibody corresponding to the arms of the Y is the Fab region. These arms feature 'hands' and, just like the myriad shapes and sizes of human hands, the Fab region's hands come in an endless variety of forms that each recognize and dock with one of billions of target molecules, such as structures on invading pathogens.

Once docked, the antibody's business end — the Fc region — goes to work. This portion corresponds to the tail of the Y. It sends signals to the body's own assassins — cells and molecules that destroy whatever the antibody has latched on to.

Bits and pieces

The Fab and Fc regions were both thought to be essential to antibodies' ability to seek out potential threats within the body and mark them for annihilation. So many scientists scoffed when, in 2000, a 28-year-old researcher named Ian Tomlinson and his colleague Gregory Winter, based at the prestigious Medical Research Council's Laboratory of Molecular Biology in Cambridge, UK, founded a company on what seemed like a hopeless premise. (Tomlinson even left the institute to focus on this project.) Winter's team had shown that it could create 'domain antibodies' consisting of only a small stump of the Fab region, and completely lacking an Fc region[2]. Tomlinson and Winter claimed that these 'abbreviated' antibodies would be able to follow paths that are off-limits to whole antibodies, such as passing through the gut into the bloodstream. Whole antibodies are too large to penetrate the gut lining and must be injected directly into the blood.

Seven years later, this play looks quite smart. It turns out that the Fc region of an antibody, which marks cells or molecules for destruction, isn't always necessary. In some cases, binding of the Fab region prevents a molecule

The disastrous trial of TeGenero's experimental antibody at Northwick Park holds important lessons.

S. CLARKE/REX FEATURES

from performing a detrimental action, making its annihilation unnecessary. (Indeed, the Fc region can even be dangerous; reactions mediated by the Fc portion of the TeGenero antibody are thought to have had a role in its extreme side effects.)

Domantis, the company founded by Tomlinson and Winters and based in Cambridge, says it has shown that its domain antibody against a molecule known as tumour necrosis factor alpha is effective in animal models of rheumatoid arthritis and chronic lung disease, and preclinical trials are under way. Last December, Domantis was bought for £230 million (US$461 million) by drug giant GlaxoSmithKline.

The acquisition of Domantis is a prominent endorsement of a burgeoning area of innovation: using pieces, parts or combinations of antibodies. Companies and academic researchers are now testing all sorts of pared-down or pieced-together antibodies in the hope that they will reach targets inaccessible to the whole molecule, or hit already accessible targets more effectively.

The varieties are seemingly endless. Some consist of the entire Fab region; six such therapies have been approved by the US Food and Drug Administration (to prevent blood clotting, for example, or treat snake bites) and more are in clinical trials. Others are just a tiny portion of the Fab region, consisting of little more than a targeting domain — that is, a single hand one of the antibody's arms. Some such fragments have been linked together in doublets or triplets that bind to multiple targets

> **"Most major pharmaceutical firms now have partnerships or in-house discovery units devoted to antibody therapies."**

or to different regions on the same target, thus binding it more tightly. Domantis is working on such linked antibodies. Even more fragments have been connected to other molecules such as stabilizing agents or drugs.

Scientists in Maryland and California, for example, have glued together small segments of a Fab region that attaches to a receptor on human cancer cells. They linked it to the chemotherapy agent doxorubicin in order to target the drug exclusively to cancer cells[3]. The question now is how effective this and the many other innovative structures will be. Most scientists suspect that this will depend on the disease being treated and the specific structure of the molecule designed to treat it.

Back-seat driver

Many researchers are focusing on the front portion of the antibody, but there is also increasing interest in the Fc region. The importance of this region was famously shown five years ago when a group of researchers based in France published a study on rituximab. This antibody binds to and wipes out B cells, which multiply in certain types of cancer. Rituximab has been a miracle cure for some patients, but fails to work in 30–50% of people with particular lymphomas.

The researchers divided patients on rituximab into groups on the basis of genetic variations in particular receptors that bind the Fc region and help to kill whatever the antibody has bound to. One group had receptors that bind the antibody tightly, and another had a form that bind it more loosely. Of those patients with the tighter-binding form, 90% benefited from rituximab, compared with only 50% of those with the looser-binding variant[4]. This

was a dramatic demonstration that interactions involving an antibody's back end can be crucial to its effectiveness — "a pretty big step," says Greg Lazar of the biotechnology company Xencor, of Monrovia, California.

Researchers and companies are now attempting to modify the rear end of antibodies so that they bind their receptors more strongly and boost a patient's immune response. Xencor, for instance, has retooled its entire business model to focus on this idea. The field is rife with competition, and companies are doing everything from changing the structure and surface properties of the Fc region to altering its amino-acid sequence. It's not clear whether these attempts will work. But it won't take long to find out, because clinical trials of antibodies with modified Fc domains have already begun.

Until now, many such therapies have been aimed at cancer cells, which have their own unique protein signature an antibody can latch onto. Autoimmune disease has been another focus, because antibodies can interfere with specific overactive proteins in the immune system.

New directions

But now companies are broadening their sights. One of the most intriguing examples of this began about ten years ago, when scientists at Genentech were trying to treat a condition known as peripheral neuropathy by injecting patients with a form of nerve growth factor, a protein that was lacking in the neurons of their skin. The company stopped developing the drug because it wasn't working well and because of potential side effects. Patients had reported pain at the injection site and throughout their entire bodies.

In 2001, Genentech spun out all of its efforts to develop treatments for the nervous system. Arnon Rosenthal, the Genentech executive who left to run the spin-off company, was
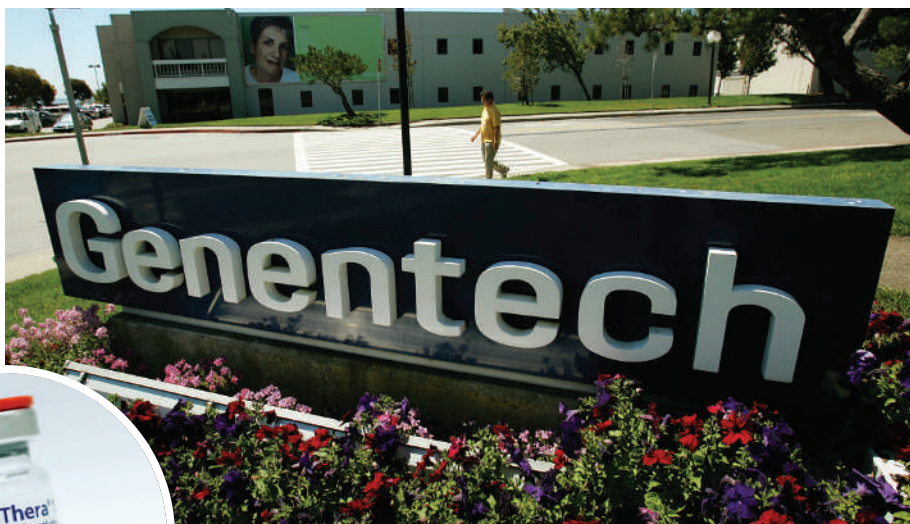
Genentech's success with rituximab drew big drug companies to antibody research.

intrigued by the perplexing side effect of the nerve growth-factor trials. If injecting the protein caused pain, he wondered whether blocking the protein might relieve it. Some thought this idea wildly impractical, because nerve growth factor is thought to be essential for the survival of neurons in the central nervous system.

Nonetheless, Ronsenthal's company, Rinat Neuroscience, decided to test the hunch — after all, the pain market is vast, and no major pain drugs had been developed in almost a decade. The company produced an antibody to block nerve growth factor and tested it in animal models of cancer and arthritis pain, then in human volunteers. Last May, one month after Rinat was acquired by the drug giant Pfizer, company scientists reported to a meeting of the American Pain Society that a single injection of the blocking antibody alleviated arthritis pain in 79 patients for up to two months.

Earlier attempts to treat nervous-system disorders with antibodies ran into difficulties. Biogen Idec and Elan's troubled drug Tysabri, for instance, is intended to prevent the immune cells that cause nerve damage in multiple sclerosis from entering nervous tissues. The antibody binds to a protein known as integrin alpha-4 on these cells. But it had been on the market only a few months when the company had to pull the drug after two patients taking it developed a rare brain infection. The companies conducted a thorough — but inconclusive — investigation to try to discover what could be causing the infections.

The Tysabri episode served as a warning that antibodies for nervous-system conditions can have unpredictable results — but it hasn't slowed the race to find them. Alzheimer's disease is a particularly attractive target, because the potential drug market is huge. The toxic

protein aggregations that cause this brain disease could, in theory, be neutralized or destroyed by an antibody.

One complication is that antibodies are large, and so rarely cross the protective blood–brain barrier. Even so, one clinical trial found hints of efficacy. The trial, conducted by Elan in conjunction with Biogen Idec, used a vaccine designed to spur production of patients' own antibodies against the toxic proteins. The trial was halted because it caused dangerous brain infections in some patients[5] — but autopsy reports suggested that the vaccine did shrink protein clumps in the brain[6]. Moreover, the company says that some treated patients lived more independently than did those in the control group.

Other companies have hurried to start work on better antibodies against Alzheimer's disease — ones that deliver antibodies directly into the body rather than trying to stimulate their production with a vaccine. Elan is farthest ahead, and is running a 240-patient trial in conjunction with Wyeth that is expected to wrap up next year. Other major companies, such as Merck and Eli Lilly, aren't far behind.

Should one of these treatments work, it would mark a watershed in Alzheimer's disease treatment, and a major milestone in the expansion of diseases that might be treated by antibodies. Designers are already thrilled by the new mileage promised by their pimped-up antibodies — and are all too eager to take them for a spin. ■

Erika Check is a senior reporter for *Nature* in San Francisco.

> **"Tysabri served as a warning that antibodies for nervous-system conditions can have unpredictable results."**

1. Hopkin, M. *Nature* **440,** 855–856 (2006).
2. Ward, E. S., Güssow, D., Griffiths, A. D., Jones, P. T. & Winter, G. *Nature* **341,** 544–546 (1989).
3. Nellis, D. F. *et al. Biotechnol. Prog.* **21,** 221–232 (2005).
4. Cartron, G. *et al. Blood* **99,** 754–758 (2002).
5. Check, E. *Nature* **422,** 370–372 (2003).
6. Nicoll, J. A. R. *et al. Nature Med.* **9,** 448–452 (2003).

Ian Tomlinson backed the idea that fragments of antibodies could be effective on their own.

Arizona State University's president Michael Crow wants to shake up the hierarchy of American universities.

THE ARIZONA EXPERIMENT

A shift in population, money and political influence to America's 'sunbelt states' is helping to reshape its research universities. The first of two features looks at the far-reaching ambitions of Arizona State University. The second asks whether a rush to create extra medical schools could spread the region's resources too thinly.

It is a hot February morning in the Arizona desert, and Walter Cronkite, the legendary American newscaster, is straining every muscle in his 90-year-old body to break the hard ground with a golden shovel.

Cronkite is in Phoenix to start construction on a new home for America's biggest journalism school, in America's largest university, in what will soon be its third-largest city. He has some generous words for his host, the president of Arizona State University (ASU). Michael Crow, Cronkite tells the crowd gathered for the ground-breaking, is "a true visionary of our time. He entered the city and took the reins of the university, and gave it direction and energy beyond what anyone could have imagined."

Strong words coming from the 'most trusted man in America'. But the energy is palpable across ASU, including its campus here in downtown Phoenix where construction cranes speak to Crow's ambition. In the past four years, since he left Columbia University in New York to take the reins in Arizona, Crow has had one goal in mind. Put simply, he wants to leave behind the Harvard template, and build a new American university for the twenty-first century.

The key to Crow's vision is to break away from the department-based model of most universi-

ties, and instead build up excellence at problem-focused, interdisciplinary research centres. US research universities, Crow argues, "are at a fork in the road: do you replicate what exists, or do you design what you actually need?" By his reckoning, centres that teach students to communicate with the public and to tackle real problems, such as water supply, are more relevant to today's needs than, say, a chemistry department.

Crow's ideas for ASU have some powerful supporters. "It has become a very different and very exciting institution," says Frank Rhodes, former president of Cornell University in New York and the one-time chair of the US National Science Board. "It is going to be a prototype for the rest of the country."

Not everyone is convinced. Some think that Crow has over-reached, attempting to turn a public university with a mixed reputation into a research hub of international repute. For instance, critics have attacked plans for a medical school in Phoenix — supported by Crow, but being built by Arizona State's erstwhile rival, the Tucson-based University of Arizona — as being extravagant and politically inspired (see page 971). In addition, Crow has been involved in noisy public disputes with ASU's student newspaper over allegations that he tried to censor

its content to please Ira Fuller, a Mormon construction magnate who has donated more than US$160 million to various university projects.

Before arriving at ASU, Crow had a reputation as a talented but headstrong university leader. A political scientist who specialized in science and technology policy at Iowa State University, he entered full-time university administration as a vice-provost at Columbia University, one of the top private research universities in the United States. There he helped to establish the Earth Institute — now led by economist Jeffrey Sachs — to tackle interdisciplinary environmental problems. He also pursued a vainglorious effort to save Biosphere 2, the Earth-sciences experiment-cum-greenhouse built in the Arizona desert and funded by Texan billionaire Ed Bass. It was his sojourns to Biosphere 2 that first drew him to the youngest and, arguably, brashest of the contiguous United States. "I liked the attitude here," Crow recalls.

Talk to any academic who has accepted or rejected a position at ASU recently — and there are plenty of them — and this attitude invariably comes to the fore. For your typical American university professor from either coast, the idea of moving to Phoenix is about as appealing as a stint in the nineteenth-century wild-west community

C. CARLSON/AP

portrayed on HBO's series *Deadwood*. Yet the size and the sheer energy of the city and the project can overcome initial misgivings. A surprising number of top-flight individuals — from Nobel-prize-winning economist Edward Prescott to the biologist and former research chief of SmithKline Beecham George Poste — have taken the plunge.

ASU was already growing its research from a modest base, with an interdisciplinary bent, before Crow turned up. In 2002, the university was involved in setting up the Translational Genomics Research Institute (TGen), a genetics research centre run by Jeffrey Trent, former scientific director of the US National Human Genome Research Institute. Today TGen has about 300 researchers, an annual research income of $60 million, and eight spin-off companies under its belt.

Crow's role has been to publicly raise the flag of bold reform, get politicians and philanthropists on board, sign up some star academics and build interdisciplinary centres to house them. Those established under his tenure include the physically spectacular, $150-million Biodesign Institute led by Poste; a School of Earth and Space Exploration (SESE), headed by geologist Kip Hodges; and the Consortium for Science, Policy and Outcomes run by Dan Sarewitz, a former Democrat staff member of the House of Representatives.

Stylish approach

The Biodesign Institute, whose building won R&D magazine's award last year for the finest new laboratory in the United States, houses 700 staff, including 100 faculty members who are collaborating, drug-industry style, on new approaches to molecular biology and genetics. Biology, computing and engineering in particular, but also law, social sciences and other specialties, are brought into the mix. Last year, the institute attracted about $60 million in public research funding. If all goes to plan, two additional laboratory buildings will be built by 2009, at a further total cost of $300 million.

Poste has the air of someone more accustomed to giving orders than following them,

Arizona State University's Biodesign Institute is set up to foster collaborations between researchers.

ARIZONA STATE UNIV.

"ASU is the most radical experiment going on in American higher education."

— George Poste

but, like many of Crow's recruits, he speaks with an almost-childlike enthusiasm for the project. ASU is "singularly the most radical experiment going on in American higher education", he says. "This is the fastest-growing metropolitan area in the United States, and its largest constellation of undergraduates. This isn't just about the research; it is about the future of these young people."

Michael Tracy, deputy director of the institute, admits that he hesitated before coming to Arizona from his previous position at Stanford Research International, a contract research group in California. But he says that he has been impressed by the extent to which Arizona's residents have bought into the university's plans. "Local people realize that the area needs a high-value proposition," he says. "They have really embraced the idea."

Kip Hodges, who came from Massachusetts Institute of Technology (MIT) in Cambridge last July to lead ASU's new School of Earth and Space Exploration, shares this enthusiasm. "It is a wonderful thing to be part of a place that is becoming, rather than a place that has been," he says of Phoenix.

SESE, which has 36 faculty from many disciplines, hopes to be in a new, purpose-built building by 2009. It aims to pull together expertise in engineering, computation and Earth and space science (ASU is a leader in Mars exploration) to obtain a better understanding of problems on this planet and farther afield. At first, Hodges suspected that ASU's emphasis on serving Phoenix was parochial — "after all, what has MIT ever done for Boston?" he asks rhetorically. "But for SESE, it is a matter

of the relationships between the Earth and society." For example, the centre plans to study how society coevolves with changes in water resources. "It is obviously important for Phoenix, but it is important for the rest of the world, too," Hodges says. Addressing such problems requires a huge collection of skills, including archaeology, the physics and chemistry of water, evolution, anthropology, human ecology, climate and palaeoclimatology. The centre aims, for instance, to build a comprehensive model of the entire Colorado river basin.

Heat islands

Water issues are also to the fore at the Global Institute for Sustainability, an interdisciplinary centre led by Chuck Redman, an anthropologist and long-time ASU faculty member. The institute brings together about 50 faculty, all of whom also have departmental appointments, to study the relationship between people and the environment, especially in urban areas. A focus of interest, Redman says, is to develop building materials and coatings that are suited to 'heat island' cities such as Phoenix, where temperatures can exceed the surrounding area by as much as 8 °C.

Unlike these other centres, the Consortium for Science, Policy and Outcomes is a unit that Crow founded in Washington DC and brought with him to ASU. On Tuesday mornings, he even teaches class there, engaging in a double act with centre director Sarewitz in a three-hour seminar with about 30 undergraduate and graduate students. Crow's a talented teacher. He pulls students into Thomas Kuhn's ideas on how scientific paradigms change by
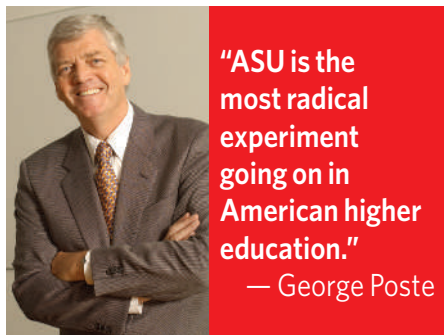
T. TRUMBLE/ARIZONA STATE UNIV.

alluding to a friend who died of the immune disease lupus, and illuminates policy questions by referring to people he knows, such as Jack Marburger, the US presidential science adviser, and Jim Collins, an ASU biologist who runs the National Science Foundation's biological-sciences directorate.

All of this is part of an overall picture that puts strong precedence on three things: high-quality, interdisciplinary research; access for large numbers of students from Phoenix's racially and socially diverse population; and relevance to the needs of the city and the region. Parts of the vision predate Crow's arrival, but the university has become very much Crow's project. "You have got to give him credit for attacking on multiple fronts simultaneously," says Trent.

Fast forward

Phoenix may be the ideal place to host the vision. "This is a university that is being built at the same time as the city is being built," says Crow. On a wall in his office, an aerial view displays the phenomenal growth of a city whose developed area already exceeds that of Los Angeles. The population of Phoenix and its suburbs has grown rapidly to nearly 4 million, and is projected to reach around 8 million within the next quarter century.

Yet for Crow and his disciples, this newness is the greatest part of the appeal. The city has no establishment to overthrow; even Los Angeles is an establishment town by comparison. Asked why people come here, Crow's answer is straightforward: "Quality of life," he says. For most of the city's youthful population that means space to live and drive, not opera or bookshops. That suits Crow, the son of an car mechanic, just fine: 'Someone once said to me, 'some of you blue-collar PhDs are quite smart,'" he recalls with glee.

The question of what kind of university will best serve this new American university is yet to be answered. The model Crow mentions most frequently is that of the University of London which, although now fraying, includes everything from the élite Imperial College to stalwart specialized colleges such as Goldsmiths and

"It is a wonderful thing to be part of a place that is becoming, rather than a place that has been."
— Kip Hodges

Path to success: ASU will provide resources for a diverse community.

Birkbeck, which provide vocational training and adult education.

Arizona, with its booming population and slim infrastructure, faces some of the same challenges as Victorian London — and it sees strong universities as part of the solution. The relationship between public universities and legislature in most American states is traditionally a difficult one. Universities are regarded by many state legislators as expensive and dangerous hotbeds of radicalism and free love (a relationship searingly portrayed by Jane Smiley — a friend and former Iowa State colleague of Crow's — in the comical novel *Moo*). But in Arizona, the free love is between the legislature and the state university system, of which ASU is the largest part.

Conservative lawmakers and the state's liberal governor, Janet Napolitano, are united in their support for the university's expansion. A state referendum in 2000 allocated $1.5 billion in sales tax revenue over 20 years to research in the state university system. Two years later, the legislature gave $400 million more.

And private money is flowing too. In a young city with a thin scholarly tradition, private philanthropy should be a hard sell. But the involvement of people like Fuller, who never attended the university but delivered newspapers near its main Tempe campus, and William Carey, a financier who gave $50 million to ASU's business school in 2003, have shown that this door, too, can be prised open.

The programme still has its detractors. Academics who don't agree with the new interdisciplinary paradigm say that they have been

trampled underfoot. Take Robert Pettit — a chemist and long-time director of the Cancer Research Institute at ASU until he lost the position in 2005. Pettit had clashed with Crow over various issues, including the relationship between his lab and new, interdisciplinary centres at the university. Still a tenured professor at ASU, Pettit says that the arrival of the new institutes "has been very destructive to faculty and student morale and has placed ASU in serious financial jeopardy".

Richard Zare, head of chemistry at Stanford University and another former chair of the National Science Board is, like Rhodes, an authority on US research universities. But he thinks that ASU's emphasis on the interdisciplinary may come unstuck. "It will learn, as others have in the past, that you can't have strong interdisciplinary programmes without strong departments," Zare says.

But to Crow, the hierarchy among established US university departments is too rigid. "If you are the 25th-best geology department, you are trapped!" he says. "Your chances of getting to be fifth-best are zero." The interdisciplinary approach, he says, offers more promise.

Meaningful measures

Measuring achievement for the new ASU falls to the university provost, Betty Capaldi. A typical Crow hire, Capaldi was already co-leader of a project that collates detailed, comparative statistics between US universities. So no one is likely to accuse her of fudging the numbers. Capaldi, however, is rather coy about how the performance of the new ASU should be measured. The centres can "use anything they find meaningful", she suggests. "We are asking every unit to talk to us about how they would measure success. If you are unique — compare yourself to yourself," she says.

It will take years to determine whether the experiment has been a success. Senior ASU faculty assert that the commitment to interdisciplinary research and broad student access have already developed deep roots at the university. And the enthusiasm that it has rekindled in such diverse and seasoned characters as Walter Cronkite and George Poste is certainly infectious.

But the entire project undoubtedly hangs on Crow's unique style of leadership. The approach is unusual, as Rhodes notes, because presidents of US universities often have to content themselves with refining what they have already got. "The polishing of the status quo is much more comfortable," he says. At ASU, he says, "there is a kind of rugged individualism that says — we will just make this happen." ∎

Colin Macilwain is a reporter and editor for *Nature* based in Edinburgh.

Modesto Maidique, president of Florida International University, unveils plans for a new medical school in Miami.

Medicinal properties

From Florida to California the story is much the same. Local political leaders, noting that there are too few doctors to serve their state's growing population, are arranging to build new medical schools. These, they say, will perform cutting-edge research and generate successful biotechnology companies, as well as training the much-needed doctors.

The projects are planned on the basis that they will do research that attracts funding from the National Institutes of Health (NIH), the main US biomedical-research agency. But the NIH's budget, which was doubled from $13 billion to $27 billion by 2003, has since stagnated — raising questions about where the funds will come from.

Critics also argue that there may not be enough qualified students who will want to enrol at the medical schools. The doubters say that there's little likelihood that the new schools will generate clusters of biotech companies and the associated economic benefits promised by their champions. In short, the critics charge, taxpayers in the cities in question have been hoodwinked into backing expensive white elephants that will weaken the states' existing medical schools, leaving medical education and research worse off than they were before.

"This is a very expensive experiment," says Joseph Cortright, a consulting economist in Portland, Oregon, and co-author of *Signs of Life*, an analysis of biotechnology clusters published by the Brookings Institution in Washington DC. "There is no way to determine if it will succeed."

Medical-school campuses, some costing several hundred million dollars, are currently planned or being built in Athens, Georgia; Houston, Texas; three different cities in Florida; Phoenix, Arizona; and Riverside, California. Of the various states pursuing this policy, Florida has proved particularly keen. During his time in office, former governor Jeb Bush championed initiatives that will expend more than $1 billion on building and equipping biomedical-research institutes in the state (see page 1112).

Class act

In the past decade, the Florida legislature has created three new medical schools — two of them in the past 12 months. Florida International University in Miami and the University of Central Florida in Orlando will each open medical schools that will accept their first classes in 2009.

"Florida is a rapidly growing state, with an increasingly affluent and ageing population that accesses health care frequently," says neuroscientist Terry Hickey, provost at the University of Central Florida. "We are convinced there is a physician shortage; by 2020, that situation will be dire."

Officials at established Florida medical schools are not impressed by the plan, however. Abdul Rao, vice-dean for research at the University of South Florida in Tampa, says existing institutions and universities are being undercut. "We need continued support for existing medical schools," Rao says.

When researchers receive a grant from an agency such as the NIH, the university to which they are affiliated gets additional funds to cover overheads. But for the past three years, the NIH's total budget for these operational costs has remained flat at $5.9 billion. That means the new medical schools that come online are simply increasing competition for fixed resources.

Some specialists also take issue with the largest single political selling point for the new schools — the need for more doctors in states with booming populations. Officials at the University of California, Riverside, for instance, argue that the planned $500-million medical school set to open in 2012 is needed to help reverse a shortage of doctors in the surrounding area.

This point of view has some sympathizers. Edward Salsberg, a veteran healthcare planner who now directs the Center for Workforce Studies at the Association of American Medical Colleges in Washington DC, says that the United States needs to increase the number of doctors it trains by 30% by 2020.

But David Goodman, a paediatrician at Dartmouth University in Hanover, New Hampshire, who leads a project that assesses US healthcare needs, takes issue with this. He

> "The free market doesn't work for physician distribution — you have to have incentives."
> — David Goodman

Medical schools under construction on sites such as Lake Nona in Florida (left) may help to produce more doctors to treat Florida's poor.

says that the problem is not how many doctors there are, but how they are distributed. "There has been fantastic spinning of the need for more physicians," he says.

In the United States, a private doctor can practise anywhere and, as a result, more tend to congregate in high-income or desirable communities, leaving some areas underserved. Miami is a prime example, as about three-quarters of its residents live in areas that are medically underserved, according to Goodman's analysis (D. C. Goodman *et al. Health Affairs* **25,** 521–531; 2006). But because of its pockets of wealth, he says, Miami actually has far more doctors than some other US cities. It has 40% more per capita than Minneapolis, Minnesota, for example. But on average, people in Miami live shorter lives, are subjected to more treatments and are less happy with their care than patients in Minneapolis, he says.

School of thought

Goodman claims that training yet more doctors will just add to the excess without addressing the needs of the medically underserved. "The free market doesn't work for physician distribution," he says. "You have to have incentives to have physicians practise in needy communities."

History also suggests that when states build medical schools, the resulting graduates won't necessarily practise in the vicinity. Doctors in the United States typically take speciality residencies that last up to six years, and can be anywhere in the country. In 2005, for example, Florida State University in Tallahassee lost half its first medical class to residencies in other states.

And another objective — that of broadening the diversity of communities from which medical students and researchers are drawn — is problematic, too. Public-health specialists all agree that recruits need to be drawn from disadvantaged communities in order to provide a better service for these groups.

In Florida, the greatest needs are in African-American and Hispanic communities. But among Florida State University's first graduating class of 27, there were only two African-Americans. And among the 50 medical students graduating this spring, only three are African-Americans and three Hispanics.

To fill up places, many medical schools in the states in question already draw students from outside the United States and pay their full tuition and stipend costs. In Florida State University's PhD programme, which began in 2004, half the 21 students are undergraduates from overseas universities. "The number of applicants was very low; kind of pathetic," says microbiologist Myra Hurt, associate dean of research at Florida State University's medical school. "But we are committed to developing a very robust research programme."

Joint venture

In Phoenix, Arizona, a new satellite campus of the University of Arizona College of Medicine, based about 200 kilometres away in Tucson, is taking shape. City leaders hope that the school, which is now recruiting its first class of 24 students, will help to invigorate downtown Phoenix, strengthen biomedical research at Arizona State University (see page 968) and spawn a biotechnology hub.

The Phoenix school will be operated under a partnership between the two state universities. Arizona governments have so far put more than $130 million into the project, which is projected to cost around $470 million. In five years' time, the campus hopes to accommodate a total of 450 medical students, nearly 500 PhD students and almost 300 postdoctoral researchers.

"If you want to crack into the top 50 US medical schools, you have to grow," says Phoenix lawyer Gary Stuart, a member of the Arizona Board of Regents. "But the University of Arizona medical school in Tucson couldn't grow, because it is land-, patient- and resource-locked. We wanted to make something bigger, better and faster. The Phoenix campus made sense at every level."

> **"If you want to crack into the top 50 US medical schools, you have to grow."**
> — Gary Stuart

Yet academics at the existing College of Medicine in Tucson view the project with some apprehension. "Everybody has concerns about how they will pay for this huge effort," says biologist Stuart Williams, who recently resigned his position as chairman of college's bioengineering department to take up a post at the University of Louisville in Kentucky.

Raymond Nagle, a pathologist and former deputy director of Arizona's only federally designated cancer-research centre in Tucson, shakes his head when he hears news reports pledging that the new campus will create a biotechnology industry in Phoenix. "I wish them well; I hope they succeed," he says. "I just haven't seen the evidence they will."

And Joe Panetta, head of the consultancy BIOCOM in San Diego, says it is too late for the areas around the new medical schools to compete with the likes of San Francisco, Boston and Research Triangle in North Carolina. Arizona lacks the venture capital, industrial partners and specialized legal and scientific services. But legislators are optimistic that one day their cities can host all of these services with a brand new medical school at their core. ■
Rex Dalton, *Nature*'s **West Coast correspondent.**

A logged forest in Borneo is better than none at all

SIR — We welcome your encouragement for integrating conservation with other land use in Borneo ("Timber and tapirs" *Nature* **446,** 583–584; 2007). However, your picture of rampant logging and forest destruction in Indonesian Borneo (Kalimantan) requires modification. Many Indonesian timber companies now contribute to conservation.

About 10% of Borneo is under strict protection. If no more than this forest is maintained, habitat loss and fragmentation will have a severe impact on many rare and wide-ranging species such as the Bornean clouded leopard *Neofelis diardi*, or the endangered Storm's stork *Ciconia stormi*. Maintenance of any additional forest offers numerous potential conservation benefits.

With half of Borneo's remaining forests, about 200,000 square kilometres, under active forestry concessions, these areas are of key conservation importance. For example, we estimate that 75% of the Bornean orangutans *Pongo pygmaeus* live in forest concessions.

Given political realities, extensive forest areas will endure only if they yield economic benefits. Production forestry in Borneo's rain forests is selective: only a few trees are removed from each hectare, and what remains is still forest. We recently reviewed how such practices affect Borneo's wildlife (E. Meijaard *et al. Life after Logging* CIFOR, 2005). We found that, for forest fauna, logged forest is considerably better than no forest.

Strict protection status currently makes little difference to forest loss in Kalimantan (L. M. Curran *et al. Science* **303,** 1000–1003; 2004). This reflects the challenge of patrolling and managing extensive areas with limited resources. In contrast, many timber companies have the capacity to manage and protect large areas of forest — and it makes good business sense to do so.

Timber companies wish to access the burgeoning 'green market' in certified timber. Four Indonesian natural forest concessions have already achieved internationally recognized Forest Stewardship Council standards, and more are trying to do so. This demonstrates a commitment to conservation-friendly management. Our monitoring of one of these concessions implies forest losses below 0.1% per year, compared with the Kalimantan average of 2% (D. O. Fuller *et al. Conserv. Biol.* **18,** 249–254; 2004).

We urge wider recognition and support for such conservation hopes. Without this support, forests will continue to disappear.

Erik Meijaard*, Douglas Sheil†
*The Nature Conservancy, Indonesia Forest Program, Balikpapan, East Kalimantan, Indonesia
†Center for International Forestry Research, Bogor, West Java, Indonesia

Millennium: invest in country statistical systems

SIR — Your Editorial "Millennium development holes" (*Nature* **446,** 347; 2007) is timely in highlighting the complexities of monitoring the ambitious development goals on which the world is focused. As you note, the uneven quality of data means that our confidence in reported or predicted achievements varies by country and by indicator. Your call for more investment in evidence-based approaches is welcome, but the solutions are more complicated than this.

A significant handicap in a country's efforts to evaluate interventions is the requirement, by multiple agencies, to monitor multiple indicators for multiple internationally led projects. Its scarce resources can be undermined by the creation of parallel reporting structures, by demands for overlapping surveys for different purposes, and by financial support that is skewed to meet the donor's needs to report internationally. This results in an ever-widening gap between national capacity and international expectations, and the influx of more international experts to fill the 'holes'.

Agencies and donors want national estimates of Millennium Development Goal indicators in order to make international comparisons and to monitor their investments. Because the underlying country data are often weak, agencies develop predictions or estimates to fill data gaps. But countries need more than national averages; they require intimate knowledge of changing disparities in indicators of access and outcomes, by administrative areas and between socio-economic groups. This knowledge can be gleaned only from empirical data collected through information systems that are reliable at sub-national level.

Additional investment is required, but it should be directed to supporting countries' efforts to strengthen their own statistical systems in order to produce the evidence they need nationally and sub-nationally. Until international agencies agree to provide coordinated support, the disparities will remain unnoticed and the 'holes' unfilled.

Sarah B. Macfarlane*, Madeleine Thomson†, Carla L. AbouZahr‡
*Global Health Sciences, University of California, San Francisco, Box 0443, 3333 California Street, San Francisco, California 94143, USA
†International Research Institute for Climate and Society, Earth Institute at Columbia University, New York, New York 10964-8000, USA
‡Health Metrics Network, World Health Organization, 27 Avenue Appia, 1211 Geneva 27, Switzerland

Millennium: big effort has produced statistical results

SIR — Your Editorial "Millennium development holes" (*Nature* **446,** 347; 2007) states correctly that better data are needed to track progress towards the Millennium Development Goals. In 2002, the United Nations secretary-general mandated UN agencies and other international organizations to provide the best available data to monitor progress toward the agreed goals. It became immediately evident that, as you state, many countries lacked the capacity to produce and report the necessary data.

Now, all but 17 of the countries involved have trend data for at least half the indicators.

The international agencies to which you refer carefully review available national data sources used in compiling the indicators, and formulate methodologies when estimates are needed to assess trends in the various regions. These methodologies are then reviewed by international and national experts.

Data gaps have been clearly identified, and efforts to assist countries in the production and use of the necessary data have been scaled up. Statisticians from UN member countries have also reviewed the quality and availability of data to monitor the goals and have provided recommendations. At the 2006 and 2007 meetings of the UN Statistical Commission, a forum for the heads of national statistical systems, more than 130 countries reported on their progress in implementing these recommendations. Although noting that deficiencies still exist, the commission agreed that real progress has been made, and called for improved funding and political commitment to support the development of statistics. We believe the national and global statistical systems have benefited immensely from these efforts.

The global statistical system has made a huge effort to improve data quality and availability, from helping to conduct censuses and surveys in difficult areas to improving vital registration systems. This has produced visible results.

In addition, rather than using lack of data as an excuse for inaction, many countries have increased the use of existing data to prepare and implement goal-based strategies and to conduct rigorous research and assessment of their programmes.

Paul Cheung
United Nations Statistics Division, 2 UN Plaza, DC2-1670, New York, New York 10017, USA

COMMENTARY

When good drugs go bad

How can we best reduce the risk of severe adverse reactions to marketed drugs? An international group of scientists argues that a global research network is needed to identify genetically at-risk populations.

Safety issues can arise throughout the life-history of a drug — from preclinical screening through to clinical trials and, importantly, after the drug is marketed and tested for the first time on the population at large[1]. Although serious adverse drug reactions (SADRs) that can cause fatalities or severe morbidity are relatively rare once a drug is marketed, they are estimated to be the fourth leading cause of death in the United States, not far behind cancer and heart disease[2]. It is estimated that each year about 2 million patients in the United States experience an SADR when using marketed drugs, resulting in 100,000 deaths. Similar numbers have been estimated for other Western countries[3].

SADRs can lead to drug withdrawals, depriving some patients of otherwise beneficial drugs, and unfortunately, may not be recognized for years after a drug has been on the market. Regulators, drug companies, physicians and their patients would all like tools to better predict the apparently unpredictable.

Nineteen drugs have been withdrawn from the US market since 1998 because of unpredictable patient fatalities (see table, page 977; ref. 4). Yet what constitutes an unacceptable SADR is not easily defined: it depends both on the seriousness of the disease being treated and the availability of safer alternative therapies. For example, for life-threatening cancers, potentially fatal SADRs will be tolerated, but for an antihistamine, even the slightest risk of an SADR is not acceptable. Regulators therefore have to review the evidence for individual drugs on a case-by-case basis.

Predictive tests

The consequences for any patient who experiences an SADR can be catastrophic. In an era when patients have many therapy choices, it is unthinkable that selecting drugs for individual patients remains an empirical exercise, and that risk factors for many SADRs — and in particular genetic risk factors — remain largely unknown.

One possible solution is to develop predictive genetic tests for SADRs. Since the sequencing of the human genome, there has been much talk about personalized medicines targeted to individual patients. We share these dreams, but it is unlikely that such research will

"It is unthinkable that selecting drugs for individual patients remains an empirical exercise."

lead to a large number of ('yes' or 'no') litmus tests for SADRs — as with all medical tests, the results will require interpretation. Most probably, a useful genetic test may reveal whether a risk factor is present — thereby increasing the probability of an SADR — or absent, reducing the probability. Such tests are far from perfect, but they can provide doctors and patients with many more options than they have now; the doctor can decide to use another drug, adjust the dose or monitor the patient more closely. In addition to generating predictive tests, we believe an important goal of pharmacogenomic studies of SADRs is to understand molecular mechanisms for developing safer drugs.

What are the possibilities for discovering genetic predictors of SADRs? An early example showing that genetics can help to determine apparently unpredictable drug toxicity was the finding during the Second World War that an inherited deficiency of glucose-6-phosphate dehydrogenase can cause severe anaemia in servicemen treated with the antimalarial drug primaquine[5]. Data gathered over subsequent decades have shown that genetic susceptibility contributes to other SADRs, and several genetic tests for predicting SADRs with known mechanisms have already been approved for clinical use[6].

Thus, early successes led to the identification of single-gene variants causing unusual drug responses. More recently, progress in genomics, including the International HapMap Project (www.hapmap.org), has created opportunities to discover additional genetic risk factors for SADRs. However, there are many serious issues and challenges, scientific and societal, that need to be addressed if the promise of applying genomics to the prediction of SADRs is to be fulfilled. The research required is expensive, with no guarantee of success. Further, even if tests are developed, they may not be cost-effective for clinical use or they may over- or under-predict the patients who are truly at risk from SADRs. It is also possible that certain SADRs do not have a significant genetic component or that risk is a complicated mixture of environmental and genetic factors that are too difficult to sort out.

Unknown mechanisms

What are the biggest research challenges to overcome? Interactions between genetic and non-genetic risk factors are unavoidable in many SADRs. For example, a patient's overall health status may place them at a higher risk of an SADR, so those with heart disease will have increased risk of drug-induced cardiac arrhythmia. Similarly, interactions between multiple drugs amplify the risk of SADRs,

contributing to several drug withdrawals in the past decade (see table). These issues also complicate identification of SADR cases; for example, when monitoring liver toxicity, should people with pre-existing liver disease be excluded?

Another challenge in discovering genetic markers for risk of SADRs is that in most cases we have little or no information about the underlying biological mechanisms. Experience shows that preclinical safety testing in animal models often cannot predict SADRs. For example, there was no hint from animal studies of the risk of severe shock and liver and kidney damage that was suffered by the six individuals who took part in the first-in-man evaluation of the monoclonal antibody TGN1412 last year.

Most SADRs can be classified into two groups: those that can be predicted because they are related to the drug's beneficial effects (for example, bleeding complications with anticoagulant therapy) and SADRs that are unrelated to the drug's effects (such as unexpected liver toxicity). In the latter case, it is much harder to identify genetic risk factors that may predict SADRs because no mechanistic information is available. In the absence of biological understanding, one potential approach is to apply genome-wide scans to patient groups with identifiable SADRs to discover genetic variations (for example, single nucleotide polymorphisms) associated with that group.

Upping the numbers

But by far the most challenging issue in identifying genetic risk factors for SADRs in marketed drugs is that the numbers of cases are too small for adequate genetic analysis. This problem is compounded by failure to identify drugs as the causative agent in many SADR cases or by under-reporting of SADR cases by health practitioners. Under existing systems of voluntary reporting, some estimates suggest that only 1–10% of cases are ever reported[7]. To acquire sufficient numbers for statistical analysis, it is usually necessary to combine

This multiple-sclerosis patient lost out when natalizumab was withdrawn from the market.

For a number of drugs, serious adverse reactions are detected only after they have gone on the market.

cases across diverse clinical settings.

Hence, identifying predictive genetic markers requires greater numbers of cases and controls in different ethnic groups than are generally available from individual medical centres. Even in the best of circumstances, with the identification of genetic markers that seem to be associated with an SADR, there is the problem of false positives, particularly when the markers are identified by genome-wide scans or testing of many candidate genes during clinical trials. Therefore, although replication of results is now the standard for such genotype–phenotype associations, this may not be possible when studying SADRs in any single patient population given their frequency and potential complexity.

In our view, a global pharmacogenomics network is needed to study SADRs. Often, no single site or even country can generate a sufficient number of cases within a reasonable time. Several multicentre networks are already being developed. The EUDRAGENE project funded by the European Commission involves a multicentre collaboration of ten European countries, thereby targeting a potential patient population of 350 million[8]. Its goal is to establish a freely shared case-control collection of DNA samples for six known SADRs — including liver injury, tendon rupture, muscle toxicity and psychosis.

Another nationwide example of a multicentre consortium for studying SADRs is the Canadian Genotypic Adjustment of Therapy in Childhood (GATC) project that builds upon Canada's national healthcare system, targeting close to 80% of children in Canada. SADRs in children are a significant problem because less than one-quarter of pharmaceuticals licensed in North America have been tested in paediatric populations. Differences in the way children metabolize drugs may cause adverse effects

that are seen less commonly in adults[9].

The experience of GATC and EUDRAGENE highlights several issues that must be faced when forming a global network. To acquire suspected SADR cases, EUDRAGENE relies on existing national or regional systems of spontaneous reporting of cases by physicians. Among cases reported to EUDRAGENE, many are excluded because they fail to meet preset criteria, patients cannot be located after the initial report, or physicians or patients refuse to participate. From these experiences it is estimated that at least 2,000 suspected cases will need to be investigated to obtain 200 to 250 useful cases and controls with DNA samples. This means that for an SADR that occurs in 20 patients per million people per year (such as torsades de pointes, a potentially fatal cardiac arrhythmia), the EUDRAGENE population would generate only 700 cases per year. Active surveillance could result in many more cases being included.

Ethnic diversity

In addition to making it possible to collect adequate numbers of cases and appropriate controls, a global network has the advantage of involving multiple ethnic groups. It is the rule rather than the exception that the frequencies of gene variants involved in the metabolism, transport and action of drugs differ among various populations, with some variants occurring almost exclusively in one ethnic group[10]. A global consortium would also be of great value in identifying environmental, as well as genetic, determinants of SADRs among diverse population subgroups.

Further, a global network would standardize criteria used to identify SADR cases. Currently, criteria for specific SADRs vary considerably among studies, making it harder to replicate individual findings and to combine cases
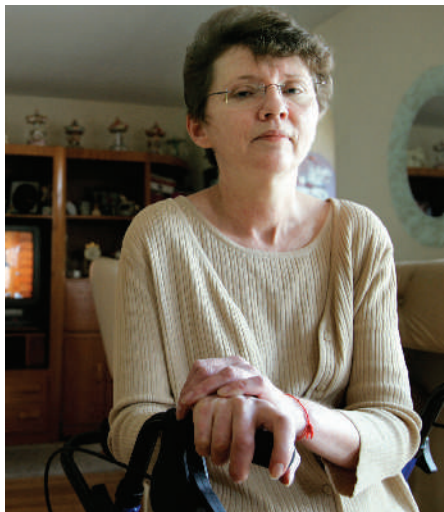
from multiple studies. For example, one of the commonest causes for US drug withdrawal in the past decade has been an adverse reaction that can lead to torsades de pointes[11]. The arrhythmia, although quite distinctive on an electrocardiogram, often requires a specialized cardiovascular practitioner to differentiate it from other cardiac disturbances. Further, some studies focus on early features of the electrocardiogram rather than on the arrhythmia itself.

Similar issues arise when defining SADR criteria for patients experiencing drug-induced liver injury[12]. Here, even significant increases in liver enzymes may not indicate clinically important liver injury: drug-induced elevations in liver enzymes may reverse, even with continued use of the drug, and may not lead to permanent injury. Therefore, identifying true cases of drug-induced arrhythmia or liver injury needs careful criteria and monitoring.

Funding the network

What needs to be done to build a global SADR network? Adequate funding for regional networks of trained doctors, pharmacists and research nurses is essential, as is technical support for medical databases and DNA banking. None of this comes cheap. But it may be achieved at a fraction of the cost of SADRs to the healthcare system. Some estimates put these costs at approximately US$40 billion to $50 billion per year worldwide. And a pharmaceutical company can lose as much as $1 billion if a marketed drug has to be withdrawn. These are costs that are currently passed on to patients.

The United States lacks a national healthcare system and a single post-marketing surveillance system for SADRs, so it is at a disadvantage compared with Canada or Europe, although this is starting to change. The Pharmacogenetics Research Network, sponsored by the National Institutes of Health (NIH), has several centres studying genetic risk factors for SADRs, and is forming partnerships with healthcare systems such as the Health Maintenance Organizations' Research Network and the Veterans Administration hospitals. The US Food and Drug Administration has also initiated collaborations with various stakeholders, including the NIH, academic institutions and the pharmaceutical industry, to focus on SADRs related to cardiovascular and liver toxicity. These multiple activities need to be expanded and coordinated to develop a coherent US strategy for the conduct and funding of SADR research.

These regional SADR networks are an important step towards a global network. Once established, a global network should identify specific SADRs that have the highest priority for study and focus on criteria for diagnosis, study design and data analysis. Genetic analyses of patient groups and healthy controls can build on the genetic data generated

> "A global SADR network will not be cheap, but it will be a fraction of the current cost of SADRs to the healthcare system."

by the HapMap Project. As low-cost genome-sequencing methodologies become available in the next 5–10 years, it will be possible to identify rare genetic variants that associate with SADRs in individual patients.

Efforts to secure sufficient resources for a global network should involve national and international funding agencies (including the NIH, the World Health Organization and the European Commission), along with other stakeholders such as the pharmaceutical industry, regulatory agencies and health-provider organizations. Recent advances provide hope for the continued development of genetic tests to predict SADRs. At least half a dozen genetic tests that can predict adverse drug reactions for a range of treatments are or soon will be ready for use in clinical practice[6]. The availability of these genetic tests are noted in revised product labels that should reduce the incidence of SADRs.

We believe that the promise of emerging genomic technologies offers an opportunity to alert and educate clinicians and the public about the problem of SADRs, as well as to enhance reporting of rare cases that may ultimately reduce the incidence of SADRs. A key first step is the establishment of a global SADR pharmacogenomics network. ■

Kathleen M. Giacomini is in the Department of Biopharmaceutical Sciences, University of California, San Francisco; Ronald M. Krauss is at the Children's Hospital Oakland Research Institute; Dan M. Roden is at the Vanderbilt University School of Medicine; Michel Eichelbaum is at the Stuttgart Institut Klinische Pharmakologie; Michael R. Hayden is at the University of British Columbia; and Yusuke Nakamura is at the University of Tokyo.

1. Institute of Medicine report *The Future of Drug Safety* www.iom.edu/CMS/3793/26341/37329.aspx
2. Lazarou, J., Pomeranz, B. H. & Corey, P. N. *J. Am. Med. Assoc.* **279,** 1200–1205 (1998).
3. van der Hooft, C. S. *et al. Drug Saf.* **29,** 161–168 (2006).
4. Huang, S. M. *et al. Toxicol. Mech. Meth.* **16,** 89–99 (2006).
5. Beutler, E. *Pharmacol. Rev.* **21,** 73–103 (1969).
6. Roden, D. M. *et al. Ann. Intern. Med.* **145,** 749–757 (2006).
7. Hazell, L. & Shakir, S. A. *Drug Saf.* **29,** 385–396 (2006).
8. Molokhia, M. & McKeigue, P. *Pharmacogenomics* **7,** 633–638 (2006).
9. Johnson, T. N. *Toxicology* **192,** 37–48 (2003).
10. Eichelbaum, M., Ingelman-Sundberg, M. & Evans, W. E. *Annu. Rev. Med.* **57,** 119–137 (2006).
11. Roden, D. M. *N. Engl. J. Med.* **350,** 1013–1022 (2004).
12. Watkins, P. B. & Seeff, L. B. *Hepatology* **43,** 618–631 (2006).

DRUGS WITHDRAWN FROM THE US MARKET SINCE 1998

Drug name	Approved	Withdrawn	Use	Risk
Mibefradil	1997	1998	High blood pressure/ chronic stable angina	Drug–drug interactions Torsades de pointes
Bromfenac	1997	1998	Non-steroidal anti-inflammatory	Acute liver failure
Terfenadine	1985	1998	Antihistamine	Torsades de pointes Drug–drug interactions
Astemizole	1988	1999	Antihistamine	Torsades de pointes Drug–drug interactions
Grepafloxacin	1997	1999	Antibiotics	Torsades de pointes
Etretinate	1986	1999	Psoriasis	Birth defects
Alosetron*	2000	2000 (2002)*	Irritable bowel syndrome in women	Ischaemic colitis; complications of constipation
Cisapride	1993	2000	Heartburn	Torsades de pointes Drug–drug interactions
Troglitazone	1997	2000	Diabetes	Acute liver failure
Cerivastatin	1997	2001 (2002)*	Cholesterol lowering	Rhabdomyolysis Drug–drug interactions
Rapacuronium	1999	2001	Anaesthesia	Bronchospasm
Levomethadyl	1993	2003	Opiate dependence	Fatal arrhythmia
Rofecoxib	1999	2004	Pain relief	Heart attack; stroke
Valdecoxib	2001	2005	Pain relief	Skin reactions (SJS)
Natalizumab*	2004	2005 (2006)*	Multiple sclerosis	Brain infection
Technetium (99m Tc) fanolesomab	2004	2005	Diagnostic aid	Cardiopulmonary arrest
Pemoline	1975	2005	Attention-deficit hyperactivity disorder	Liver failure
Pergolide	1988	2007	Parkinson's disease	Valvulopathy
Tegaserod	2002	2007	Irritable bowel syndrome with constipation	Angina; heart attack; stroke

* Remarketed with restrictions. Data from this table are from ref. 1 and were updated by Shiew-Mei Huang.

BOOKS & ARTS

In a hole in the ground...

What happens when you find a hobbit — or a unicorn?

The Discovery of the Hobbit: The Scientific Breakthrough that Changed the Face of Human History
by Mike Morwood & Penny Van Oosterzee
Random House: 2007. 326 pp. Aus$34.95.
To be published as *A New Human* by HarperCollins/Smithsonian in May.

Henry Gee

The unicorn, wrote Jorge Luis Borges (in *Kafka and His Precursors*), is universally regarded as a supernatural being of good omen. But there's a problem: despite its folkloric familiarity, we wouldn't know how to recognize a unicorn if we met one in real life. It "does not figure among the domestic beasts, it is not always easy to find, it does not lend itself to classification," Borges continues. "It is not like the horse or the bull, the wolf or the deer. In such conditions, we could be face to face with a unicorn and not know for certain what it was."

Most science consists of elaborations on well-tested themes. The discovery of something genuinely unexpected, however, kicks over the traces. Even the discoverers find it hard to accommodate the new arrival in a picture of the world that they have, perhaps inadvisedly, taken for granted.

Many manuscripts received by *Nature* are full of the confidence of scientists who know precisely what they have found and why it is important. But a paper that landed unannounced on my desk on 3 March 2004 was surprising, not only for the extraordinary discovery that it reported, but for the matter-of-fact, almost muted, tones in which it was described. Reading between the lines, it seemed as if the discoverers of *Sundanthropus floresianus* weren't entirely clear in their own minds about what manner of unicorn they had unearthed.

The report, although hard to interpret, is easily summarized. Peter Brown of the University of New England in Australia and his colleagues described a hitherto unknown and extinct member of the human family from a cave called Liang Bua on the island of Flores in Indonesia. The skeleton was archaic in form and extremely small — smaller even than Lucy, the primitive, pre-human hominid discovered in Ethiopia whose species (*Australopithecus afarensis*) died out around 3 million years ago.

But there were two big surprises, as Mike Morwood, co-principal investigator of the joint Indonesian and Australian project on Flores, relates in his book *The Discovery of the Hobbit*.

Home to a hobbit: *Homo floresiensis* was found in Liang Bua cave on the island of Flores in Indonesia.

The first is that the Flores hominid lived as recently as 18,000 years ago. The second is that the form of the remarkably tiny skull, especially its teeth, suggested a more derived ancestry, not from *Australopithecus* or any other (unknown) primitive hominid, but from a form of *Homo* — perhaps a population of *Homo erectus* that was marooned on Flores a million years ago and had undergone the kind of endemic dwarfing seen in the native, pony-sized elephants already found on Flores. The prior existence of stone tools dating back 800,000 years from elsewhere on Flores suggested the second scenario. And yet the jarring anachronism of the Flores hominid — literally so, given its recent date and weird morphology — prompted the authors to place their creature in its own genus, at least for the present, as *Sundanthropus*, or Sunda man.

The referees responded with one accord. To be sure, the creature was strange, but the strangeness might be a consequence of its size. The skull, though, was clearly that of a member of our own genus, *Homo*. In addition, one referee commented specifically on the specific name, *floresianus*, noting that generations of students would dub it 'flowery anus'. The authors duly changed the generic and specific name to *Homo floresiensis* and, after several

iterations, that was the name attached to the fossil when the discovery was published in *Nature* on 28 October 2004.

The rest, as they say, is history. In *The Discovery of the Hobbit*, Morwood and science writer Penny Van Oosterzee relate the discovery, its background and the fallout. To be fair, it's not great literature, but it is of vital importance, standing as the first, book-length, 'official' account of the discovery. Here you will learn how and why the researchers came to be in such a remote place, what they found, and the considerable logistical and administrative obstacles they faced.

Disbelief and doubt were not confined to the researchers and their circle. As Morwood recounts, hardly had the finds been made public than commentators queued up to declare *Homo floresiensis* not a new species, but a pathological specimen of modern humans, *Homo sapiens*, perhaps suffering from some form of microcephaly. This was despite the presence of not one specimen but several, now known to have existed perhaps as long as 95,000 years ago.

Such reaction is common in the wake of new hominid discoveries, which are routinely dismissed either as pathological humans (*Homo neanderthalensis*) or apes (*Australopithecus*

REUTERS

africanus and *Sahelanthropus tchadensis*). Such reactions say less about the facts than the mindsets of commentators, who might be unwilling to have their comfortable views of the world so forcibly changed. Confronted with what might be a genuine unicorn, many would prefer to see a pantomime horse with a spike glued to its head.

After that, research fades into hypothesis, and thence into myth. It was perhaps inescapable that the researchers chose to name their find 'hobbit', after those small, secretive and entirely fictitious residents of J. R. R. Tolkien's Middle-earth. It may be telling that the researchers couldn't even agree on this soubriquet, but it was wished upon them by the press.

And what of *Homo floresiensis*? Although the most parsimonious view is that it is *Homo erectus* writ small, Morwood reminds us that evolution is not obliged to be parsimonious,

and that the very existence of *H. floresiensis* reminds us that much remains undiscovered. In which case, perhaps the researchers were right to begin with in supposing that their creature represented a hitherto unknown and primitive hominid that arrived in southeast Asia, from places unknown, long before *Homo* evolved.

The unicorn remains as it always did, frustratingly elusive. This year, the researchers will return to Liang Bua to see if they can discover more. But stories such as this demand a mythological beast altogether less serene. It is as if the researchers had set out to discover some new form of fossil mouse, only to find that they had grabbed a dragon by the tail instead. And as any devotee of Harry Potter will remind you: *Draco dormiens nunquam titillandus.* ■

Henry Gee is a senior editor of *Nature* and the author of *The Science of Middle-Earth*.

A user's guide to technology

The Shock of the Old: Technology in Global History Since 1900
by David Edgerton
Profile/Oxford University Press: 2007.
320 pp. £18.99/$26

Andrew Nahum

As an academic discipline, the history of technology has spent many decades refining models of technological change. There has been a move from a relatively deterministic 'engineering darwinism' to considering technologies, at least in part, as being elicited and selected in the light of social values, aspirations and prejudices.

The polemical thrust of David Edgerton's book *The Shock of the Old*, however, is that this analysis of the succession of technologies, techniques and machines has the wrong focus. It is hopelessly biased towards innovation and novelty, and implicitly creates a tidy 'timeline of progress'. Moreover, things are not always as they seem. For example, B-52 bombers, an important part of the US Air Force today, are almost 50 years old and might be flown by the grandsons of their first pilots. In the Second World War, more than half-a-million horses dragged the 'mechanized' German army into Russia, while Britain's wooden Mosquito aircraft outpaced metal Messerschmitts.

Counter-intuitive examples such as these are the firecrackers of Edgerton's original and timely book, which has been promoted and packaged on the basis of the title and these old–new teasers. However, the author does not fully complete this advertised mission, for his real agenda is to promote an analysis of technology that is rooted in use and in practice, rather than in innovation and invention.

"One particularly important feature of use-based history of technology," Edgerton argues,

"is that it can be genuinely global." Rather than focusing on the dichotomy between the technology-rich developed world and what Edgerton calls the 'poor world', usually presumed to be deficient in technology or furnished only with second-hand or debased versions, user-based history engages "with all the world's population, which is mostly poor, non-white and half female".

This reading of the dislocation between the present reality and our "reheated futurism" is typified by a section on urban development entitled 'Not Alphaville but bidonville', which notes that by the end of the twentieth century, "most of the largest cities in the world were poor places where once Paris, London and New York led in scale and opulence". Whether built

from the oil drums (bidons) of North Africa, or the crates used to import tractors to Durban, these urban centres are home to distinctive, and effective, local 'creole technologies' about which we understand far too little. The book is filled with powerful and thought-provoking examples derived from Edgerton's enormous experience of the mainstream history of technology, but also, unusually, of South America and other parts of the 'poor world'.

With his emphasis on use, the author takes us into areas that most accounts of technology overlook, such as slaughterhouses and execution chambers. In one of the most valuable essays he reminds us that many people (perhaps most?) engage with technology through repair and maintenance. According to this perspective, the technological story of Toyota or Mercedes is not exemplified by the production lines and industrial culture of Nagoya or Stuttgart, emblematic though these may be in technological history. Equally important are the communal repair sites of Ghana known as 'magazines', where tools are rudimentary — "hammers, spanners in incomplete sets, files and screwdrivers" — but where the vehicles are reworked to be indefinitely maintainable, within their local system, and where those who repair them get to know them much more intimately than do drivers or mechanics in any rich country.

The account is a compelling *tour de force* and a corrective. My own institution, the Science Museum in London, does not entirely escape; he scolds it for an 'innovation timeline' that orders the showcase gallery "grandly" entitled 'Making of the Modern World'. However, as curators, we believe it is most effective to engage with audiences, initially at least, through what they know about technology. We may regret that an older historiography of technology dominates public understanding, but this is the case nonetheless, so icons

A hands-on approach: mechanics in Cuba maintain old American cars indefinitely using basic tools.

such as Stephenson's Rocket help us to engage a wide audience. Of course, we also aim to add a deeper and more nuanced understanding, and those who penetrate a little farther into the gallery find an exactly parallel (and very popular) timeline display on the technology of everyday life — familiar and once-familiar artefacts that have earned a presence through use at home, in the workplace and at play.

Icons of invention do exist in our imaginations, but Edgerton seems unprepared to accept that humans navigate their culture by identifying exceptional events and discontinuities. Nor is he concerned with exploring the intriguing mental and cultural processes that make certain inventions or episodes memorable and emblematic. But this is not to deny the great importance, and the interest, of what he has started here.

Ultimately, the pursuit of a more global or democratic history is not unproblematic. What would a use-based history of technology be like, and how would it be more than a natural history? It would not be a replacement for other accounts, perhaps, but one running in parallel.

Finally, there is more to understanding than argument alone, and the author is often most persuasive when he is, perhaps unknowingly, being impressionistic and descriptive. One passage in particular was especially evocative and, almost in itself, seemed to validate the whole historiographic attempt: "Travelling through the poor world it is hard to miss… tiny metal-working shops where the most complex bit of machinery may well be an oxyacetylene, or electric, torch for welding." It is from such shops that "at dusk, bright intermittent light from welding illuminates streets all over the world." ∎

Andrew Nahum is principal curator of technology and engineering at the Science Museum, Exhibition Road, London SW7 2DD, UK.

The fall of a wonder drug

Penicillin: Triumph and Tragedy
by Robert Bud
Oxford University Press: 2007. 344 pp. £30, $55

Hugh Pennington

Alexander Fleming discovered penicillin in September 1928. He sent his initial observations to the *British Journal of Experimental Pathology* on 10 May 1929 and they were published a month later. The paper made little impact. It is very different now, of course. Antiquarian booksellers salivate at the thought of a copy and value it not far short of the price of a first edition of a James Bond book by another Fleming.

The circumstances surrounding the discovery of penicillin have been described many times. Most accounts are celebratory, some hagiographic. The best one, which is neither, is Ronald Hare's *The Birth of Penicillin and the Disarming of Microbes* (Allen & Unwin, 1970). Hare was there, in the Inoculation Department of St Mary's Hospital in London, when Fleming made his discovery.

Robert Bud's book *Penicillin* is the sequel to Fleming's story and is far from celebratory. The development of antibiotic resistance gets more space than tales of therapeutic success; these are the tragedy and triumph of its subtitle. A major theme is penicillin as a brand. Driving the book is the notion that the casting of penicillin as a 'wonder drug' has done more harm than good. The case is persuasive. By the 1950s, penicillin was cheap enough and safe enough to be used freely and on a grand scale by general practitioners and veterinarians, and the perception of its impact caused not only the general public but many medical specialists to imagine they were living in an era that would soon see the end of infectious disease as a public-health problem. Penicillin was given much of the credit for the undoubted decline in infection that had taken place in Europe and North America. But it was the abundantly available clean water for drinking and washing that had seen off cholera, typhoid and typhus.

Penicillin was promoted as a miracle cure in the Second World War.

Better diet was responsible for the decline in tuberculosis that had been going on for a century. Immunization had made smallpox and diphtheria historical diseases. With the benefit of hindsight we can see that the *Streptococcus* bacterium that caused scarlet fever — an organism particularly susceptible to penicillin — was becoming less virulent on its own. But of course, from the moment penicillin became generally available, profligate prescription began promoting the spread of resistant pathogens.

Bud shows that the creation of penicillin as a brand owed much to the Second World War and that the strength of its image served to typify the power of the science that had led to the Anglo-American victory. But he also covers the controversies: how influential politicians, journalists and newspaper owners canonized Fleming, rather than Howard Florey, who led the team that turned penicillin into a drug; and how penicillin joined the Frisch–Peierls memorandum on the manufacture of the atomic bomb and the cavity magnetron as British scientific achievements from which it could be said that the Americans had unreasonably benefited. Bud goes on to recount the development of penicillin derivatives designed to deal with resistant bacteria and public policy responses to the same problem. He ends on a pessimistic note: in essence, the strength of the brand has been outmatched by bacterial evolution.

A focus on the people and institutions involved — scientists, policy-makers, propagandists, 'big pharma' and governments — gives a story full of human interest that general readers will enjoy. But this is only the half of it. By and large, the microbes are missing, although in fairness that would have needed another volume. The abiding triumph of penicillin over *Treponema pallidum*, which turned syphilis from a common scourge to a clinical curiosity, and its difficult relationship with *Staphylococcus aureus*, which was the organism on Fleming's famous Petri dish and continues today as MRSA, are big stories too. ∎

Hugh Pennington is in the School of Medicine, University of Aberdeen, Aberdeen AB9 2ZD, UK.

A sense of proportion

A painting of Federico Zuccari, the founder of the academy of arts in Rome, clearly demonstrates his credentials in geometry.

Martin Kemp

Why, in the portrait shown here, is the founding father of the Roman academy of arts, Federico Zuccari, depicted displaying some rudimentary geometric diagrams? He shows us a point, parallel and perpendicular lines, angles of 60 and 45 degrees, triangles, a circle, a semicircle and a square. Below the diagrams is a sectioned human figure with an oddly detached arm.

He has drawn these diagrams on a board or primed canvas with the white chalk gripped in its metal holder. His right hand, emerging from its white linen cuff, is poised with the precise refinement of a Victorian lady sipping tea from the finest porcelain.

The posthumous portrait of Zuccari, which can be seen in the exhibition 'Dürer e l'Italia' (Dürer and Italy) at the Scuderie del Quirinale in Rome until 10 June, was painted by Giovanni Maria Morandi in 1695 to mark the centenary of the Accademia di San Luca. St Luke was the patron saint of painters, which is why the Italian art academy bears his name. The large creased document in the painting is a Papal bull assigning the Church of Santa Martina to the painters as the base for their devotions.

Zuccari's pens and brushes protrude through the thumbhole of his painter's palette, held by a knotted ribbon. We can easily read the Latin inscription, "AEQUA POTESTAS", a phrase from *Ars poetica* by the Roman author Horace that alludes to the "equal daring" of the poet and painter in creative invention. The painter, then, is the equal of the poet and operates with a high level of papal favour. But why the geometry?

The answer lies with what had become the stock opening sections of Renaissance treatises on geometry for painters. Leon Battista Alberti wrote the first of these, *De pictura*, in 1435. He began with the definitions of point, line, surfaces and triangles, taking his cue from Euclid. Step by step, he built towards the three-dimensional construction of painter's perspective according to the rules of geometric optics.

This euclidian grounding became a long-lasting tradition. When the English landscape painter J. M. W. Turner was appointed professor of perspective at the Royal Academy of Arts in London in 1807, he schooled himself and his charges in the basics of Euclid's *Elements*. To a modern spectator familiar with Turner's evanescent landscapes suffused with veils of colour,

this dedication of the geometric basics comes as something of a surprise.

Zuccari was an intellectually ambitious author on the visual arts, well versed in all the leading ideas. The sectioned figure testifies to his knowledge. It is taken from *Four Books on Human Proportion* by Albrecht Dürer, the German painter and print-maker. Dürer was the first to formulate systems of proportion for figures of all types: fat and thin, tall and short, stocky and attenuated, and even children.

Dürer's books on proportion and geometry have a claim to be the most influential of all artists' books on the worlds of art and science, as they pioneer several practical techniques in geometry. Zuccari has copied the figure, whose head is 1/7 of his total height, larger than the norm.

Geometry was not just useful for perspective and proportions, but was something that should be seen to be known. Emerging in the Renaissance from a predominantly artisanal tradition, artists were keen to brandish the 'science' of their art. Geometry played a crucial role in their quest for more elevated status, both intellectually and socially.

As with any profession that harbours growing pretensions, the public parade of expertise is important. They sold their skills, so there was a little point in painters having knowledge of geometry unless they were seen to have it.

Martin Kemp is professor of the history of art at the University of Oxford, Oxford OX1 1PT, UK. His new book, *Seen | Unseen*, is published by Oxford University Press.

SCIENCE IN CULTURE

Putting the pieces **together**

Disappearing act

The bizarre absence of certain gene classes in eukaryotes is key to understanding their evolution and complex links with prokaryotes.

James A. Lake

Eukaryotic evolution is something of a Gordian knot. Using single genes to unravel it won't work, as the genomes of eukaryotes (animals, plants, fungi and protists) are derived from those of several prokaryotes (eubacteria and archaebacteria). So the focus has shifted towards analysing flows of gene populations, and even of entire prokaryote genomes, into eukaryotes. These more holistic studies are revealing the complex genetic and evolutionary connections between eukaryotes and prokaryotes.

Until recently, everyone assumed, based on a single ribosomal RNA gene, that eukaryotes descended from archaebacteria — extremophilic prokaryotes distinct from 'true' bacteria, or eubacteria. Now we know that's not the case. More than two-thirds of the nuclear genes of the yeast *Saccharomyces cerevisiae*, for instance, are derived from eubacteria, and the balance from archaebacteria. What we know of gene losses and gains also indicates that the eukaryotic genome probably resulted from the fusion of archaebacterial and eubacterial genomes, effectively turning the tree of life into a ring of life. But how did evolution come up with the strange distribution of eubacterial and archaebacterial genes we see in eukaryotes today?

In prokaryotes there are two major gene classes: operational and informational. Operational genes are involved mainly in day-to-day processes of cell maintenance, and code for amino-acid and nucleotide biosynthesis as well as related functions. Informational genes feature primarily in transcription, protein synthesis, DNA replication and other processes to convert information from DNA into proteins.

Because eukaryotes are derived from archaebacteria and eubacteria, one might expect to find an archaebacterial and a eubacterial copy of each nuclear gene. But strangely, archaebacterial operational and eubacterial informational genes are almost completely absent from eukaryotes, even though the first eukaryote contained two sets of informational and operational genes.

This well-documented correlation between phylogenetic origin and gene disappearance is paradoxical because no one understands how these classes of genes left the scene. I call this correlation, which provides an important clue to the early evolution of eukaryotes, the Janus

paradox. Like the two faces of the Roman god Janus, thought to represent the Moon and the Sun, the phylogenetic origins of informational and operational genes in eukaryotes are as different as night and day. Finding a gene distribution such as this is the statistical equivalent of finding that a coin tossed at night (Janus's archaebacterial face) always comes up heads (informational genes), and tossed during the day (Janus's eubacterial face) always comes up tails (operational genes).

But before we look at possible causes of the Janus paradox, we need to understand the interactions between operational and informational genes. In some ways, gene transfers between prokaryotes mimic patterns of telephone use. Some people only call their family and friends, for instance. This is similar to the pattern of transfer of informational genes between closely related prokaryotes. Others add distant associates to their basic phone lists — analogous to the broader transfer of operational genes between prokaryotes. The 'complexity hypothesis' attributes the reduced transfer rate seen in informational genes partly to the observation that their proteins are often deeply integrated into large complexes such as the ribosome. So to function, transferred informational genes must fit into complex pre-existing structures, effectively restricting their transfers to closely related prokaryotes (family and friends). Transferred operational genes, as members of smaller, less complex structures, fit in more easily, allowing them to function when transferred over larger phylogenetic distances (family, friends and associates).

Combining two genomes into one nucleus would not have been simple. We know, for example, that two sets of ribosomal RNA operons — genes controlled as a unit — cannot coexist in the same cell. Recent attempts to fuse a cyanobacterial genome into the genome of a host *Bacillus subtilis*, a common soil bacterium, were successful only when the cyanobacterial ribosomal RNA operons were removed from the fusion chromosome. We also

know from experiments on reconstituting ribosomal subunits that, consistent with the complex interactions within the ribosome, even a single damaged protein can completely inactivate protein synthesis.

This suggests a possible explanation for why the eubacterial informational genes disappeared. Because two types of ribosomal genes cannot exist in the same nucleus, the archaebacterial ribosome may simply have been the lucky survivor when one of the components in the eubacterial ribosome was inactivated. Once this chance inactivation occurred, it was probably only a matter of time until all eubacterial informational genes were eliminated, given the extensive interactions between the ribosome and informational proteins.

Unfortunately, I have no good suggestion for why the archaebacterial operational genes were eliminated. I hope that this will motivate some readers to think of hypotheses and experiments. It could be that somehow, the ready availability of operational genes within the eubacterially derived cellular organelles led to the preponderance of eubacterial operational genes.

Whatever explanations of the Janus paradox are unearthed, it will be exciting to follow the quest. How the eukaryotic cell came to be is one of the greatest enigmas in biology. It is a story so complex that no single gene can tell it. Only entire genomes can.

James A. Lake is distinguished professor in the Departments of MCD Biology and Human Genetics, and in the Molecular Biology Institute, University of California, Los Angeles, 242 Boyer Hall, Los Angeles, California 90095, USA.

FURTHER READING
Martin, W. & Embley, T. M. *Nature* **431,** 134–136 (2004).
Hey, J., Fitch, W. M. & Ayala, F. (eds) *Systematics and the Origin of Species On Ernst Mayr's 100th Anniversary* (The National Academies Press, Washington DC, 2005).
Jain, R., Rivera, M. C. & Lake, J. A. *Proc. Natl. Acad. Sci. USA* **96,** 3801–3806 (1999).
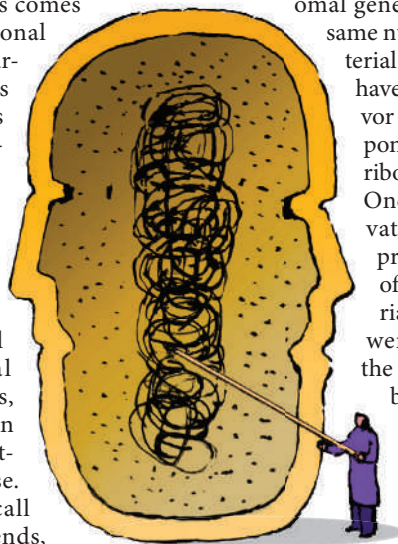
J. KAPUSTA/IMAGES.COM

CONNECTIONS

NEWS & VIEWS

PLANT ECOLOGY

Resourceful invaders

Tim Seastedt

Plant species that colonize new environments tend to favour habitats with ample water and nutrients. But invasive plants can be more efficient in their use of resources than that observation might imply.

There are few places on Earth that are not susceptible to invasions by plants intentionally or accidentally introduced into areas outside their native ranges. Such species can wreak ecological and economic havoc in their new habitat, hence the continuing search for long-term solutions to control them[1]. With the findings of Funk and Vitousek[2], however, described on page 1079 of this issue, that endeavour has just become more difficult.

Invasive plants must first colonize and then persist in their new environment. A small percentage of these may then become abundant and dominant components of the plant community. For example, an annual grass, cheatgrass (*Bromus tectorum*), was accidentally introduced from Asia into grasslands throughout the world, and now occupies many millions of hectares in the environments into which it was introduced. An ornamental plant called the cartwheel flower (*Heracleum mantegazzianium*) escaped its garden environments; it is now known as 'giant hogweed' and is a serious problem on three continents. The Monterey pine (*Pinus radiata*), originally from

a small area in the western United States, was planted for wood production worldwide. This tree is now considered a weed in many regions. These examples illustrate the diversity of growth forms that can characterize plant invaders.

Until now, most invasive species were thought to be largely opportunistic, exploiting changing environmental conditions and human or natural disturbances. A survey of available studies[3] showed that most native species equalled or outperformed invaders for those traits known to be beneficial to the survival and reproduction of the species. That same analysis suggested that low-resource environments are least susceptible to invasions. Low-resource environments are those that provide little in the way of light, water or nutrients such as nitrogen and phosphorus. A limitation in one or more resources has the potential to exclude plant species from occupying such habitats. Although exceptions to invasion of low-resource habitats are known[4], findings from studies on ecosystem restoration indicate that low-resource environments have the lowest abundances of unwanted species[5].

In their paper, however, Funk and Vitousek[2] show that invasive plants also have the ability to compete in low-resource environments. In a study that included a broad spectrum of different forms of plant growth, and different habitats, the authors show that, over short time periods, invasive species tend to be more efficient at capturing energy per unit of leaf mass or per unit of leaf nitrogen. Water-use efficiencies are similar between natives and invaders. Native plants tend to retain their leaves longer than invaders, so after longer time intervals the efficiencies for exploiting light, water and nitrogen, with the exception of one resource in one habitat, are also similar. Nonetheless, the demonstrated efficiencies of the invaders, accompanied by other traits such as slightly higher seed production or reduced mortality compared with the natives, would make these species superior competitors in low-resource environments.

Funk and Vitousek's conclusion that diverse plant groups — grasses, ferns, trees and garden ornamentals (Fig. 1) — demonstrate high resource-use efficiency is strong evidence that invasive plants can indeed fare well under

Figure 1 | Pretty, but not wanted. Hawaii is especially vulnerable to invasive species, which include these two ornamental plants — firethorn (*Pyracantha angustifolia*, left) and montbretia (*Crocosmia pottsii* × *aurea*). These species were included in Funk and Vitousek's study[2] demonstrating that invasive plants are competitive in low-resource environments.

low-resource conditions. The generalization that native species have a 'home field advantage' in such conditions is not valid. Reconciling these results with existing studies is possible, but the outcome broadens our view of what characterizes successful invaders.

First, plant invaders include the usual suspects — weeds capable of exploiting high-resource, disturbed conditions. These appear in gardens, along roadsides and, occasionally, in natural areas. But such species do not usually pose a long-term problem in natural areas because they cannot compete once resources such as soil nutrients become more limiting.

A second group of invaders can be viewed as those species that, in the process of being introduced into new habitats, have escaped enemies and/or formed fortuitous associations with other organisms. These associations include beneficial soil microbes that assist the invader in obtaining resources[6,7]. The loss of enemies and the gain of mutualist organisms can provide invaders with more resources than they would have obtained in their native lands[8]. Some of these plants can out-grow and out-reproduce native species in a wide range of habitats.

Funk and Vitousek[2] now show that there are also many species that are well adapted to low-resource conditions. This group might also benefit from being attacked by fewer pathogens and fewer herbivores, and from more beneficial associations with soil organisms. If this is the case, genetic constraints may — at least in the short term — limit the value of these benefits[9]. But the ability of these species to invade and persist in new, low-resource environments is at least undiminished, and may be enhanced.

Funk and Vitousek's take-home message is less than encouraging for conservation biologists and restoration ecologists. Attempts to exclude invasive species, by reducing the availability of resources for example, will not protect these habitats from non-native species with the traits shown in this study. Management activities that create or enhance low-resource habitats are unlikely to prove effective barriers to invaders, unless they are accompanied by the planting of large numbers of equally adapted, native species. ■

Tim Seastedt is in the Department of Ecology and Evolutionary Biology, and INSTAAR, University of Colorado, Boulder, Colorado 80309, USA.
e-mail: timothy.seastedt@colorado.edu

1. D'Antonio, C. M., Jackson, N. E., Horvitz, C. C. & Hedberg, R. Front. Ecol. Environ. 2, 513–521 (2005).
2. Funk, J. L. & Vitousek, P. M. Nature 446, 1079–1081 (2007).
3. Daehler, C. C. Annu. Rev. Ecol. Syst. 34, 183–211 (2003).
4. Reed, H. E., Seastedt, T. R. & Blair, J. M. Ecol. Appl. 15, 1560–1569 (2005).
5. Blumenthal, D. D., Jordan, N. R. & Russelle, M. P. Ecol. Appl. 13, 605–615 (2003).
6. Kilronomos, J. N. Nature 417, 67–70 (2002).
7. Mitchell, C. E. & Power, A. G. Nature 421, 625–627 (2003).
8. Mitchell, C. E. et al. Ecol. Lett. 9, 626–640 (2006).
9. Blumenthal, D. M. Ecol. Lett. 9, 887–895 (2006).

ASTROPHYSICS

The answer is blowing in the wind

Yousaf M. Butt

A source of astoundingly energetic γ-rays associated with a star cluster might provide a clue to a century-old question: where do the cosmic rays that constantly bombard Earth come from?

Massive stars have extreme lifestyles. They are born in clusters of up to several thousand members, blow fierce charged-particle winds during their short lives, and die — more or less together — in powerful supernova explosions. Now comes word from the High Energy Stereoscopic System (HESS) collaboration, to be published in *Astronomy and Astrophysics*[1], that γ-rays of very high energy have been spotted coming from the powerful young stellar association Westerlund 2 located in the southern sky[1] (Fig 1). This emission is of a higher energy than ever seen before from a group of stars, and pushes the limits of our understanding of the processes behind it.

Stars typically emit light around the visible part of the spectrum, where photons have an energy of a few electronvolts (eV). The γ-rays that HESS detected have energies in the range of tera-electronvolts (TeV), or 10^{12} eV. Previously, TeV γ-rays have been seen emanating from only a handful of exotic celestial objects. These include energetic pulsars (rapidly spinning and highly magnetized neutron stars just 30 or so kilometres across); the huge interstellar shock waves associated with the remnants of powerful supernovae; binary systems of a neutron star or a black hole coupled with a regular star; jets from distant 'active galaxies'; and the supermassive black hole thought to lurk at the centre of our Galaxy. So the HESS collaboration's discovery[1], dubbed HESS J1023–575, amounts to finding a completely new species of celestial γ-ray source. In fact, another TeV

source discovered recently[2] might also be a member of the same species. Designated TeV J2032+4130, this source is probably related to a subgroup of powerful stars in the Cygnus OB2 stellar association[3], but this identification is not quite as firm as in the case of Westerlund 2.

The most likely model for the origin of these highly energetic γ-rays is that multiple, supersonic winds of charged particles blowing from the dozens of massive stars (for our purposes, stars bigger than 8 solar masses) create violent plasma motions within Westerlund 2. This turbulence can accelerate particles to TeV energies (ref. 4 and references therein), and these particles can then interact with the ambient material and light to produce the detected γ-rays. This type of turbulent particle acceleration process is called the second-order Fermi mechanism, or Fermi-II acceleration for short. First-order Fermi (Fermi-I) acceleration is thought to be at work in the better-formed interstellar shock waves created by isolated supernova explosions.

Could such an isolated supernova remnant be behind the HESS J1023–575 detection? This possibility is rendered unlikely by the presence of a great deal of turbulence caused by the massive stars of the Westerlund 2 association. The evolution of a supernova remnant would be greatly perturbed in such an environment, and it could hardly be considered as 'isolated'.

On the other hand, the possibility that one supernova remnant or more could have added to the turbulence created in Westerlund 2 by the massive stars resident there, and thus provided

Figure 1 | Turbulent association. A composite infrared image from NASA's orbiting Spitzer observatory shows the Westerlund 2 stellar association. According to observations from the HESS telescope[1], it seems that turbulent processes at work in the fierce winds of the massive stars in Westerlund 2 create γ-ray photons 10^{12} times more energetic than visible light. The same processes might cause the acceleration of the cosmic rays that constantly hit Earth.

NASA/JPL–CALTECH/E. CHURCHILL (UNIV. WISCONSIN)

a further power source for the Fermi-II process most probably operating there, certainly can't be discounted. Although Westerlund 2 is a young stellar association (it is roughly 2 million years old), massive stars 'live fast and die young': they burn quickly and brightly. A supernova might well, therefore, already have occurred in Westerlund 2 — which might also explain a deficit discovered[5] in counts of the most massive members of Westerlund 2. There is no tell-tale sign of any supernova remnant in the area, but we should not necessarily expect there to be[6]: Westerlund 2's medium is so hot and rarefied that any remnant would leave a barely detectable signature.

This brings us to the most important aspect of the HESS findings[1]: their bearing on the origin of cosmic rays. Cosmic rays are very energetic particles — mostly protons, but also heavier nuclei and electrons — that continually rain down on Earth's atmosphere from space. No one knows for sure where in our Galaxy, or beyond, they are accelerated.

It is widely believed that supernova remnants are the main energy source powering the acceleration of cosmic rays. But the acceleration mechanism itself can be quite different, depending on whether the remnant is isolated (where Fermi-I acceleration dominates) or is part of an interacting system of several remnants embedded in a turbulent medium such as a stellar association (mainly Fermi-II acceleration)[4,7]. Because the massive stars that lead to supernovae are born together in stellar clusters, roughly 80% of supernova explosions are expected to take place near others within a relatively short period of time[8,9] — much like a closing fireworks display — creating a large 'superbubble' filled with hot, tenuous, turbulent plasma[10–13].

Cosmic rays, it has been argued[4,7–9], are much more likely to be accelerated in such superbubbles predominantly via the Fermi-II process, rather than — or, possibly, in addition to — in isolated supernova remnants where the Fermi-I mechanism holds sway. The HESS findings[1], which seem to confirm the viability of turbulent, Fermi-II acceleration, whether by massive stars alone or by supernova remnants in concert with massive stars, will be welcome news to the proponents of the superbubble mechanism of cosmic-ray acceleration. But there is considerable work still ahead.

At the theoretical end, it would be useful to develop the details of superbubble models[4], such that concrete predictions of emissions from realistic sources, from radio to TeV energies, can be made. Observationally, we must now aim to go beyond examining only the isolated supernova remnants as possible cosmic-ray acceleration sites. Even if the intriguing TeV-emitting isolated supernova remnants that are already known can be shown to be accelerating cosmic rays, it would not solve the general cosmic-ray acceleration problem for two reasons. First, the well-known and catalogued isolated supernova remnants are a

distinct minority of the supernova remnants in the Galaxy. Second, the precise mechanisms of acceleration at work in the isolated supernova remnants and superbubbles are probably different.

Luckily, a suite of sensitive ground- and space-based γ-ray observatories with acronyms such as VERITAS, MAGIC, GLAST and AGILE is now coming online. Together with HESS, these hold the promise of a rich harvest of celestial γ-ray sources. Detecting the large, diffuse and possibly overlapping superbubbles will be very challenging and time-consuming, and perhaps downright impossible with some current instrumentation. But such objects ought to be pursued with these and other lower-frequency observatories if we are to understand how and where cosmic rays are accelerated. ■

Yousaf M. Butt is at the National Centre for Physics, Quaid-i-Azam University, Islamabad,

Pakistan and at the Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA.
e-mail: ybutt@cfa.harvard.edu

1. Aharonian, F. et al. Astron. Astrophys. in the press; preprint available at www.arxiv.org/astro-ph/0703427 (2007).
2. Aharonian, F. et al. Astron. Astrophys. **431,** 197–202 (2005).
3. Butt, Y. et al. Astrophys. J. **597,** 494–512 (2003).
4. Bykov, A. Space Sci. Rev. **99,** 317–326 (2001).
5. Ascenso, J. Astron. Astrophys. **466,** 137–149 (2007).
6. Tang, S. & Wang, Q. D. Astrophys. J. **628,** 205–209 (2005).
7. Cesarsky, C. & Montmerle, T. Space Sci. Rev. **36,** 173–193 (1983).
8. Parizot, E. et al. Astron. Astrophys. **424,** 747–760 (2004).
9. Higdon, J. C. & Ligenfelter, R. E. Astrophys. J. **628,** 738–749 (2005).
10. Bruweihler, F. C., Gull, T. R., Kafatos, M. & Sofia, S. Astrophys. J. Lett. **238,** 27–30 (1980).
11. Heiles, C. Nature **441,** 1056–1057 (2006).
12. Ehlerová, S. & Palouš, J. Astron. Soc. Pacif. Conf. Ser. **331,** 67–74 (2005).
13. McClure-Griffiths, N. M. Astron. Soc. Pacif. Conf. Ser. **352,** 95–104 (2006).

NEUROSCIENCE

The brain's garbage men

Helmut Kettenmann

Microglial cells, the immune elements of the brain, are activated in disease or following injury. New findings indicate how these cells are switched on to remove damaged cells and cellular debris.

Brain function is generally considered as the activity of the neuronal network. So brain dysfunction or damage is thought to be caused by a disturbance to this network through the loss or malfunction of neurons. But the brain also contains another population of cells called glia, which in humans outnumber the neurons. In the disease context, a subtype of glial cells — the microglia — has attracted attention as sensors of any brain-damaging event. In response to injury, for example, microglia are activated, resulting in their interaction with immune cells, active migration to the site of injury, release of pro-inflammatory substances, and the engulfment of damaged cells and cellular parts by a process known as phagocytosis. On page 1091 of this issue, Koizumi et al.[1] report that microglia express a particular kind of cell-membrane receptor — a subtype of the purinergic family — which mediates their phagocytic activity*.

Generally, members of the purinergic family of receptors are activated by the nucleotide ATP, which not only serves as an intracellular energy substrate but is also an extracellular signalling molecule. Although neurons and astrocytes — another subtype of glia — normally release ATP, under disease conditions any damaged cell can release this molecule.

So far, at least 15 purinergic receptors have

been identified, and these are separated into two groups — the P2X ion-channel receptors and the P2Y metabotropic (non-ion-channel) seven-transmembrane-domain receptors. Microglia express several receptors of each group, which control their various functions under both resting and activated states. The $P2Y_{12}$ receptor has been identified as an essential control element for the movement of microglial projections, as indicated by imaging studies in live animals[2,3]. This receptor is downregulated when microglia are activated following injury to the brain.

Now, Koizumi and colleagues[1] report that the opposite is true for $P2Y_6$ receptors — that is, following injury, these receptors are upregulated and their activation triggers phagocytosis (Fig. 1). The authors found that, on stimulation with the $P2Y_6$-specific agonist (the UDP nucleotide), microglia grown in cell culture engulf fluorescently tagged particles. They then studied cell death caused by the injection of kainic acid into mouse brains, which produces damage to the hippocampal region. They observed an increase in the expression of the $P2Y_6$ receptor in the activated microglia.

Koizumi et al. also found that, in response to the kainic-acid-induced injury, natural levels of UTP increased. They therefore speculate that UTP and its degradation product UDP might even be the natural ligands for $P2Y_6$ receptors. They propose that both ATP and UTP are released by damaged cells, but that each mediates a distinct function of microglia:

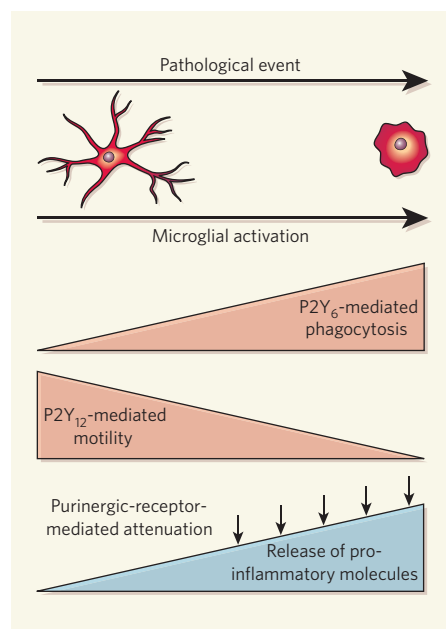*This article and the paper concerned[1] were published online on 4 April 2007.

Figure 1 | **Microglial activation.** Microglia are activated in response to pathological events such as disease or injury, and change from a ramified to an amoeboid morphology. This enables them to migrate to the site of injury. In the course of microglial activation, $P2Y_{12}$ receptors, which mediate the motility of microglial projections, are downregulated[3]. Koizumi *et al.*[1] now report that, by contrast, $P2Y_6$ receptors are upregulated and trigger phagocytosis.

ATP regulates cell motility through the $P2Y_{12}$ receptor — and potentially later on during the activation process via other purinergic receptors — whereas UTP triggers phagocytosis through $P2Y_6$ receptors. In other words, ATP seems to send a 'find-me' signal from a damaged cell, whereas UTP relays an 'eat-me' message, thereby alerting and activating the microglia, respectively.

The findings of Koizumi *et al.*[1] indicate that nucleotides and purinergic receptors constitute a core series of switches that control microglial function. Purinergic receptors also regulate other aspects of microglial activity, for example the release of the pro-inflammatory molecules, such as tumour-necrosis factor α and interleukin-6 (ref. 4; Fig. 1). However, these receptors are not the only elements that control microglial activity, as other functions of these cells, such as migration, are controlled by several other elements. Following brain injury for instance, migration of microglia into the degenerating cell layer is mediated by a protein called CXCR3 (ref. 5). In CXCR3-deficient animals, not only is microglial invasion impaired but so also is the removal of the neuronal processes that have been deprived of electrical input and thus are non-functional. Therefore, it seems that microglia can remove cellular parts, as well as whole cells[1,5].

The identification[1] of the $P2Y_6$ receptor as a mediator of microglial phagocytosis could aid the development of therapeutic agents to interfere with microglial activity in brain

diseases. However, whether microglial activation is a beneficial event in the context of neuronal degeneration and regeneration is still debated. For example, it may be necessary to remove electrical-input-deprived neuronal projections for new ones to sprout and make fresh connections. Similarly, irreversibly damaged cells might need to be removed to make space for newly generated neurons. On the other hand, partially preventing microglial activation represses the development of a condition called experimental autoimmune encephalomyelitis in mice, indicating that microglia may augment brain damage under some circumstances[6]. Nonetheless, one issue

is clear: the activation, and particularly the phagocytic activity, of microglia must be under stringent control because the brain has to be protected from an unwanted response by these cells. ■

Helmut Kettenmann is at the Max Delbrück Center for Molecular Medicine, Robert-Roessle-Strasse 10, 13092 Berlin, Germany.
e-mail: kettenmann@mdc-berlin.de

1. Koizumi, S. *et al. Nature* **446,** 1091–1095 (2007).
2. Honda, S. *et al. J. Neurosci.* **21,** 1975–1982 (2001).
3. Haynes, S. E. *et al. Nature Neurosci.* **12,** 1512–1519 (2006).
4. Färber, K. & Kettenmann, H. *Pflugers Arch.* **452,** 615–621 (2006).
5. Rappert, A. *et al. J. Neurosci.* **24,** 8500–8509 (2004).
6. Heppner, F. L. *et al. Nature Med.* **11,** 146–152 (2005).

BIOGEOCHEMISTRY

Iron findings

Philip W. Boyd

A huge phytoplankton bloom in the Southern Ocean yields estimates of how a continuous supply of iron affects oceanic carbon sequestration. But iron is not the only factor — nutrient supply is crucial too.

The ocean is a daunting place to study, where investigations must contend with a wide range of scales — from intracellular to ocean basins, from nanoseconds to seasons[1]. The difficulties are evident in the variety of approaches used to study the ecological productivity of its microscopic algae, or phytoplankton. Small-scale perturbation experiments, for example incubating seawater samples in small bottles with added elements such as iron, provide useful information on phytoplankton physiology and other intrinsic processes. They cannot, however, represent processes occurring across entire ecosystems, as the sampled phytoplankton are enclosed and isolated. In contrast, 'mesoscale' *in situ* studies, which perturb patches of ocean on the scale of hundreds of square kilometres, can address ecosystem-scale questions[2]. But even month-long experiments at this scale have flaws, such as pronounced mixing of the enriched patch with surrounding waters, altering its properties.

Elsewhere in this issue, Blain *et al.* (page 1070)[3] overcome these scaling issues by investigating a naturally occurring phytoplankton bloom covering an area of 45,000 km². Such large blooms photosynthetically convert so much carbon into an organic form that they have a marked effect on the atmospheric carbon dioxide concentration, and hence the global climate: a significant proportion of the carbon thus 'fixed', known as particulate organic carbon, is sequestered in the ocean depths. Iron is now recognized to be of equal importance to nutrients such as nitrate[4] in stimulating the development of these blooms. A larger supply of iron to the surface ocean — from dust deposition, for instance, as recorded from the

geological past[5] — can increase phytoplankton productivity and thus carbon sequestration and CO_2 drawdown[6].

Blain and colleagues[3] used satellite images to pinpoint an annually recurring bloom near Kerguelen, an island archipelago in the Southern Ocean south of, and at a longitude about equidistant from, South Africa and Australia. This 'natural laboratory' was sustained for months through constant iron and nutrient enrichment from below (Fig. 1, overleaf). The larger spatial and temporal scales of the bloom permitted Blain and colleagues to address questions inconclusively covered by previous mesoscale iron-enrichment studies in polar regions (see ref. 2 for a review). Chief among these was how much of the carbon fixed by the bloom was sequestered into the ocean depths.

The determinant of whether iron enrichment can alter global climate is the magnitude of carbon sequestration per unit of iron added. During their 30-day study, Blain *et al.* report 10 to 100 times more carbon export per unit of iron supplied than was estimated during the previous studies. These higher ratios are particularly significant, because they indicate that the higher iron supply evident during glaciation maxima[5] had a greater impact on atmospheric CO_2 drawdown than has generally been assumed. The contribution of iron enrichment to the total glacial–interglacial shift of 80 parts per million (p.p.m.) in atmospheric CO_2 might therefore approach the upper bound of 24 p.p.m. cited recently[6].

What are the reasons for this discrepancy[2,3]? Blain *et al.*[3] provide two explanations. First, the polar mesoscale iron-enrichment measurements[2] underestimated carbon export, because

50 & 100 YEARS AGO

50 YEARS AGO

In 1920 a German, Freiherr von Pohl, propounded the theory that certain rays emanate from the Earth which are injurious to the health of man... He followed this with the publication of a book... in which the use of the divining rod is recommended for detecting the presence of such rays. Credence was given to his theory by an astonishing number of people, particularly in Germany, Switzerland and the Netherlands... Spread of the belief, however, gave scope for the activities of large numbers of diviners, professional and otherwise. They appeared mostly to act in good faith, although one feels that the lady who claimed for her divining rod the ability not only to detect the presence of Earth rays, but also to extract them from the ground and deposit them in a ditch or rubbish heap... was making rather a good thing out of her neighbours' troubles... In the Netherlands, agriculture was at first unaffected by von Pohl's theory, until about ten years ago intensive propaganda on the part of the diviners proclaimed the efficiency of the protective boxes for controlling animal and plant disease. The propaganda was so successful... that uneasiness was felt in scientific circles.
From *Nature* 27 April 1957.

100 YEARS AGO

In addition to the usual record of measurements taken during the year 1906 at Epsom College, the report of the college Natural History Society for last year gives the average height, weight, and chest girth of all boys who have been measured in the ten years 1897–1906... On the whole, the average Epsom College boy would appear to be rather superior in physique than inferior to the average public-school boy. One marked exception is evident in the curves for the ages 17 yr. 10 mo. to 18 yr. 4 mo., though... the number of observations on which the curves are based is, for these months, much smaller than the rest.
From *Nature* 25 April 1907.
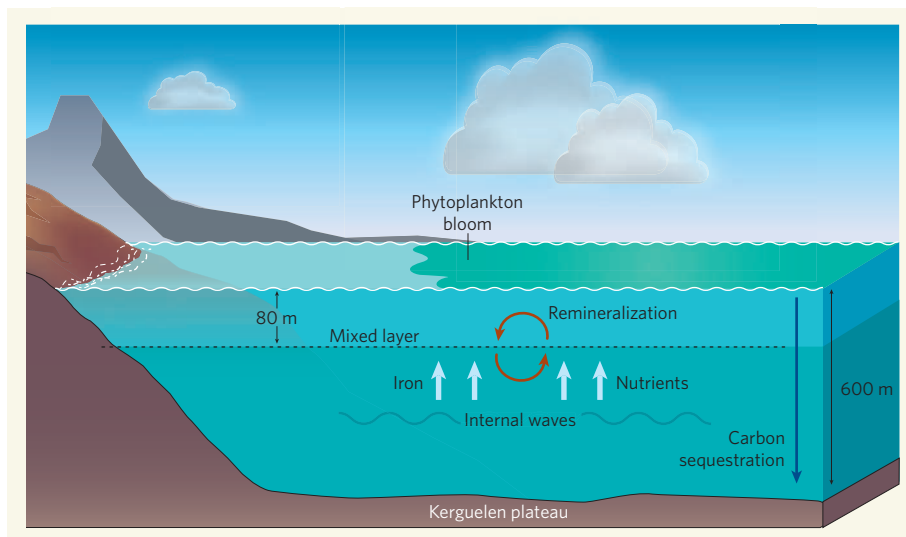
Figure 1 | Kerguelen blooming. The Kerguelen plateau is about 600 m under the ocean surface; in this region, internal waves enhance the vertical mixing of the deep waters above the plateau, which have higher iron and nutrient concentrations, with those in the 80-m-thick surface mixed layer. Up to half of the particulate iron and other nutrient elements were broken down (remineralized) to dissolved forms in the upper ocean. Both of these processes supplied continual nourishment to the phytoplankton studied by Blain *et al.*[3], which — through photosynthesis and the subsequent sinking of organic carbon into the deep ocean over several months — contribute to higher than previously reported[2] sequestration of atmospheric CO_2 per unit iron supplied.

they were too short-term (lasting just weeks) to observe its full extent. Second, they overestimated iron supply: pulses of extra iron into the surface ocean are prone to rapid removal, for example by sticking to sinking particles.

The ratios of carbon export to iron supply estimated by Blain *et al.* for particles sinking from the Kerguelen bloom are very similar to carbon–iron ratios in phytoplankton in high-iron laboratory cultures[7], pointing to little biological modification of these ratios between photosynthesis and subsequent sequestration. The similarity of the ratios is difficult to reconcile with recent reports that the organic carbon on sinking particles is broken down into dissolved forms, or remineralized, more rapidly than is iron[8]. The authors also report high stocks of zooplankton grazing on the Kerguelen blooms that would be absent from the laboratory cultures. The presence of zooplankton aids the remineralization of both iron and carbon, and thus reduces carbon export while resupplying iron to the phytoplankton.

The phytoplankton bloom at Kerguelen, fuelled by a sustained supply of iron and nutrients, was of exceptional duration, lasting some months. Although it used up virtually all of the iron and silicic acid in surface waters, it did not deplete its nitrate stock. Under high-iron conditions, a bloom should use equal amounts of nitrate and silicic acid[9]. The implication is that, despite the continuous vertical supply of nutrients characteristic of the Kerguelen site, the growth rate of the resident bloom is probably suboptimal owing to insufficient iron.

Together with other measurements[2], Blain and colleagues' results provide a powerful tool for modellers investigating the effects of the mode of iron supply on ocean biogeochemistry.

The main modes, in the geological past, have been episodic iron enrichment of the uppermost ocean through dust deposition and/or sustained enrichment of overlying waters through the upwelling of deep waters. The sustained iron and nutrient supply through internal wave activity at Kerguelen means that the intensity — and, more importantly, the ratio of its iron and nutrient supplies — may differ from those of polar upwellings. A quantification of the effects of upwelling and internal waves on this ratio is needed to determine whether the Kerguelen data are a proxy for the polar ocean during the glacial maxima.

But does Blain and colleagues' evidence[3] of more carbon export per unit iron supply mean that iron enrichment is a viable short-term climate-mitigation strategy? The authors say no: the enhanced export resulted from bloom longevity that was driven not just by sustained iron enrichment, but also by continuous nutrient enrichment. Moreover, the ratio of carbon export to iron supply is notoriously difficult to measure, and only a fifth of the phytoplankton's requirements were accounted for in the study's iron budget.

Nevertheless, the work is a novel and valuable addition to the library of phytoplankton-biogeochemistry studies. A final testament to the challenges of marine research, and the technical difficulties in assessing the efficacy of iron enrichment as a climate-mitigation strategy, is given by the story of Blain and colleagues' sediment traps, particle interceptors used to measure carbon sequestration at great depth. These could not in the first instance be recovered, but have just finally been salvaged — one year on.

Philip W. Boyd is at the NIWA Centre for

Chemical and Physical Oceanography,
Department of Chemistry, University of Otago,
Dunedin, New Zealand.
e-mail: pboyd@alkali.otago.ac.nz

1. Margalef, R. *Our Biosphere* (Ecol. Inst., Oldendorf, 1997).
2. Boyd, P. W. *et al. Science* **315,** 612–617 (2007).
3. Blain, S. *et al. Nature* **446,** 1070–1074 (2007).
4. Morel, F. M. M. & Price, N. M. *Science* **300,** 944–947 (2003).
5. Martin, J. H. *Paleoceanography* **5,** 1–10 (1990).
6. Sigman, D. M. & Boyle, E. A. *Nature* **407,** 859–869 (2000).
7. Sunda, W. G., Swift, D. G. & Huntsman, S. A. *Nature* **351,** 55–57 (1991).
8. Frew, R. D. *et al. Glob. Biogeochem. Cycles* **20,** doi:10.1029/2005GB002558 (2006).
9. Hutchins, D. A. & Bruland, K. W. *Nature* **393,** 561–564 (1998).

ATMOSPHERIC CHEMISTRY

A forest air of chirality

Euripides G. Stephanou

A sophisticated survey of certain volatile organic compounds in the air over forest ecosystems shows how such work can reveal varied emission patterns of different chiral, or mirror-image, forms of these compounds.

An unseen component of forest ecology is the emission, by both plants and animals, of a cocktail of volatile organic compounds into the atmosphere. Emissions from plants exceed those from animals by several orders of magnitude, and have various purposes, such as repelling herbivorous animals or attracting specific insects for pollination. A little-understood aspect of the chemistry of these compounds is the possibly different emission patterns of different chiral (mirror-image) forms, known as enantiomers, of the same compound. With a Lear jet as one of their platforms, Williams *et al.*[1] have undertaken an ecosystem-scale sampling of air over tropical forest. As they report in *Atmospheric Chemistry and Physics*, they find distinct differences between the enantiomeric composition of a common volatile compound emitted by tropical forest compared with boreal forest.

Plant emission of volatile organic compounds occurs on a vast scale, estimated at 1,150 Tg carbon (1 Tg = 10^{12} g) annually[2]. Many of these compounds are built up from isoprene (C_5) subunits and are traditionally given the collective name terpenes[3]. Together, isoprene and its monoterpene derivatives ($C_{10}H_{16}$) make up 55% of the estimated emissions[2]. Many monoterpenes are produced in two enantiomeric forms: α-pinene, a common example, occurs as both (+)-α-pinene and (−)-α-pinene.

The light-induced oxidation of natural and man-made volatile organic compounds affects the chemical and physical properties of the atmosphere[4,5]. But although the importance of different enantiomers for the biological activity of a compound is well known, the implications for atmospheric chemistry have not been established[6]. Indeed, the global effect of compounds such as monoterpenes has been estimated by measurements and models that consider the enantiomeric forms as one[2]. The work of Williams *et al.*[1] is an attempt to use a detailed enantiomeric characterization of monoterpene patterns in the atmosphere as a tool to enable a better understanding of the mechanisms underlying their emission from plants.

This is where the Lear jet came in. Williams *et al.* took air samples from an aircraft making flights of 3 hours, with a range of 1,800 km, over the tropical forest of the northeast coast of South America, and subsequently analysed the monoterpene content. They also sampled ambient air at the level of the forest canopy at a site in Suriname, within the region surveyed. The other sites of investigation were boreal pine forest in Finland, samples being taken at a height of 8 m from the ground, and the greenhouse of the botanical garden of the Johannes Gutenberg University in Mainz, Germany, where 8,500 plant species from all over the world are cultivated.

Most notably, monoterpenes emitted from the tropical forest showed an excess of the (−) enantiomer of α-pinene, and also of the (+) enantiomer of β-pinene. The monoterpene emissions from a mixture of three tropical species in the greenhouse of the botanical garden likewise showed an excess of (−)-α-pinene. Laboratory studies[6] have shown that both enantiomers of α-pinene react at the same rate with atmospheric oxidants (hydroxyl radicals and ozone). So the dominance of one enantiomer over the other does not imply preferential atmospheric removal of one of them. Rather, the explanation probably lies in the factors influencing emissions by plants. In that respect, studies of leaf chemistry[2] have shown that the strength of isoprene emission is a function of both light and temperature. In contrast, although monoterpene emission rates increase with temperature, plant species show differential sensitivity to temperature; different monoterpenes can also exhibit differing temperature dependence.

To interpret their observations, Williams *et al.* compared the concentrations of isoprene, (−)-α-pinene and (+)-α-pinene measured over tropical forest. They observed a surprising correlation between (−)-α-pinene and isoprene, but no correlation between its mirror image (+)-α-pinene and isoprene. The converse applied to emissions from the boreal forest: there was a clear predominance of (+)-α-pinene and (−)-β-pinene. A previous study[7] conducted in the same boreal forest showed that rates of emission of non-enantiomerically resolved α-pinene were temperature dependent. On the basis of both their own results[1] and this earlier work[7], Williams and colleagues propose that over the tropical forest, emission of (−)-α-pinene is — like that of isoprene — light-dependent, whereas that of (+)-α-pinene over the boreal forest is temperature-dependent.

Volatile organic compounds have a fundamental role in the coexistence of the flora and fauna in ecosystems. But there is still much to learn about the relationships and interactions between species that can be related to an effect of naturally produced compounds such as monoterpenes. Studies[8] of the biosynthesis and the enantiomeric composition of monoterpenes in forest trees has shown that different enzymes (terpene cyclases) are responsible for the production of the two enantiomers. This supports the observation that enantiomeric composition varies in different tissues and in different individual trees[8]. Research aimed at correlating the enantiomeric composition of monoterpenes from different parts of specific trees with the corresponding enantiomeric composition of emissions by the same trees under specific environmental conditions — light, temperature and humidity — will be necessary for a fuller understanding of the mechanisms of the emission of volatile organic compounds into the atmosphere.

Microbial degradation and plant metabolism might be selective towards a specific enantiomer; but processes in the atmosphere such as photo-oxidation are not. So chiral analysis will be a powerful additional tool in atmospheric chemistry for investigating the sources of emission and the fates of chemical compounds in various environments[9]. The main message of Williams and colleagues' work is that ecosystem-scale field studies involving sophisticated sampling and analytical techniques are required to make sense of the biochemistry and ecology of these natural emissions, and of other physical and chemical observations in the atmosphere. ■

Euripides G. Stephanou is in the Environmental Chemical Processes Laboratory, School of Sciences and Engineering, University of Crete, 71003 Voutes-Heraklion, Greece.
e-mail: stephanou@chemistry.uoc.gr

1. Williams, J. *et al. Atmos. Chem. Phys.* **7,** 973–980 (2007).
2. Guenther, A. *et al. J. Geophys. Res.* **100,** 8873–8892 (1995).
3. Ruzicka, L. *Proc. Chem. Soc.* 341–360 (1959).
4. Andreae, M. O. & Crutzen, P. J. *Science* **276,** 1052–1058 (1997).
5. Kavouras, I. G. *et al. Nature* **395,** 683–686 (1998).
6. Nunes, F. M. N. *et al. Atmos. Environ.* **39,** 7715–7730 (2005).
7. Tarvainen, V. *et al. Atmos. Chem. Phys.* **5,** 989–998 (2005).
8. Norin, T. *Pure Appl. Chem.* **68,** 2043–2049 (1996).
9. Harner, T. *et al. Environ. Sci. Technol.* **34,** 218–220 (2000).

COMPUTING

The security of knowing nothing

Bernard Chazelle

'Zero-knowledge' proofs are all about knowing more, while knowing nothing. When married to cryptographic techniques, they are one avenue being explored towards improving the security of online transactions.

Modern scientists, not unlike medieval monks, keep their knowledge firmly grounded in trust and authority. Unless it is part of their job description to probe such matters, they take it on faith that Genghis Khan defeated the Khwarezmid Empire, that praying mantids moonlight as sexual cannibals, and that man landed on the Moon. There is only so much a person can check: trust is the oxygen of the scientific community.

Unless, of course, it has to do with online shopping. Computer transactions tend to bring out the paranoid in all of us. By all accounts, that is a healthy reaction. We might be putting our faith in online auctions, e-voting, computer authentication and privacy-preserving data mining, but — with very different aims in mind — so are the bad guys. More than a decade ago, Oded Goldreich, Silvio Micali and Avi Wigderson showed how to make virtually any cryptographic task secure[1]. But unfortunately, their

otherwise remarkable scheme breaks down in 'concurrent' settings, which is another way of saying that it fails where it really shouldn't — namely, on the Internet. Work by Boaz Barak and Amit Sahai in recent years, however, offers a way out of this bind[2].

The secret to secure online transactions is the mastery of 'zero knowledge': the art of proving something without giving anything else away. Can I convince you that I am the better chess player without ever playing a game, that I am younger than you without divulging my age, or that I can prove a hard theorem without revealing my hand about the proof? Can a referendum take place on the Internet that leaks no information about voters' preferences? The concept of zero knowledge[3], introduced in the mid-1980s, helps us to formalize these questions.

To illustrate the principle, let us say Petra (the prover) and Virgil (the verifier) are shown

the subway map of a large metropolitan area (Fig. 1). Blessed with superior mental powers, Petra claims to see right away that it is possible to visit every stop exactly once without leaving the subway system, thus forming what is called a hamiltonian path. Poor Virgil sees nothing of the sort — the reason being his inability to solve conundrums like the one at hand, known as NP-complete problems.

Such problems have solutions that can be verified in a number of steps proportional to a polynomial in the size of the input data. Whether all NP-complete problems can actually also be solved within that same time is an open question, arguably one of the most pressing in all science. The answer is believed to be no: this is why Virgil badly needs Petra's help if he is to be convinced of her claim.

A zero-knowledge proof takes the form of a question-and-answer session between Petra and Virgil that will leave Virgil convinced of the
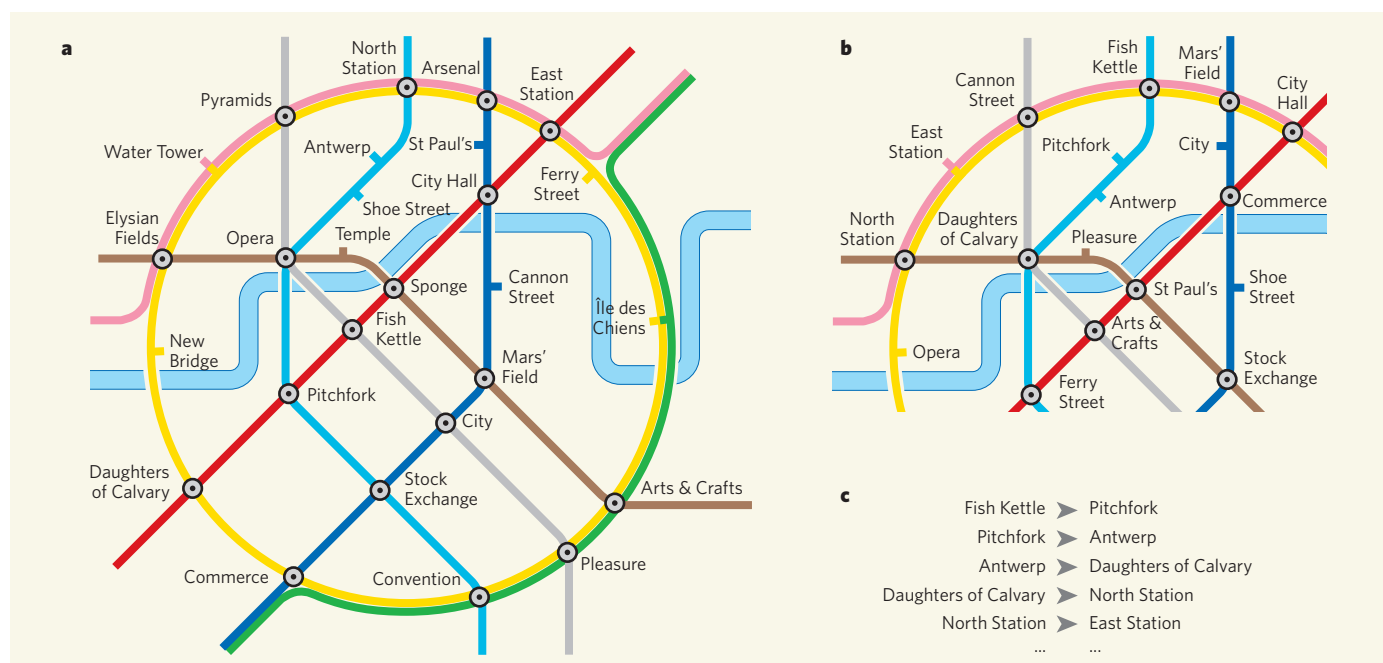
Figure 1 | Tube tour. After decades of steady investment, Orbiville's underground-railway system (**a**) is finally blessed with its own hamiltonian path, allowing one to visit every station exactly once without leaving the system. Or so Petra believes. To convince Virgil of that development, she renames the stations on the Orbiville metro map at random (**b**): North Station becomes Fish Kettle, Antwerp becomes Pitchfork, Shoe Street becomes Antwerp, and so on. She then hides the renamed map and the permutation table in a safe. Next, Virgil tosses a coin: heads, Petra opens the locked box and Virgil checks that the map was properly renamed, with each station appearing exactly once; tails, Petra reveals only the pairs of names on the relabelled map (**c**) that form the hamiltonian path. She also grants Virgil restricted access to the safe by revealing to him only the path on the map formed by the pairs of names. In this way,

Virgil can check that the alleged path not only visits each station once, but also actually exists in the subway system — all the while being denied access to the rest of the map. If Petra can perform this test a few times without raising eyebrows (using freshly permuted names at each round), Virgil will leave utterly assured of the existence of a hamiltonian path, yet clueless about its layout. Why? If Petra did not have a valid path or cheated in the renaming, she would have at least a 50% chance of getting caught at each round: either the encryption or the path would have to be faulty. Heads could catch the first case, and tails the second one. Repeating the test boosts Virgil's confidence in Petra's claim beyond any reasonable doubt — or shatters it, as the case may be. At each round, Virgil learns the encrypted map or the alleged hamiltonian path with the renamed stations, but never both. This ensures zero knowledge.

existence of a hamiltonian path, but that will reveal nothing about it[4]. If Petra is mistaken or in a cheating mood, Virgil will spot a contradiction in her responses and reject her 'proof'. If Petra is correct, however, not only will Virgil believe her claim, but he will learn nothing else that he could not figure out on his own without Petra's help. Better still, no amount of deviousness in posing the questions on Virgil's part can alter that fact. Zero knowledge is a mechanism for enforcing honest behaviour on both sides: cheating would bring no benefits to Virgil and it would expose Petra to the embarrassment of getting caught. Little wonder that zero knowledge is often held up as the 'holy grail' of secure computing.

But do zero-knowledge proofs even exist? Under widely accepted assumptions, the answer is yes for any NP-complete problem[5]. Were Petra to hand over the hamiltonian path, Virgil would learn more than the mere truth of her claim: he would know the itinerary itself, something, we just argued, he could not find on his own. To achieve zero-knowledge status requires, as so often in life, dialogue, commitment, and a bit of luck. To keep Petra from deceiving him with inconsistent answers, Virgil will ask her to commit to her claim once and for all. This is the software equivalent of hiding the hamiltonian path in a locked box that neither party can access until both jointly decide to open it. In fact, the path must be relabelled randomly, so that merely looking at it will not help Virgil trace it in the original map. Why so

much suspicion? Because distrust is the very ailment that zero knowledge seeks to cure: if Virgil trusted Petra, after all, her good word alone would be sufficient and no proof would be needed.

The surprise is that, to enforce honesty among distrustful parties, randomness must be thrown into the mix. A nasty side effect is that the conversation might go awry and lead Virgil wrongly to conclude that a hamiltonian path exists when none is to be found. This should happen with a probability of below 50%. (Correct claims will never be rejected, however.) Isn't this being inordinately lax? No: by repeating the dialogue a few dozen times, one can easily reduce the error probability to one in a trillion.

For technical reasons, to carry on several such conversations at the same time, as might happen on the Internet, is a big no-no. The issue is subtle, but to see why concurrency facilitates deception is not. Ever care to 'beat' your local chess-club champ? Here is how you do it: arrange for Garry Kasparov to play a match with you while you play with the local champion simultaneously — and do it online so neither one can see what you're up to. The trick is to feed each player the other's moves: in all likelihood, you will lose to Kasparov, but win against your neighbourhood champion.

The beauty of Barak and Sahai's work[2] is that they are able to overcome the pitfalls of concurrency by deploying a sophisticated

arsenal of cryptographic techniques. One of them is to relax the definition of zero knowledge by enhancing Virgil's power to simulate any potential dialogue with Petra. Another is to squeeze long messages into shorter ones without losing their security properties. Although fairly complex, the Barak–Sahai technology takes us one step closer to true security on the Internet.

In 1962, US President John F. Kennedy dispatched Dean Acheson to Paris to offer Charles de Gaulle photographic evidence of Soviet missiles in Cuba. The French president declined to see it, saying: "The word of the president of the United States is good enough for me." Zero knowledge is so blissfully easy in a climate of trust. The challenge is to deal with liars and cheaters. The continuing work of those such as Barak and Sahai[2] is giving us new tools to do that on the Internet. ∎

Bernard Chazelle is in the Department of Computer Science, Princeton University, 35 Olden Street, Princeton, New Jersey 08540-5233, USA.
e-mail: chazelle@cs.princeton.edu

1. Goldreich, O., Micali, S. & Wigderson, A. Proc. 19th ACM Symp. Theor. Comput. 218–229 (1987).
2. Barak, B. & Sahai, A. Proc. 46th IEEE Symp. Fut. Comput. Sci. 543–552 (2005).
3. Goldwasser, S., Micali, S. & Rackoff, C. SIAM J. Comput. 18, 186–208 (1989).
4. Goldreich, O. Foundations of Cryptography (Cambridge Univ. Press, 2001).
5. Goldreich, O., Micali, S. & Wigderson, A. J. Ass. Comput. Machin. 38, 691–729 (1991).

CHEMICAL BIOLOGY

Dressed-up proteins

Gijsbert Grotenbreg and Hidde Ploegh

Proteins aren't just defined by their constituent amino acids — structural modifications can yield complex mixtures of protein forms. An approach that controls the addition of such modifications may help to define their role.

Polypeptides freshly minted from the cell's protein-assembly apparatus are by no means ready for active service. Both the peptide backbone and its side chains may need to be altered by post-translational modifications (PTMs), the covalent attachment of chemical groups that change the properties, and hence the function, of newly generated proteins. PTMs also control the degradation of aberrant proteins and those at the end of their lifespan. Such modifications dramatically expand the compositional and functional complexity of these molecules. Not satisfied with leaving everything to nature's whim, chemists are taking great strides in developing PTM-mimics. On page 1105 of this issue[1], van Kasteren et al. describe strategies that offer precise control over the attachment of carbohydrates and other PTM-mimics to polypeptides.

In nature, any given protein may exist as a mixture of forms, each incorporating different PTMs. Assessment of the contribution made by each PTM to that protein's function demands that the individual components of these mixtures should be isolated and identified. But separating the constituents from these complex — and possibly dynamic — mixtures is a dauntingly arduous task. The ability to uniformly modify proteins at specific sites in the molecule, using chemical approaches that mimic the original PTM, is therefore essential for progress in this area.

A few examples of this approach are already known. One clever and synthetically straightforward example was reported recently[2], in which the side chains of cysteine amino acids in proteins were selectively targeted to mimic a particular PTM — the modification of lysine

amino-acid side chains with methyl groups. The resulting molecules were used to assess the role of this PTM in DNA–protein complexes found in cell nuclei.

One intricate form of PTM that remains a prize target for chemists is the attachment of carbohydrates — glycans — either to nitrogen atoms (N-linked) in the side chains of asparagine amino acids, or to oxygen atoms (O-linked) on the side chains of serine or threonine. In living cells, glycosidase and glycosyl transferase enzymes first trim N-linked glycans, then extend them with sugars that can branch in several directions, generating numerous variations on a theme. Glycosyl transferases also act on O-linked glycans, imposing similar extensions and modifications. To add to the complexity, each sugar can bear different chemical groups, and the linkages between sugars have specific orientations. Many biological processes, such as cellular differentiation and development, cell adhesion, immune surveillance and inflammation, rely to varying degrees on the correct decoration of proteins with such glycans.

Van Kasteren et al.[1] have now responded to the clarion call for improved synthetic methods to modify proteins. Not only have they installed carbohydrate PTM-mimics at specific locations in a protein, but they have also incorporated another important motif — a sulphated
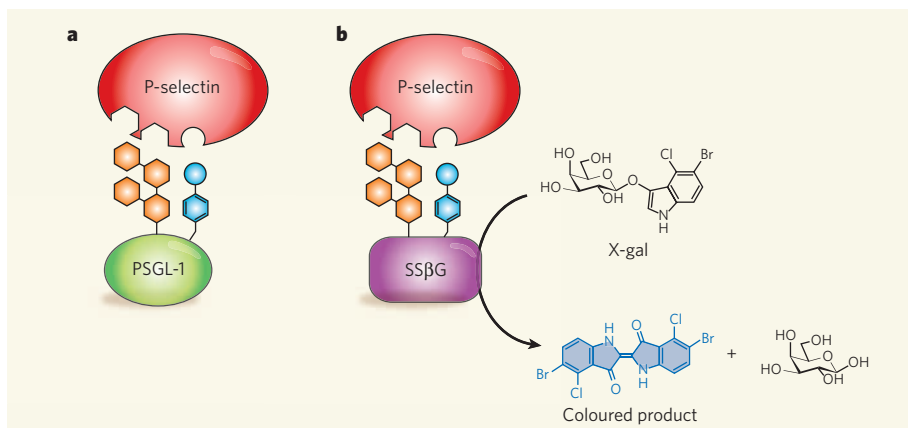
Figure 1 | Dressing up proteins. a, The protein P-selectin recognizes and binds to its ligand, PSGL-1, using covalently bound chemical groups (known as post-translational modifications, PTMs) at the ligand's surface. The PTMs are a carbohydrate (sialyl-Lewis-X, orange hexagons) and a sulphated amino acid (sulphotyrosine, blue hexagon with circle attached). **b**, Van Kasteren et al.[1] show that the unrelated SSβG protein can be fitted with mimics of these PTMs so that it is recognized by P-selectin. SSβG enzymatically converts X-gal molecules into a coloured product. This reporter function of the modified SSβG was exploited to visualize P-selectin in tissue samples.

tyrosine amino acid — at another defined site in the same molecule. To accomplish such a feat of directed chemistry, a creative synthetic strategy was required. The authors specifically appended their sulphotyrosine mimic to the side chains of cysteines using a selective 'disulphide exchange' reaction. To attach the carbohydrate PTM-mimic, van Kasteren et al.[1] incorporated chemical groups (alkynes and azides) into the glycan and the protein. These groups react exceptionally well with each other under near-physiological conditions[3,4], but are indifferent to all other groups present.

To target their PTM-mimics to precise locations, the authors used genetically engineered bacteria to create proteins either with cysteine or with azide-containing or alkyne-containing amino acids at the desired surface positions. Azide-containing and alkyne-containing amino acids do not occur in nature, so these were incorporated by tricking the cell's biosynthetic machinery into installing them at positions where a methionine amino acid would otherwise have occurred.

The uptake and incorporation of azide-containing building-blocks into biomolecules has previously proven its worth, as azide groups can be selectively labelled by chemical means[5,6]. Similarly, the tethering of carbohydrates — from simple glucose to the complex

carbohydrate known as sialyl-Lewis-X — to proteins is not unprecedented[7]. But what makes van Kasteren and colleagues' work unique is that two distinct PTMs have been installed on a single polypeptide. This successful application of two distinct chemical processes, involving reaction centres that recognize each other, but which ignore the dense concentration of chemical groups surrounding them, represents a formidable achievement.

As a target to which they could apply their technique, van Kasteren et al.[1] chose the protein P-selectin, whose presence in tissue is indicative of inflammation and is also a useful hallmark of cerebral malaria. P-selectin recognizes another protein, PSGL-1, by means of two distinct PTMs — a sulphated tyrosine residue and a sialyl-Lewis-X carbohydrate — found at separate locations on the PSGL-1 backbone (Fig. 1). The authors took an enzyme (*Sulfolobus solfataricus* β-glycosidase, SSβG) that normally doesn't bind to P-selectin, and decorated it with PTM-mimics to resemble PSGL-1. The presence of SSβG can be revealed when it converts a molecule known as X-gal to a blue product. Would SSβG, dressed up as PSGL-1, now allow the visualization of P-selectin? The answer was a resounding yes. The mere presence of the PTM-mimics positioned appropriately on the surface of SSβG allowed it to be recognized by P-selectin, as reported by the blue enzymatic reaction product. In a convincing demonstration of the technique's utility, PTM-bearing SSβG protein was used to visualize chronic inflammation and cerebral malaria pathology in tissue sections.

This work artfully brings together several chemical tools to obtain pure proteins

---

APPLIED PHYSICS

Weight inside

How can tiny particles such as molecules be weighed? A cantilever, if small and flexible enough, will bend under the weight of a molecule adsorbed onto its surface. A rather more sensitive mass measurement is given by the shift in the cantilever's resonance frequency as mass is adsorbed. This tiny resonance signal can be read out electronically by integrating the cantilever onto a silicon chip.

Previous demonstrations of chip-based mass sensors include the detection of a single virus weighing as little as 9.5 femtograms (1 femtogram is 10^{-15} grams) in air (A. Gupta *et al. Appl. Phys. Lett.* **84,** 1976–1978; 2004), and weighing a cluster of about 30 xenon atoms — equivalent to a mere 7×10^{-21} grams — in a vacuum (Y. T. Yang

et al. Nano Lett. **6,** 583–586; 2006). But where the particles to be weighed are suspended in solution (as is the case, for example, in many biological settings), the technique hits a snag: the viscosity of the fluid damps the resonator, and significantly decreases its sensitivity.

Elsewhere in this issue, Scott Manalis and colleagues present what might be described as a radical solution to the problem: putting the fluid inside the resonator (T. P. Burg *et al. Nature* **446,** 1066–1069; 2007). They have designed a vacuum-sealed silicon microcantilever with hollow channels (pictured), connected to pressure-controlled inlets and outlets for fluid delivery.

The device works in two modes. In the first, which is particularly suited to selective detection of biomolecules,

a solution is continuously run through the channels and particles can adsorb on the channels' inner surface — which must be specially prepared for the purpose. The authors demonstrate how the mass change can be followed by monitoring shifts in the resonance frequency of the cantilever in real time as proteins in solution become bound to appropriate receptor molecules that have been grafted onto the tube surface.

In a second mode, particles are detected in transit through the channels; this is useful for weighing particles in dilute solutions. An experiment in this mode determined the distribution in masses, with a resolution of one femtogram, for two types of live bacterial cell, of average masses 110 and 150 femtograms.

The work is an example of the steady progress that is being made in designing practical, inexpensive and portable lab-on-a-chip diagnostic devices. Although further advances are required to demonstrate a 'killer application' for Burg and colleagues' fluidic sensor — a medically relevant specific detection of viruses in blood samples, for example — it is already an elegant method for weighing tiny particles in solution.
Liesbeth Venema

carrying specific, complex-carbohydrate-based PTMs. It remains to be seen whether proteins other than conveniently sized reporters such as SSβG will tolerate such manipulation with similar ease. Nevertheless, by combining the genetic manipulation of proteins with chemical reactions to selectively install biologically active structures, van Kasteren et al.[1] have taken a crucial step towards unravelling the contributions of complex PTM patterns to protein function.   ■

Gijsbert Grotenbreg and Hidde Ploegh are at the Whitehead Institute for Biomedical Research

and the Department of Biology, Massachusetts Institute of Technology, 9 Cambridge Center, Cambridge, Massachusetts 02139, USA.
e-mail: ploegh@wi.mit.edu

1. Van Kasteren, S. I. *et al. Nature* **446,** 1105–1109 (2007).
2. Simon, M. D. *et al. Cell* **128,** 1003–1012 (2007).
3. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. *Angew. Chem. Int. Edn* **41,** 2596–2599 (2002).
4. Tornoe, C. W., Christensen, C. & Meldal, M. *J. Org. Chem.* **67,** 3057–3064 (2002).
5. Saxon, E. & Bertozzi, C. R. *Science* **287,** 2007–2010 (2000).
6. Kiick, K. L., Saxon, E., Tirrell, D. A. & Bertozzi, C. R. *Proc. Natl Acad. Sci. USA* **99,** 19–24 (2002).
7. Fazio, F. *et al. J. Am. Chem. Soc.* **124,** 14397–14402 (2002).

MATHEMATICAL PHYSICS
Added dimensions to grain growth

David Kinderlehrer

A long-standing mathematical model for the growth of grains in two dimensions has been generalized to three and higher dimensions. This will aid our practical understanding of certain crucial properties of materials.

Most engineered materials, including metals and ceramics, are polycrystalline — that is, they are composed of myriad, small, crystalline grains, separated by interfaces or boundaries (Fig. 1). Yet despite the ubiquity of such grains, we know surprisingly little about how they grow. On page 1053 of this issue, MacPherson and Srolovitz[1] supply a precise answer in three dimensions — indeed, in any number of dimensions. The result is not merely novel, but also intellectually exciting: and it is exciting both for materials scientists, who now know that this question has a succinct and elegant answer, and for mathematicians, who can witness recondite methods of integral geometry resolving a question of material significance.

The study of crystalline microstructure is ancient. Our ancestors fashioned stone implements from tool-stones that had special microstructures, and the excavation of such tool-stones has informed our ideas about human migration[2]. In the world of modern materials, cellular networks are implicated in many, varied material properties at many different length scales, including lifetime properties such as fracture toughness, and functional properties such as electrical conductivity. Preparing arrangements of grains and boundaries suitable for a given application is a central problem of materials science: the problem of microstructure.

Over time, and especially with exposure to elevated temperatures, grain boundary networks coarsen, and thus grains grow. Promoting or preventing this process is of paramount technological importance in many engineered materials. Simultaneously, a grain configuration must completely fill space and, in some way, minimize energy[3]. In other words, it

is subject to both topological and thermodynamic constraints.

The simplest coarsening mechanism is local: each point on a boundary wall moves perpendicularly, with a rate proportional to its surface tension. But to understand the evolution of a network, this microscopic description is unsatisfactory: one cannot inspect the surface tension at every point. Hence the central question:
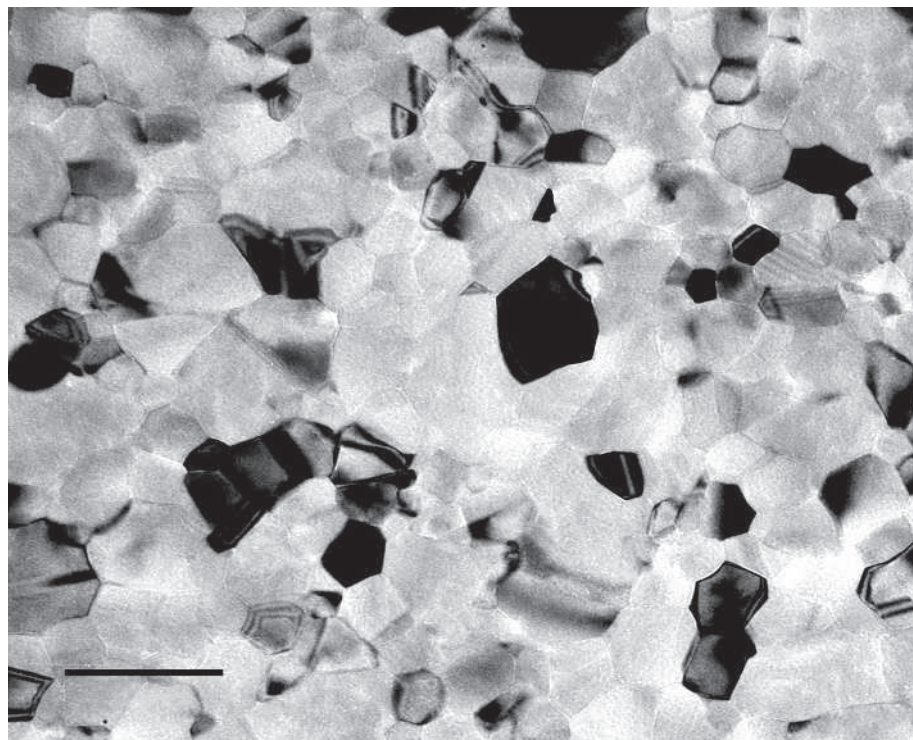

Figure 1 | With the grain. The way in which the area of a two-dimensional granular structure — here, a transmission electron micrograph of an aluminium thin-film structure — varies with time has been known for more than 50 years. Only now have MacPherson and Srolovitz[1] managed the jump to three and higher dimensions. (Scale bar, 200 nm; image courtesy of K. Barmak.)

how does the volume of a cell (which is easier to measure) change? In the two-dimensional case, the answer was given by John von Neumann[4] in 1952 and William Mullins[5] in 1956. Often just called the '$n-6$ rule', it states that the rate of change of area, dA/dt, of a two-dimensional cell with n sides is given by $dA/dt = \gamma(n-6)$, where γ is a material constant. Thus, cells with more than six sides grow (positive change in area with time) and those with fewer than six sides shrink (negative change). The familiar pattern of regular hexagonal tiles is an equilibrium configuration.

The equation is, in fact, a disguised version of the familiar geometric fact that the sum of the exterior angles of a polygon is 2π (360°). As an irregular cell grows, its sides are assumed to maintain equal surface tension and to meet at vertices of interior angle 120° — as indeed they do in the regular hexagon. It is interesting to note here that the dimensionality of the measurement implied in the equation (the number of sides, which has no dimension) is two less than what is being measured (a two-dimensional change in area). The ($n-6$) notation contains all the ingredients for measuring and counting cell features that extend to the three-dimensional structures.

In this respect, MacPherson and Srolovitz's result in three dimensions has a familiar face, as it were. It takes the form

$$dV/dt = \gamma(\sum_{i=1}^{n} e_i - 6L)$$

and thus relates the rate of change of (three-dimensional) volume, dV/dt, of a cell to two

one-dimensional quantities: the cell's linear size, L, and the sum of the lengths of its edges, e_i, which is also one-dimensional. Just as in the two-dimensional case, this equation accounts for growth that maintains equalized surface tension and constant angles at which grain facets meet. The rate of change of volume is, however, no longer a constant for a given cell, but varies as its size changes. The n-dimensional result maintains this property: the rate of change of n-dimensional volume is related to $n − 2$-dimensional features of the cell and no others — in particular, no lower-dimensional ones.

Why this is the case is a signature problem in integral geometry and geometrical probability that also has an 'ancient' history. It dates back at least to the Buffon needle problem formulated in 1777 — in essence, a method for using chance and geometry to compute π — and was followed up by efforts from many great names since[6]. One of these was the twentieth-century Swiss mathematician Hugo Hadwiger. In a celebrated theorem[7–9], he stated that there is a unique way to measure n-dimensional features — such as the linear size of a grain or a grain facet as

depicted in a two-dimensional image — in a space of equal or higher dimensionality, such as three-dimensional space, that satisfies the natural properties of additivity (roughly, that the size of the union of disjoint objects is the sum of their sizes) and homogeneity.

The Hadwiger measure was 'hiding in plain sight' to solve the problem in hand, but, in studying time-dependent evolution, MacPherson and Srolovitz have extended it and employed it in a dramatic new way. Their result will be a crucial factor in meeting the challenge of finding a complete theory of microstructural evolution that satisfies both topological and thermodynamic constraints, as is required[3].

MacPherson and Srolovitz's answer[1] to one central question in materials science raises a second equally fundamental question. A physical grain network is beset by inhomogeneities and anisotropy, to name just two impediments to ideal growth. Even in the abstract, the effect of such features is unknown: we enter the realm of stochastic analysis through simple combinatorial events such as cell or facet deletion[10]. Such analysis, implemented with automated data acquisition in the laboratory and

large-scale simulation at the desk, will be the future direction of the subject. The remarkable and profound collaboration of mathematicians and materials scientists exemplified by MacPherson and Srolovitz's advance looks set to continue. ■

David Kinderlehrer is in the Center for Nonlinear Analysis and the Department of Mathematical Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213-3890, USA.
e-mail: davidk@andrew.cmu.edu

1. MacPherson, R. D. & Srolovitz, D. J. Nature 446, 1053–1055 (2007).
2. Mellars, P. Science 313, 796–800 (2006).
3. Smith, C. S. in Metal Interfaces 65–108 (Am. Soc. Metals, Cleveland, 1952).
4. Von Neumann, J. in Metal Interfaces 108–110 (Am. Soc. Metals, Cleveland, 1952).
5. Mullins, W. W. J. Appl. Phys. 27, 900–904 (1956).
6. Santaló, L. Integral Geometry and Geometric Probability (Cambridge Univ. Press, 2004).
7. Hadwiger, H. Vorlesungen über Inhalt, Oberfläche, und Isoperimetrie (Springer, Berlin, 1957).
8. Klein, D. A. & Rota, G.-C. Introduction to Geometric Probability (Cambridge Univ. Press, 1997).
9. Chen, B. Geom. Ded. 105, 107–120 (2004).
10. Barmak, K. et al. in Progress in Nonlinear Differential Equations and Their Applications Vol. 68 (eds Dal Maso, G., De Simone, A. & Tomarelli, F.) 1–11 (Birkhäuser, Boston, 2006).

STEM CELLS

Blood lines from embryo to adult

Hiroo Ueno and Irving L. Weissman

Does blood formation in mammalian embryos and adults have separate origins or a common source? The most recent investigations into the question add a further chapter to this long-running story.

Thirty years ago the consensus view was that blood formation — haematopoiesis — originates in stem cells in the embryo yolk sac, with that same cell lineage going on to supply the fetus and adult. This was supplanted by the now-conventional wisdom that the yolk sac serves only the embryo, and that an independent origin of blood-forming stem cells begins in the region of the developing fetus known as the aorta–gonad–mesonephros (AGM). In the latest turn of events, however, Samokhvalov and colleagues (page 1056 of this issue[1]) provide powerful evidence that yolk-sac haematopoiesis does indeed supply the adult as well.

In mammals, blood cells are the first differentiated cell types generated in early-stage embryos. The first blood cells are found in mice in the yolk sac at embryonic day (E) 7.0–7.5 , just 2 or 3 days after the undifferentiated conceptus implants in the uterus. They first appear as islands of blood cells, which are later surrounded by the endothelial cells that will form blood vessels. The yolk sac supports primitive blood formation, with these early blood cells being mainly nucleated red blood cells carrying the embryonic and fetal-type haemoglobins

that are presumably optimal for the developing embryo and then the fetus — a transition that is marked by the formation of distinct organs.

One and a half days after these blood islands appear, at about the time the yolk-sac blood vessels connect to the embryo proper, blood formation in the embryo is first observed at or near the dorsal aorta[2]. Blood-forming activity attributed to haematopoietic stem cells (HSCs) is evident in this region at around E10.5 (ref. 3). The site of blood formation then migrates to the fetal liver, thymus and spleen, and eventually to the bone marrow.

Is there a lineage relationship between yolk-sac blood-forming cells and those found later in the embryo, or are they generated separately? In 1970, the answer seemed clear[4]. In tissue culture, E7, 8 and 9 yolk-sac cells were found to produce a subset of blood-cell types in myeloerythroid colonies, which contain a mixture of red blood cells and myeloid cells, but not lymphocytes. Yolk-sac cells from similar stages also made the same type of colonies in the spleens of irradiated mice 7 days after injection into the mice. No such activity was present in E7–8 embryos, and in culture the

embryo gained such cells only if co-cultured with yolk sac.

In subsequent work[5], in which E8–9 yolk-sac cells were transplanted into the yolk sacs of genetically distinct embryos at the same stage, the cells entered 'blood islands'. When these mice reached adulthood, their bone marrow contained yolk-sac, donor-derived marrow cells that, when transplanted, gave day 10 myeloerythroid spleen colonies. These yolk-sac-cell recipients also contained donor-derived T lymphocytes in the thymus. So it seemed that the issue was settled, as only HSCs produce myeloerythroid and T cells for life.

However, experiments involving a new method of organ culture[2,3] apparently contradicted these findings[4,5]. In these assays, the AGM contained blood-forming progenitors, but much lower or no activity was observed in the yolk sacs of fetuses of the same age. It was also found that the blood-forming cells in the yolk sac cannot engraft in marrow in lethally irradiated adults in the way that they can in same-age fetuses. In these studies[2,3], the transplantable cells were first found in the AGM. Hence, the current wisdom is that
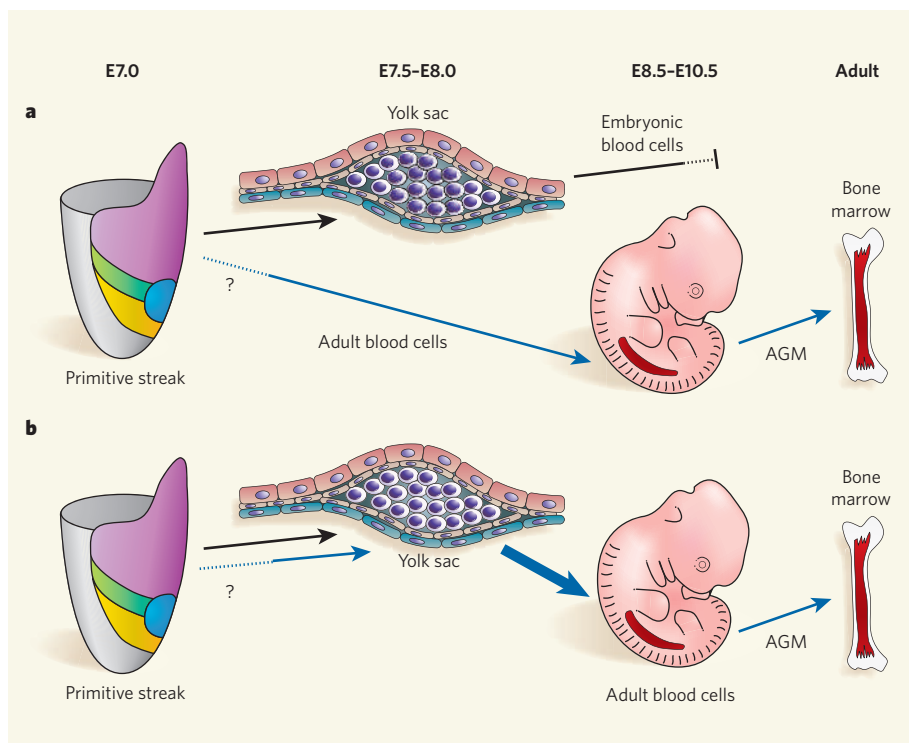
Figure 1 | Two models of embryonic and adult blood formation. a, The 'separate' model. Embryonic blood-forming cells, generated in a region called the 'primitive streak', proliferate in the yolk sac, migrate to the embryo proper, but eventually die out. The adult blood-forming cells, whose origin in the early-stage embryo is unknown, are separately generated in the aorta–gonad–mesonephros (AGM) region and later seed the adult bone marrow. **b,** The 'common' model. Adult blood-forming cells in the AGM regions at least partly originate from *Runx1*-positive progenitors in the yolk sac. Black and blue arrows show pathways of embryonic and adult blood-forming cells, respectively; bold blue arrow is the pathway described by Samokhvalov and colleagues[1]. Regions of cells at the primitive-streak stage: notochord (blue); somites (yellow); heart and cranial mesoderm (green); extraembryonic mesoderm (pink)[11].

the yolk-sac-derived, primitive blood-forming cells are transient and die, whereas all the adult blood-forming cells originate from the AGM (Fig. 1a).

Nevertheless, it remained possible that HSCs in the AGM migrate from the yolk sac to the embryo. In that case they would need to undergo further maturation to successfully make their way to the liver, where blood formation next occurs, before they could acquire the ability to engraft in adult bone marrow. In fact, in mice lacking a homing receptor for adult HSCs, the integrin $\alpha_4\beta_1$ (ref. 6), haematopoiesis never proceeds beyond the yolk sac to the fetal liver[7,8]. In addition, co-culture of E8.5 yolk-sac cells with AGM-derived cells for 4 days produces adult-type HSCs that can reconstitute adult bone marrow[9]. Like early yolk-sac cells, embryonic-stem-cell (ESC)-derived blood-forming cells do not engraft in adult marrow. However, introduction of developmental genes (*Cdx4* and *Hoxb4*) into ESC-derived haematopoietic cells gives cells that have gained some ability to become engrafted in adult bone marrow[10]. These studies[6–10] suggest that, to contribute to adult haematopoiesis, yolk-sac blood cells have to be subject to further events as they mature in embryos.

The origin of haematopoietic cells in early embryos has been examined by fate-mapping studies that allow the history of various cell lineages to be followed. Progenitors of yolk-sac blood cells are generated in a defined region in early-stage embryos[11], and common progenitors of blood and endothelial cells, expressing the genes *brachyury* and *Flk1*, are reported to be produced in this region[12]. The yolk-sac blood islands are, however, formed by multiple progenitors that are mainly already lineage-committed to a blood or an endothelial fate. At most, 15% of these yolk-sac blood cells could derive from local progenitors that assume both blood and endothelial fates; put another way, only about 30% of blood cells could originate from *Flk1*-positive progenitors[13]. Thus there seems to be a heterogeneous population of haematopoietic and endothelial progenitors at this stage, and their relationship to later HSCs is unclear.

Samokhvalov *et al.*[1] now report experiments involving an elegant, inducible cell-labelling technique, which they use to follow cells expressing an essential gene for blood formation called *Runx1* or *aml1* (later in life this gene is implicated in causing acute myeloid leukaemia, hence the latter name). Samokhvalov *et al.* labelled *Runx1*-positive cells at E7.5, before the establishment of vascular connections between the yolk sac and embryo, and before *Runx1*-positive cells are

generated within the embryo itself. At that stage, it is mainly yolk-sac cells that are *Runx1*-positive, although some *Runx1*-positive cells appear slightly later in the allantois and chorion, precursors to the placenta[14].

The authors[1] show that these *Runx1*-marked cells can contribute to adult HSCs (but not to vascular endothelial cells, indicating early segregation of the two lineages) that persist for at least 15 months after birth. Because almost all *Runx1*-positive cells are in the yolk sac at the time of lineage tracing, this outcome indicates that at least some adult HSCs come from the yolk sac, contrary to the view that yolk-sac and intraembryonic haematopoiesis are completely separated in mammals (Fig. 1).

In Samokhvalov and colleagues' results[1], the contribution of *Runx1*-positive yolk-sac cells to adult haematopoiesis is partial. But because their method of cell marking may itself be partial, and *Runx1*-negative yolk-sac cells may also contribute to adult HSCs, the contribution of yolk-sac cells to adult haematopoiesis might be greater — even 100%. On the other hand, it is possible that progenitors of AGM HSCs really do originate only in part from yolk-sac cells. So it will be necessary to find out whether the yolk-sac/AGM pathway is essential for normal haematopoiesis, or if it is redundant and there are two distinct pathways. Even if all the progenitors for AGM HSCs go through the yolk sac, the origins of embryonic and adult haematopoietic cells in early embryos could be different (Fig. 1b). It is also not clear whether the later-developing allantois/chorion–placenta is an independent source of haematopoiesis or whether it also arises from yolk-sac cells[14]. Although these issues are yet to be resolved, the yolk-sac origin of at least some adult haematopoiesis is now back on the books. ∎

Hiroo Ueno and Irving L. Weissman are at the Institute of Stem Cell Biology and Regenerative Medicine, Departments of Pathology and Developmental Biology, Stanford University School of Medicine, Stanford, California 94305, USA.
e-mail: hueno@stanford.edu

1. Samokhvalov, I. M., Samokhvalova, N. I. & Nishikawa, S. *Nature* **446,** 1056–1061 (2007).
2. Cumano, A., Ferraz, J. C., Klaine, M., Di Santo, J. P. & Godin, I. *Immunity* **15,** 477–485 (2001).
3. Medvinsky, A. & Dzierzak, E. *Cell* **86,** 897–906 (1996).
4. Moore, M. A. & Metcalf, D. *Br. J. Haematol.* **18,** 279–296 (1970).
5. Weissman, I., Papaioannou, V. & Gardner, R. in *Differentiation of Normal and Neoplastic Hematopoietic Cells* (eds Clarkson, B., Marks, P. A. & Till, J. E.) 33–47 (Cold Spring Harbor Laboratory Press, New York, 1978).
6. Wagers, A. J., Allsopp, R. C. & Weissman, I. L. *Exp. Hematol.* **30,** 176–185 (2002).
7. Potocnik, A. J., Brakebusch, C. & Fassler, R. *Immunity* **12,** 653–663 (2000).
8. Arroyo, A. G., Yang, J. T., Rayburn, H. & Hynes, R. O. *Cell* **85,** 997–1008 (1996).
9. Matsuoka, S. *et al. Blood* **98,** 6–12 (2001).
10. Wang, Y., Yates, F., Naveiras, O., Ernst, P. & Daley, G. Q. *Proc. Natl Acad. Sci. USA* **102,** 19081–19086 (2005).
11. Kinder, S. J. *et al. Development* **126,** 4691–4701 (1999).
12. Huber, T. L., Kouskoff, V., Fehling, H. J., Palis, J. & Keller, G. *Nature* **432,** 625–630 (2004).
13. Ueno, H. & Weissman, I. L. *Dev. Cell* **11,** 519–533 (2006).
14. Zeigler, B. M. *et al. Development* **133,** 4183–4192 (2006).

OBITUARY

John Backus (1924–2007)

Inventor of science's most widespread programming language, Fortran.

John Backus, who died on 17 March, was a pioneer in the early development of computer programming languages, and was subsequently a leading researcher in so-called functional programming. He spent his entire career with IBM.

Backus was born on 3 December 1924 in Philadelphia, and was raised in Wilmington, Delaware. His father had been trained as a chemist, but changed careers to become a partner in a brokerage house, and the family became wealthy and socially prominent. As a child, Backus enjoyed mechanical tinkering and loved his chemistry set, but showed little scholastic interest or aptitude. He was sent to The Hill School, an exclusive private high school in Pottstown, Pennsylvania. His academic performance was so poor that he had to attend summer camps to catch up on his studies. He fared no better at the University of Virginia, where he enrolled for a degree in chemical engineering in 1942, and from which he was sent down for poor attendance.

Backus was drafted into the US Army in early 1943, where he was initially put on an anti-aircraft programme. Aptitude tests revealed his high intelligence, and he was sent on a specialist engineering course at the University of Pittsburgh, Pennsylvania. He also tried a pre-medical training course, but lost enthusiasm and dropped out. After leaving the army in May 1946, he enrolled at Columbia University, New York, initially as a probationary student because of his poor academic record. In mathematics, he at last found his calling, graduating with a BA in 1949 and obtaining a master's degree the following year.

In 1948, IBM had completed its first experimental electronic computer, known as the Selective Sequence Electronic Calculator (SSEC). It was a huge machine with 13,000 tubes and 23,000 relays. IBM displayed the machine in the showroom of its New York headquarters, where it was visible to passers-by and attracted much media attention. Backus went to see the SSEC, and persuaded IBM to hire him as a programmer. He spent the next year of his life calculating lunar positions, which he found a delight.

IBM introduced its first commercial computer product, the model 701, in 1952. The computer had to be programmed at a level very close to basic binary machine code, which was not intrinsically difficult but time-consuming and error-prone. Backus devised an automatic programming system called Speedcoding for the 701, which made the task much easier. Like similar systems being developed elsewhere around the same time, however, Speedcoding produced programs that were uneconomically slow.

Backus proposed to his manager the development of a system he called the Formula Translator — later contracted to Fortran — in autumn 1953 for the model 704 computer, which was soon to be launched. The unique feature of Fortran was that it would produce programs that were 90% as good as those written by a human programmer. Backus got the go-ahead, and was allocated a team of ten programmers for six months. Designing a translator that produced efficient programs turned out to be a huge challenge, and, by the time the system was launched in April 1957, six months had become three years. The Formula Translator consisted of 18,000 instructions, which was not especially long as programs went, but it embodied fiendishly complex algorithms for code optimization.

Fortran was eagerly taken up by users of the IBM 704, and other computer manufacturers also produced Fortran systems so that their machines' software would be compatible with IBM's. Fortran became the *lingua franca* of scientific computing, which it remains 50 years on — today, for example, the UK Meteorological Office's computer model for climate change consists of a one-million-line Fortran program. Fortran has often been criticized for its inelegance, but as Backus subsequently explained, when it was being designed efficiency was paramount and very little thought was given to the language itself. And of course, no one could have dreamed that the language would still be going strong half a century later. Because of the need for backward compatibility, Fortran has never escaped the path-dependency of those early decisions.

In the late 1950s Backus became a member of the Algol Committee, which was designing an international scientific programming language, named Algol 60. The language was specified in Backus–Naur form, a notation that Backus devised in collaboration with the Danish computer scientist Peter Naur. Algol 60 was an elegant language that was popular in Europe in the 1960s and 1970s. However, it was never able to replace Fortran, primarily because of the investment already made in the Fortran code. Nonetheless, Algol was hugely influential in programming-language design — most modern programming languages, such as C and Java, can trace their roots to the language, and the Backus–Naur form is part of the computer-science canon.

After the pizzazz of the Fortran years, Backus's subsequent career was calmly scientific. In 1960 he became a staff member of IBM Research, Yorktown Heights, New York. In 1963 he was made an IBM fellow and spent the remainder of his career, until he retired in 1991, with IBM in both Yorktown Heights and San Jose, California.

At IBM Research, Backus led a research group in functional programming, a new programming paradigm that aimed at computation through the evaluation of mathematical functions. The idea was to bypass what Backus called the 'von Neumann bottleneck', after John von Neumann, one of the inventors of the computer. As Backus explained it, a computer consisted of a processor and a memory; the object of a program was to change the state of the memory, using the processor, but this had to be done painfully slowly, one instruction at a time. But although functional programming became (and remains) a major computer-science research topic worldwide, it has never broken through to the mainstream.

Fortran remained Backus's lasting contribution to computing, for which he was awarded the National Medal of Science in 1975 and the Turing Award of the Association of Computing Machinery in 1977 — computer science's highest honour. ∎

Martin Campbell-Kelly
Martin Campbell-Kelly is in the Department of Computer Science, Warwick University, Coventry CV4 7AL, UK.
e-mail: m.campbell-kelly@warwick.ac.uk

IBM/AP

GLYCOCHEMISTRY & GLYCOBIOLOGY

natureinsight
GLYCOCHEMISTRY & GLYCOBIOLOGY
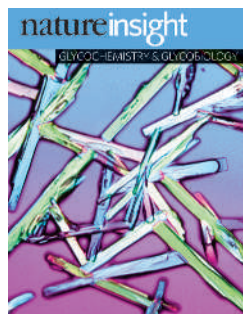
Cover illustration
Polarized light micrograph showing crystalline glucose, one of the simplest sugars. (Courtesy of Science Photo Library.)

Carbohydrates have long been underappreciated by the scientific community, and many scientists approach the complex structures and elaborate nomenclature of carbohydrates with trepidation. Like amino acids and nucleic acids, sugars are abundant in nature: many natural products contain oligosaccharides that are vital for their biological activity, and carbohydrates have key roles in a broad range of biological processes, including signal transduction and immune responses.

Although fewer scientists work with carbohydrates than with other biopolymers, researchers in this field have been prolific. Chemists and biochemists have developed new methods to rapidly synthesize oligosaccharides, enabling them to generate complex polysaccharides and analogues of natural products that have increased activity *in vivo*. Biologists have explored the physiological roles of various sugars, discovering that many have essential roles in all of the major organ systems and are involved in several disease states. In addition to extending our knowledge of how the natural world works, these findings have been used to develop carbohydrate-based drugs and vaccines, some of which show great promise for treating or preventing various diseases, including malaria, cancer and AIDS.

The number of chemists, biochemists and biologists in this field is steadily increasing, and this interest underscores the fact that there are so many discoveries to be made. To quote Ajit Varki (see page 1023), it is "a fertile area for the new generation of young scientists".

In this Insight, we present a collection of reviews that highlights some of the hottest areas in glycochemistry and glycobiology, including the chemical synthesis of carbohydrates, their biological functions and the therapeutic potential of carbohydrate-based drugs and vaccines. We hope that you enjoy it.

Joshua Finkelstein, Senior Editor

REVIEWS

nature insight

Chemical glycosylation in the synthesis of glycoconjugate antitumour vaccines

Danica P. Galonić[1] & David Y. Gin[2]

Therapeutic vaccines derived from carbohydrate antigen–adjuvant combinations are a promising approach for cancer immunotherapy. One of the critical limitations in this area is access to sufficient quantities of tumour-associated carbohydrate antigens and glycoconjugate adjuvants. At present, availability of the complex oligosaccharide constructs that are needed for the systematic design and evaluation of novel vaccine formulations relies on *de novo* chemical synthesis. The use of both state-of-the-art and emerging glycosylation technologies has led to significant advances in this field, allowing the clinical exploration of carbohydrate-based antigens in the treatment of cancer.

The identification of distinct glycoprotein and glycolipid constructs that are overexpressed on the cell surfaces of malignant cells[1–3] has spurred intense research into exploiting these tumour-associated antigens for the development of anticancer vaccines[4–6]. Tumour-associated carbohydrate antigens are anchored to the cell surface either by a lipid tail (for example, the glycolipids Globo-H, GM2, GD2 and GD3) or by a protein component (for example, glycoproteins that have an *N*-acetyl-galactosamine core (GalNAc) oligosaccharide, such as T_N, T and sialyl-T_N (ST_N))[5]. Although glycoprotein and glycolipid carbohydrate epitopes have been used in anticancer vaccine investigations, these molecular subunits typically induce only weak T-cell-independent B-cell (antibody) responses, wherein only a small population of glycolipids might be presented to T cells. However, augmentation of antibody response can be accomplished with three-component vaccine formulations. These constructs involve the covalent conjugation of various carbohydrate antigens to an immunocarrier protein such as keyhole limpet haemocyanin (KLH)[7]. When this protein is processed and presented by antigen-presenting cells a strong T-cell immune response results, leading to cytokine cascades that increase antibody response, not only to KLH but also to the less immunogenic carbohydrate antigens to which it is attached. Additional immune response potentiation of these antigen conjugates is accomplished through co-administration with an immunological adjuvant such as QS-21 (refs 8–10). Marked eradication of circulating tumour cells and micrometastases in preclinical models, together with prolonged disease-free periods and survival after primary treatment (such as radiation or surgery), highlighted the early potential of carbohydrate vaccines in cancer immunotherapy[8,11–13]. Several additional three-component anticancer vaccine formulations have since advanced to clinical trials[14–20], and large randomized, multicentre trials of polyvalent vaccines such as these will soon address the clinical impact of immunization against carbohydrate antigens.

Low expression levels of tumour-associated carbohydrate antigens in tissue cultures of cancer cells, as well as difficulties associated with isolating homogeneous material from such sources, prompted extensive research into the chemical synthesis of these complex molecules as their primary means of access[21–24]. Various synthetic strategies have been developed for the preparation not only of these complex oligosaccharide antigens but also of complex carbohydrate immunoadjuvants. The

formidable intricacies of regiochemical and stereochemical control in oligosaccharide assembly establish complex carbohydrate construction as one of the most challenging forefronts in chemical synthesis. Here we present an overview of selected glycosylation processes and strategies applied to the synthesis of tumour-associated carbohydrate antigens (TACA) and immunoadjuvants that show promise in anticancer therapeutic vaccines.

Glycosylation strategies

Acetal exchange

Many of the advances in complex carbohydrate synthesis revolve around methods to form the glycosidic bond, because this is the primary means by which monosaccharide building blocks are assembled into more complex oligosaccharide structures[25–32]. For more than a century, methods developed for glycosylation have overwhelmingly favoured an approach in which the carbohydrate coupling partner contributing its anomeric carbon in the anomeric linkage (glycosyl donor) serves the role of an electrophile. The corresponding coupling partner (glycosyl acceptor) thus functions as the nucleophilic counterpart. This strategy typically relies on the use of a selectively protected glycosyl donor (**1**, Fig. 1a), which incorporates at its anomeric centre a latent leaving group (Lg). In the presence of a suitable electrophilic 'activator' (El+), the anomeric functionality is rendered highly electron deficient (**2**), thereby allowing anomeric substitution by the nucleophilic glycosyl acceptor (Nu–H) to form the glycoconjugate **3**. This acetal exchange process governs most existing chemical glycosylation processes, and is effective for coupling not only with simple nucleophiles but also complex oligosaccharide, peptide and lipid glycosyl acceptors.

One of the more direct approaches for glycosylation involves the class of C1-hydroxy glycosyl donors (**4**, Fig. 1b), in which an unprotected anomeric hydroxyl is exchanged under a controlled dehydration process. Various dehydrating reagents have proved effective, including cationic metal Lewis acids, sulphonium and phosphonium salts, and activated sulphonyl halides. Although this approach minimizes the number of distinct anomeric derivatizations in the glycosylation protocol, alternative strategies involving functionalization of the anomeric hydroxyl into an isolable glycosyl donor before anomeric activation and coupling are often desirable. Among the earliest latent leaving groups used in

[1]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA.

chemical glycosylations are anomeric halides (**5**), wherein glycosyl fluorides, chlorides, bromides and iodides can be prepared and activated for glycosylation with their respective halophilic reagents. Other anomeric latent leaving groups involving oxygen-derived functionalities include trichloroacetimidates (**6**), phosphates/phosphites (**7**), esters/carbonates/thiocarbonates (**8**) and various aryloxy groups (**9**), all of which can be efficiently activated by the corresponding oxo-, aza- or thiophilic reagents. Also among the anomeric O-derived glycosyl donors is the 4-pentenyl glycoside **10**, in which electrophiles with high affinity for π-electrons (such as electrophilic halogen reagents) have proved to be useful selective activators. Like the 4-pentenyl glycosyl donors, glycosyl sulphides and sulphoxides (**11**) have been shown to be exceedingly useful not only as latent leaving groups but also as stable anomeric protective groups before the glycosylation event.

Glycal oxidation

In addition to carbohydrate donors with anomeric heteroatom derivatives, the use of glycals (**12**, Fig. 1c) as glycosyl donors has been explored extensively in complex carbohydrate synthesis. The presence of the 1,2-alkene functionality in this substrate allows the use of various electrophilic oxidants (El^+) that are reactive to enol ether nucleophiles (**12**). The resulting activated glycosyl donor **13** is then poised to receive an appropriate nucleophilic glycosyl acceptor (Nu–H) to form the glycoconjugate **14**. These methods allow the introduction of various functionalities 'Z' at the C2-position in conjunction with anomeric bond formation. For example, C2-oxygen transfer to glycal donors has proved useful, involving dimethyldioxirane (DMDO)-mediated[33] or sulphonium-mediated[34] 1,2-epoxidation of glycals, followed by anomeric substitution, to generate C2-hydroxy glycosides (**15**, Fig. 1d). Similarly, I(III)-containing reagents have recently been applied to the generation of selectively protected C2-acyloxyglycosides (**16**)[35]. The synthesis of 2-amino-2-deoxyglycosides has drawn considerable attention owing to the abundance and importance of this class of glycoside in naturally occurring glycoconjugates. The venerable C2-azidonitration reaction of glycals[36] to afford C2-azido pyranose derivatives (**17**) has been, and continues to be, a favoured method by which to introduce the C2-*N*-functionality. Other reactions have subsequently been developed for C2-nitrogen transfer onto the glycal donor, including stereoselective installation of a sulphonamide group (for instance, **18**, by iodosulphonamido glycosylation)[33], as well as less-used protocols to install a C2-carbamate functionality[37] and the naturally occurring C2-acetamido group[38].

This series of acetal exchange couplings and glycal glycosylations (Fig. 1) has been used, as have others, in the chemical synthesis of highly complex oligosaccharide conjugates. However, so far no single coupling method has proved to be broadly effective for all glycosylations, no doubt as a result of the high structural and functional group variability of carbohydrate substrates. This is the case for the exceedingly complex glycoconjugates identified for potential use in cancer immunotherapy, in which chemical glycosylation processes used to prepare these molecules are often highly substrate-specific.

Synthesis of tumour-associated carbohydrate antigens

Globo-H

Globo-H (**19**, Fig. 2a) is a cell-surface glycosphingolipid expressed on a number of epithelial tumours, including those of the breast, prostate and ovary[39–41]. Its complex hexasaccharide has been a target of many total synthesis approaches. The DMDO-mediated glycal assembly method[33] was elegantly applied to the first total synthesis of a Globo-H oligosaccharide[42,43] (Fig. 2b). This approach is illustrated by the stereoselective α-epoxidation of galactal **20** followed by regioselective C3-O-glycosylation of galactal diol **21**. The resulting glycal disaccharide **22** is then immediately poised for glycosylation with a selectively protected fucose donor to afford the glycal trisaccharide **23**. Synthetic incorporation of 2-amino-2-deoxyglycoside derivatives often presents a challenge in complex carbohydrate synthesis, wherein stereoselective nitrogen-transfer to glycal substrates has proved effective. In this late-stage transformation, activation of the glycal enol ether in **23** with iodonium di-sym-collidine
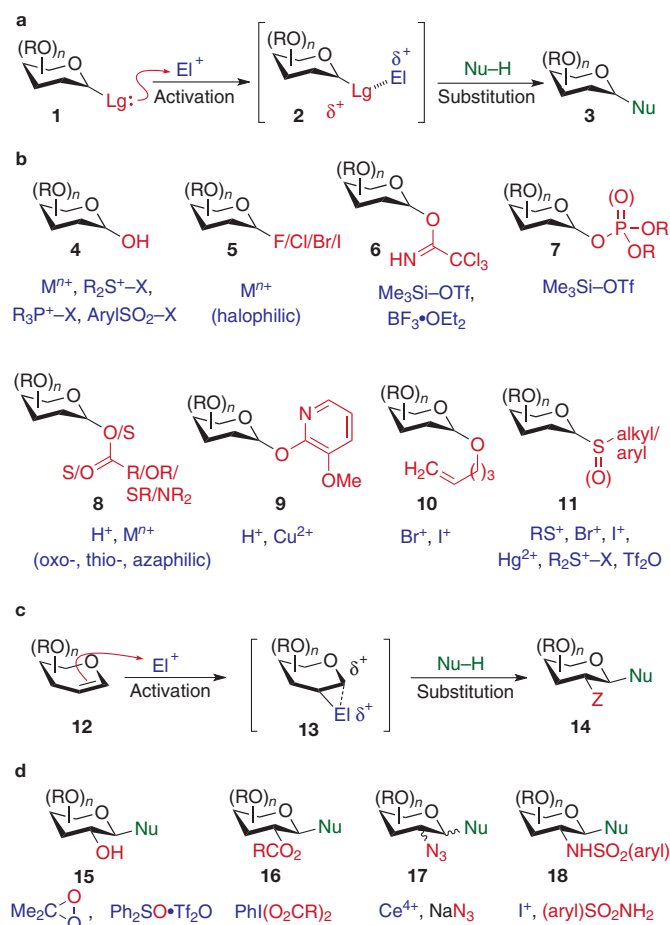

Figure 1 | Glycosylation methods. a, b, Glycosylation of acetal-derived glycosyl donors. Activation of the anomeric leaving group (Lg, red) with an electrophilic promoter (El^+, purple) is followed by nucleophilic attack of the acceptor (Nu–H, green) on the resulting electron-deficient anomeric carbon of the carbohydrate donor. **c, d,** Glycosylation with glycal donors. Activation of glycals with various electrophiles (El^+) is followed by coupling with a glycosyl acceptor (Nu–H) at the anomeric carbon. These glycosylations result in functionalization of both the C1 and C2 positions of the donor. M, metal; R, various substituents; Tf, trifluoromethanesulphonyl; X, various leaving groups; Z, various functionalities.

perchlorate (I(coll)$_2$ClO$_4$) generates a transient β-iodonium intermediate that rapidly receives a PhSO$_2$NH$_2$ nucleophile at the anomeric position. The resulting glycosyl sulphonamide group is then transferred to the C2-position in the presence of a base (lithium hexamethyldisilazide, or LHMDS) through the generation of a putative aziridine intermediate that is subsequently opened by ethane thiol. The resulting trisaccharide **24**, which incorporates the requisite protected C2-aminogalactoside donor, is ready for thioglycoside coupling with an appropriate trisaccharide acceptor (ROH) to afford the hexasaccharide **25**, which can be readily elaborated to the Globo-H antigen (**19**).

Subsequent syntheses of the Globo-H oligosaccharide used the concept of orthogonal glycosylation. By capitalizing on differences in reactivity of various anomeric leaving groups, in conjunction with the control of nucleophilicity of the acceptor, a convergent synthesis of the Globo-H hexasaccharide was accomplished (Fig. 2c). From five distinct carbohydrate building blocks (**26–30**) with either an anomeric thioalkyl, thioaryl or fluoride functionality, sequential chemoselective anomeric activations were conducted in an orthogonal two-directional glycosylation approach[44]. For example, the ethylthio fucoside donor **26** could be selectively activated in the presence of the less reactive anomeric phenylthio group of the acceptor **27**. After this initial glycosylation of **27** with **26**, the remaining anomeric phenylthio group was activated (with
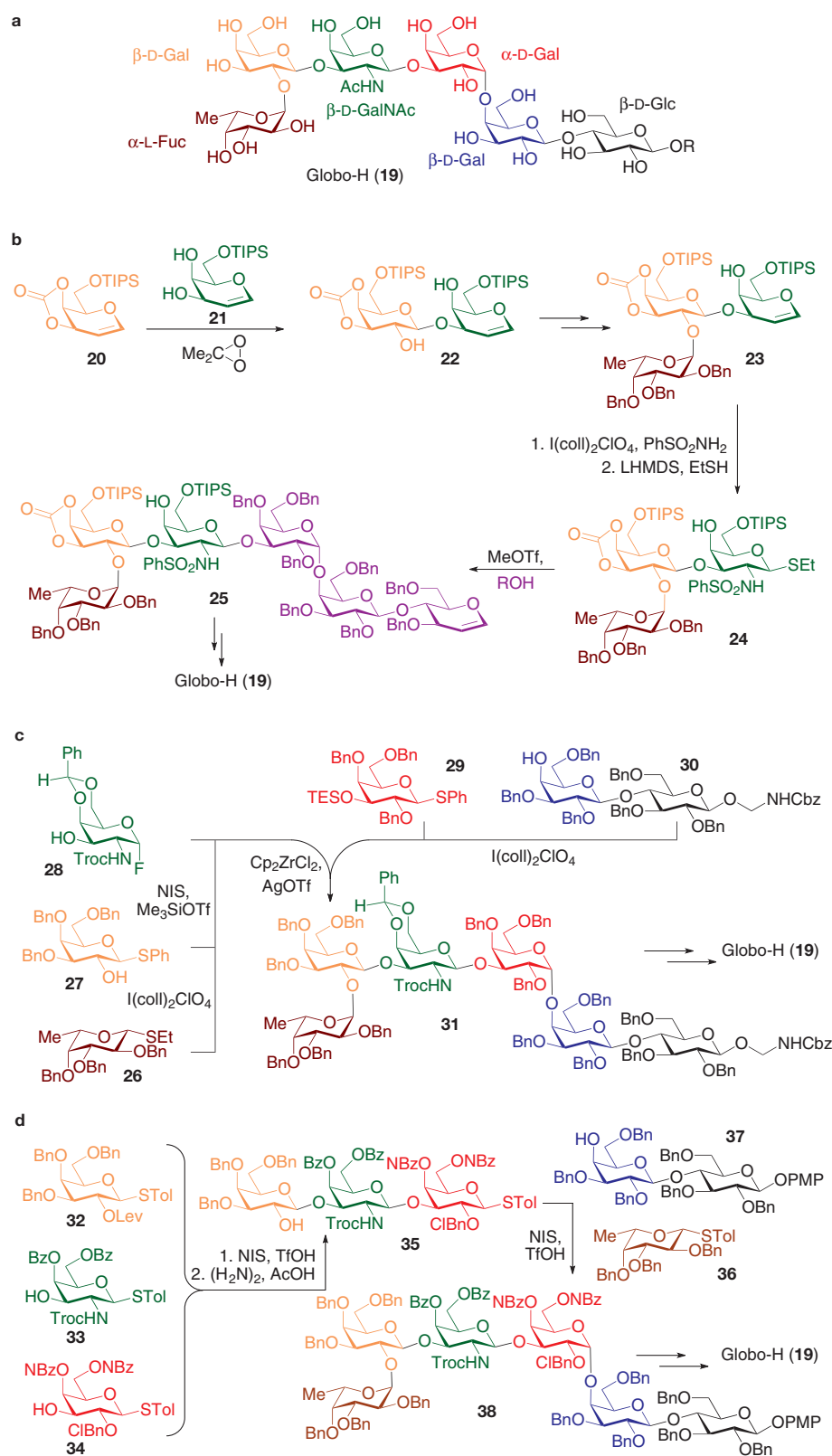
Figure 2 | Selected syntheses of tumour-associated Globo-H antigen hexasaccharide. a, The chemical structure of Globo-H. **b,** Synthesis through glycal assembly using oxidative C2-hydroxyglycosylation and oxidative C2-sulphonamidoglycosylation. **c,** Synthesis by an orthogonal two-directional glycosylation strategy using thioglycoside and glycosyl fluoride donors. **d,** Synthesis through a reactivity-based one-pot multiple glycosylation strategy. Bn, benzyl; Bz, benzoyl; Cbz, benzyloxycarbonyl; ClBn, 2-chlorobenzyl; coll, 2,4,6-collidine; Cp, cyclopentadienyl; Fuc, fucose; Lev, laevulinoyl (4-oxopentanoyl); LHMDS, lithium hexamethyldisilazide; NBz, 4-nitrobenzoyl; NIS, *N*-iodosuccinimide; PMP, 4-methoxyphenyl; TES, triethylsilyl; TIPS, triisopropylsilyl; Tol, toluyl (4-methylphenyl); Troc, trichloroethoxycarbonyl.

N-iodosuccinimide (NIS) in the presence of the galactosyl acceptor **28**, whose anomeric fluoride group remained unreactive to the thiophilic NIS reagent. The resulting glycosyl fluoride trisaccharide was then used as the donor for the glycosylation of the remaining Galα1-4Galβ1-4Glc trisaccharide fragment (where Gal denotes galactose and Glc denotes glucose) itself obtained from the glycosylation of the protected lactose acceptor **30** with the thiogalactoside **29**. In the late-stage convergent step, fluoride activation (by biscyclopentadienyl zirconium dichloride (Cp_2ZrCl_2) and silver trifluoromethanesulphonate (AgOTf)) with concomitant triethylsilyl (TES) deprotection allowed stereoselective coupling of both trisaccharide fragments to provide the protected hexasaccharide **31**, which served as a suitable advanced substrate for the completion of the synthesis. Notably, no anomeric protecting group interconversions were necessary after any of the glycosylation steps in the synthetic sequence.

The orthogonal reactivity concept was exploited on a different dimension in the reactivity-based one-pot strategy[45] for the construction of Globo-H hexasaccharide[46] (Fig. 2d). This effort relied on known relative reactivities of various thioglycoside donors, the reactivities of which were finely tuned by the careful choice of proximal protective groups[47]. The synthesis was initiated through the use of three distinct monosaccharide donors (**32**–**34**), all of which incorporated the anomeric thiotoluyl latent leaving group. When this mixture of thioglycosides was treated with electrophilic activators (trifluoromethanesulphonic acid (TfOH) and NIS), the electron-rich thioglycoside donor **32** was activated most rapidly to condense with the most nucleophilic alcohol acceptor **33**. Subsequent glycosylation of **34** with the remaining anomeric thiotoluyl group on the resultant disaccharide in the one-pot reaction generated the trisaccharide **35** after selective removal of the laevulinate (Lev) protective group. A similar one-pot multiple glycosylation process was applied to a mixture of the trisaccharide **35**, the fucosyl donor **36**, and the lactose-derived acceptor **37**, securing the hexasaccharide **38** for eventual advancement to the Globo-H hexasaccharide. This one-pot strategy resulted in the assembly of the hexasaccharide core with minimal isolation and purification of oligosaccharide intermediates[48]. Other notable syntheses of Globo-H have involved exclusive application of glycosyl trichloroacetimidates[49] and glycosyl phosphates[50], demonstrating the power and versatility of these classes of glycosyl donor.

Sialylated gangliosides
Sialylated gangliosides, exemplified by GM2 (**39**), GD3 (**40**), and GD2 (**41**) (Fig. 3a), are cell-surface glycolipids expressed in a number of neuroectodermal cancers

(including melanoma, neuroblastoma, sarcoma and small-cell lung cancer) and also, in the case of GM2, in several epithelial cancers (breast, prostate, ovary and colon)[3,51]. In addition to the typical challenges associated with the chemical synthesis of glycolipids, an added difficulty is the incorporation of sialic acid residues such as neuraminic acid (NeuAc) into complex oligosaccharides[52]. Glycosylations with sialic acid donors are often plagued by low yields, because anomeric couplings must occur at the sterically encumbered C2-position through a ketal exchange process. Moreover, naturally occurring sialosides incorporate the α-C2-stereochemical configuration, which is the contra-thermodynamic equatorial isomer, devoid of electronic stabilization through anomeric effects. Anomeric leaving groups in sialylation reactions[52] have taken the form of thio-derivatives, halides and phosphites, as well as, to a lesser extent, sialic acid-2,3-glycals and the underivatized hemiketal[53]. Efforts to control α-anomeric selectivity in these couplings have used removable neighbouring auxiliary groups (either at C3 or at C1), or beneficial solvent participation effects in the anomeric substitution event.

An early synthesis of the GM2 (**39**, Fig. 3b) antigen made effective use of the β-sialyl phosphite donor **42** in the regioselective sialylation of the C3-hydroxyl of the lactoside acceptor **43** (ref. 54). By capitalizing on the use of nitrile solvents to effect α-sialylations, presumably through a transient β-nitrilium intermediate[55–57], Sn(II) catalysed glycosylation in acetonitrile provided the α-sialoside **44** in good yield and α-stereoselectivity. Attachment of the remaining protected galactosamine residue **45** was accomplished with the trichloroacetimidate glycosylation[28] of the lone hydroxyl group in **44** to provide the tetrasaccharide **46**, which was subsequently converted to GM2 (**39**) by straightforward protective group manipulations.

Further enhancement of reactivity in sialic acid glycosylations is evident in a synthesis of GD3 (**40**, Fig. 3c), wherein the C5-acetamido functionality in neuraminic acid is replaced with a C5-trifluoroacetamido group in the glycosylation substrates (such as **47**)[58]. With this modification, construction of the NeuAcα2-8NeuAc linkage was possible without the need for installation of neighbouring auxiliary groups. Thus, treatment of the thiosialoside donor **47** with NIS in acetonitrile solvent allowed α-selective C8-O-glycosylation of the sialyl acceptor **48**, also derivatized with the C5-trifluoroacetamido substituent. The beneficial effects of these modified amide substrates in both **47** and **48** are thought to arise from the improved reactivity in **47** as a more electron-deficient electrophilic donor and from the attenuation of a putative C5-amido-C8-hydroxyl hydrogen bond in the acceptor **48** that might otherwise compromise its role as a nucleophile. The attainment of **49** allowed elaboration to GD3 (**40**), after glycosylation of an appropriate lactoside acceptor and facile late-stage exchange of the trifluoroacetamide groups to their native acetamido counterparts.

Mucin-associated glycans

Mucin motifs make up another well-studied class of tumour-associated carbohydrate antigen[59]. Mucins are high-molecular-weight glycoproteins expressed on the surfaces of many epithelial cells that are
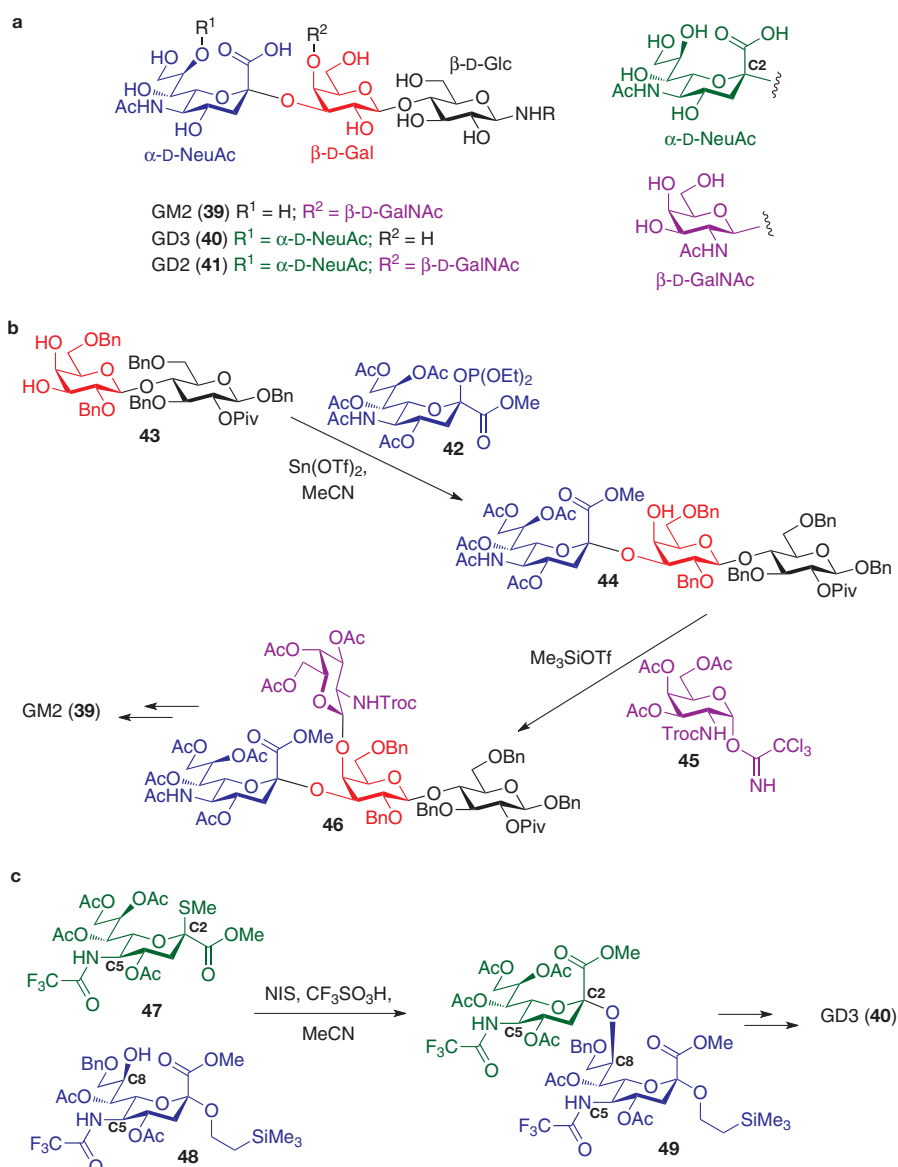
Figure 3 | Selected syntheses of tumour-associated sialylated glycosphingolipid oligosaccharides. **a**, GM2, GD3 and GD2 have a similar trisaccharide core. **b**, Synthesis of GM2 tetrasaccharide through α-selective sialyl phosphite donor glycosylation. **c**, Synthesis of GD3 tetrasaccharide by thiosialoside glycosylation. Incorporation of the C5-trifluoroacetamide group in both donor and acceptor enhances efficiency of the coupling. Piv, pivaloyl (2,2,2-trimethylacetyl).

characterized by the presence of GalNAc moieties on the hydroxyl side chains of Ser and Thr residues (which are often clustered) in the protein[60]. Further glycosyl transferase-controlled extension of the carbohydrate moiety results in the formation of oligosaccharides, which, among other functions, provide protection from proteolytic degradation and microbial infection. As a result of altered glycosyltransferase expression in tumour cells, premature termination of oligosaccharide biosynthesis leads to the formation of shortened, often sialylated, saccharide antigens as well as the exposure of peptide epitopes. Examples of tumour-associated carbohydrate antigens (Fig. 4a) include T$_N$ (**50**), T (**51**), ST$_N$ (**52**) and 2,6-ST (**53**)[10,61,62]. A number of these tumour-associated structural alterations represent a basis for the design of vaccines for selective eradication of tumours.

The chemical synthesis of mucin-derived TACAs involves a glycosylation challenge beyond carbohydrate–carbohydrate coupling: that of the construction of the linkage at the interface between the saccharide and the mucin-derived peptide. Formation of the GalNAcα1-O-Ser/Thr connection (Fig. 4a) highlights one of the key difficulties in achieving

C1,C2-*cis*-glycosylation. Indeed, a C2-acetamido group within a carbohydrate donor is overwhelmingly biased, through oxazoline intermediate formation, to afford the C1,C2-*trans* glycoside[63], a linkage with the anomeric configuration opposite to that of the mucin-related tumour-associated antigen.

To address this challenge, access to the GalNAcα1-*O*-Ser/Thr substructure is often accomplished through the glycosylation of the side-chain hydroxyl group of protected Ser or Thr derivatives with C2-azido donors (**56**; Fig. 4b), in which the non-participating nature of the azide substituent allows the stereoselective formation of the desired α-anomer. 2-Deoxy-2-azidogalactopyranose derivatives (**55**), obtained by azidonitration[36] of protected galactal substrates (**54**), can be converted to various glycosyl donors (**56**), including anomeric bromides and chlorides[64,65], trichloroacetimidates[66], thioglycosides[67,68] and pentenyl glycosides[69], all of which can be used for Ser/Thr glycosylation to generate the carbohydrate–amino-acid conjugate **58**. Importantly, these couplings are equally amenable to the construction of oligosaccharide–Ser/Thr conjugates[66,70–72]. After amino-acid glycosylation, the C2 azido group in **58** is converted, through reductive acetylation, to the naturally occurring acetamide group in **59** to allow subsequent use in iterative peptide synthesis.

A conceptually distinct approach (Fig. 4c) to the construction of mucin-type amino-acid–carbohydrate linkage involves the formation of the critical sugar–amino-acid bond from the potassium tert-butoxide-promoted conjugate addition of protected Ser and Thr nucleophiles (**61**) to a

Figure 4 | Mucin-related tumour-associated carbohydrate antigens. a, Structures of T_N, T, ST_N and 2,6-ST carbohydrate antigens. *O*-α-carbohydrate–amino-acid linkages can be synthesized by either (**b**) glycosylation of suitably protected Ser or Thr derivatives with various galactose-derived C2-azido donors, or (**c**) conjugate addition of amino-acid alkoxides into nitrogalactals. Boc, tert-butoxycarbonyl; Bu^t, tert-butyl; CAN, ceric ammonium nitrate; Y, various carbohydrate C1 groups; Z, various carbohydrate C2 groups.

2-nitrogalactal donor (**60**)[73]. Subsequent C2-nitro reductive acetylation and protective-group modification in **62** results in the preparation of glycosylated building blocks **63**, which can be readily elaborated for glycopeptide synthesis.

Once formed, the suitably protected glycosylated amino acids can be incorporated into a growing peptide chain in a modular fashion, allowing the preparation of clustered carbohydrate antigen displays, a motif common to tumour cell surfaces. Various oligopeptide–tumour antigen conjugates have been prepared in this manner, including T_N antigen constructs[74,75], ST_N antigen glycopeptides[64,76], and 2,3- and 2,6-ST conjugates[68,77].

Synthesis of carbohydrate immunoadjuvants
QS-21A$_{api}$
A third component that is required in clinically viable anticancer vaccines is a potent immunological adjuvant, a substance that is itself non-immunogenic but that significantly augments the immune response when administered together with the antigen–carrier conjugate. Among the most potent molecular immunological adjuvants used in antitumour vaccines is QS-21A (ref. 9; Fig. 5), a plant-derived complex saponin from the South American tree *Quillaja saponaria* Molina. Microgram quantities of this amphiphilic substance in combination with the antigen–carrier conjugate lead to enhancement in both antibody and cell-mediated immune response in a host of promising anticancer and antiviral vaccines[78]. The structure of QS-21A (Fig. 5a) is that of a complex triterpene–oligosaccharide–normonoterpene conjugate, incorporating quillaic acid as a central lipophilic core flanked by a branched trisaccharide, a linear tetrasaccharide, and an extended glycosylation diester side chain along its periphery. QS-21A has two constitutional isomers, in which the terminal saccharide residue on the linear tetrasaccharide substructure is either D-apiose (QS-21A$_{api}$, **64**) or D-xylose (QS-21A$_{xyl}$, **65**)[79]. Unfortunately, acquiring sufficient quantities of these natural products in pure form is fraught with technical challenges, because they exist in a mixture of more than 100 distinct amphiphilic congeners that are typically only partly purified by repeated high-performance liquid chromatography separation.

The first report of the chemical synthesis of both oligosaccharide cores of the natural product showcased expert control in stereoselective glycosylation in the synthesis of fully protected versions of both the branched trisaccharide fragment and the linear tetrasaccharide components of QS-21A$_{api}$ (**64**)[80]. Two glycosylation methods were used: NIS activation of thioglycoside donors and trimethylsilyl triflate activation of trichloroacetimidate donors. Because all of the glycosidic linkages in **64** are composed of the 1,2-*trans*-linked relative configuration, anomeric stereocontrol was effected through liberal use of C2-ester protective groups that capitalize on neighbouring group participatory effects[81].

More recently, a completed synthesis of QS-21A$_{api}$ (**64**) was accomplished[82] (Fig. 5b) that relied on minimal use of C2-ester protective groups to avoid potential difficulties in late-stage selective ester deprotection in the presence of the hydrolytically labile C4-*O*-fucosyl ester side chain[78] in **64**. In this effort, five of the glycosidic bonds in **64** were constructed with the sulphoxide-mediated dehydrative glycosylation (Ph$_2$SO•Tf$_2$O) using hemiacetal donors[83]. It is also worth noting that recently developed oxidative glycosylation protocols were applied in the construction of advanced monosaccharide substrates. For example, I(III)-mediated oxidative bis(C1,C2-acyloxylation)[35] of tribenzylgalactal (**66**) served to install benzoate groups stereoselectively at both the C2 and C1 positions, with the latter being selectively removed by aminolysis to afford the hemiacetal **67**. Thus, neighbouring group participatory effects were bestowed on this donor (**67**) in the subsequent glycosylation, yet the need for excessive ester protective group exchanges in the late stages of the synthesis was minimized. Ph$_2$SO-mediated dehydrative glycosylation of the β-glucuronic acid (β-GlcA) derivative **68** provided disaccharide **69**, which was advanced to the branched trisaccharide trichloroacetimidate **70**. The stereoselective glycosylation of the quillaic acid triterpene (QA–OH) with trisaccharide **70** proved to be among
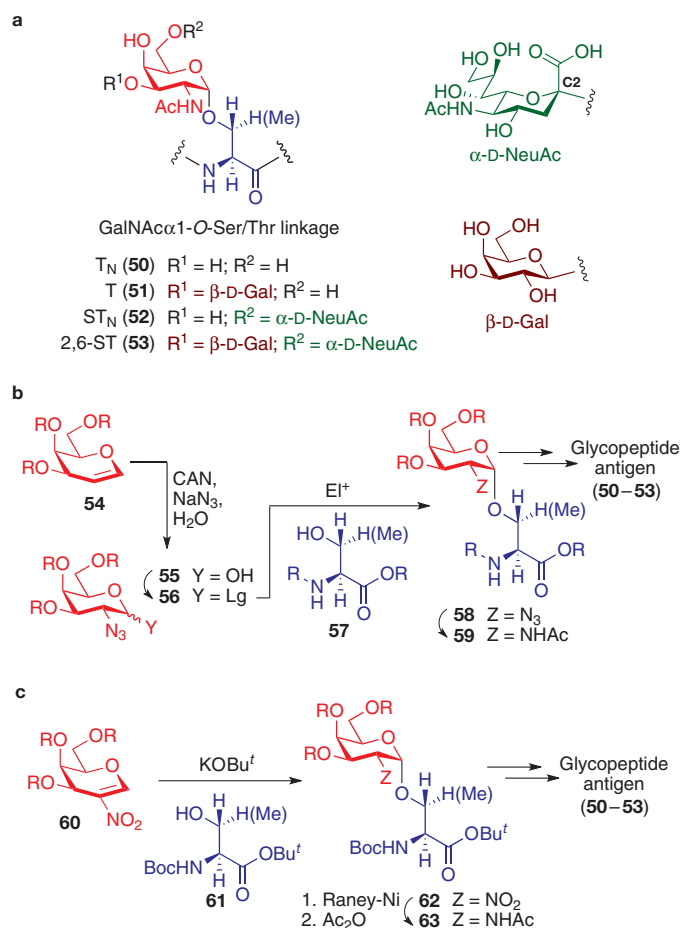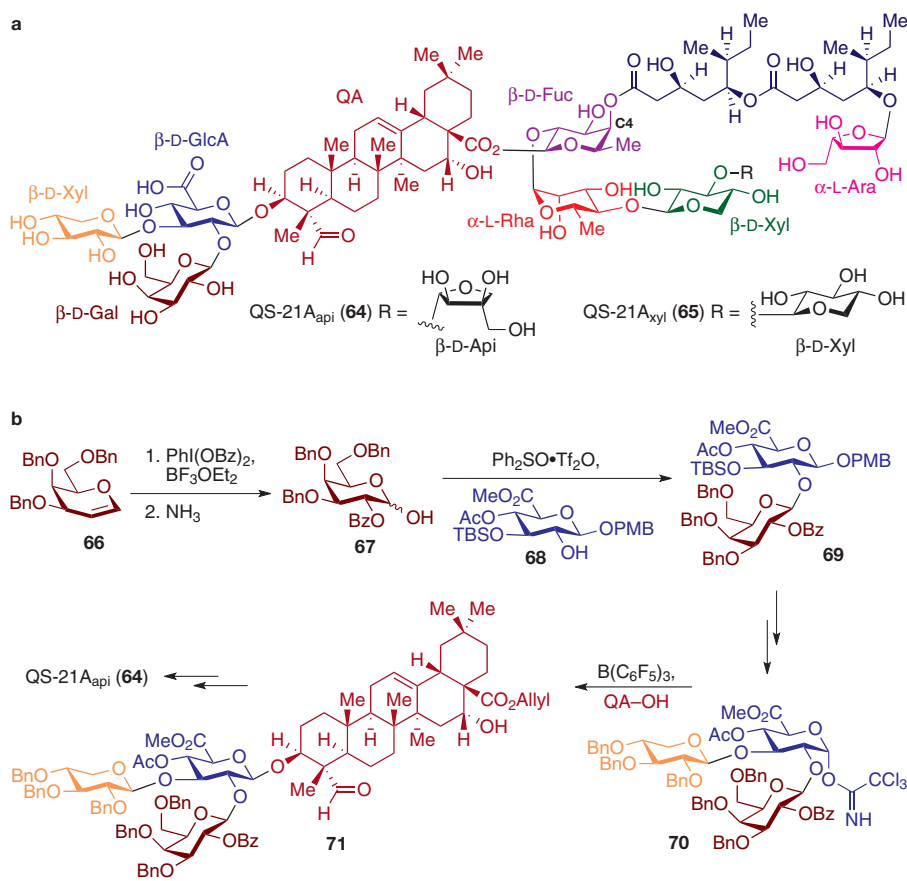
Figure 5 | Immunological adjuvant QS-21A. a, QS-21A is a heavily glycosylated saponin isolated from *Quillaja saponaria* Molina in both apiose (QS21A$_{api}$) and xylose (QS21A$_{xyl}$) forms. **b**, Synthesis of QS-21A$_{api}$. Glycosidic linkages were constructed by the use of

Ph$_2$SO•Tf$_2$O-promoted dehydrative glycosylation and glycosyl trichloroacetimidate coupling in the construction of the trisaccharide–triterpene substructure. Ara, arabinose; PMB, 4-methoxybenzyl; Rha, rhamnose; TBS, tert-butyldimethylsilyl.

the most difficult to secure, owing to the significant steric demands at the attachment sites of both coupling partners. Nevertheless, the desired β-glycosidic linkage in **71** was constructed through the use of the less common B(C$_6$F$_5$)$_3$ Lewis acid as the catalyst[84] in the trichloroacetimidate glycosylation. The carbohydrate–triterpene conjugate **71** was then advanced to the natural product **64** by way of late-stage conjugation to protected forms of the linear tetrasaccharide fragment and glycosylated acyl side chain, each of which was prepared by *de novo* multistep synthesis. *In vivo* immunological evaluation of synthetic QS-21A in vaccine formulations is currently in progress.

Conclusion

Advances in chemical glycosylation have allowed organic synthesis to fulfil its role as a supplier of rare carbohydrate antigens identified as potential targets for cancer immunotherapy[6,20]. Although several classes of antigen (for example, GM2, fucose-GM1, GD2 and GD3) can be obtained from natural sources in acceptable quantities, the acquisition of sufficient amounts to investigate others (such as Globo-H, ST$_N$, T$_N$ and T) still relies on chemical synthesis. However, chemical synthesis has evolved far beyond this initial role, as it also makes it possible to design novel versions of otherwise inaccessible antigen and adjuvant constructs, molecules that might hold the key to overcoming critical challenges in the induction of potent yet selective cellular and humoral responses.

Pursuit of non-natural structural modification of carbohydrate antigens that forcibly enhance the non-self identity of specific oligosaccharide conformations has shown promise, although guidelines for such designs are often empirically derived. For example, synthetic TACAs

in the form of sialic acid lactone derivatives of GD2 (ref. 15) and GD3 (ref. 14), or N-propionylated derivatives of polysialic acid[14,15,85], show potent serological response in immunized patients compared with their 'natural' unmodified counterparts. Moreover, as transformed cells have varying degrees of heterogeneity in the type and distribution of antigens on their surfaces[86], it is postulated that polyvalent constructs of various antigens could serve as a better mimic of cancer cells[19]. An approach to incorporate controlled epitope heterogeneity within homogeneous antigen constructs is exemplified by recently prepared pentavalent neoglycopeptides[87], whose KLH conjugates have shown encouraging antibody responses in preclinical evaluations in mice.

Further advances on these fronts will probably entail modulation of epitope selection or spacing in clustered constructs, as well as the development of novel molecular platforms for antigen display[88–90]. These efforts will rely not only on the development and application of new glycosylation technologies but also on innovations in chemoselective ligation reactions for access to increasingly complex and diverse immunogenic molecular arrays. However, chemical synthesis and modification of the immunological adjuvant component is a much less explored area in synthetic vaccine development[91,92]. Indeed, there have been only limited investigations regarding the chemical modification of QS-21A adjuvant on the basis of natural product degradation and derivatization[93,94], providing the first glimpses into its structure–activity relationship profile. The recent synthesis of QS-21A$_{api}$ will undoubtedly allow the role of chemical synthesis to extend beyond the arena of antigen construction to that of molecular adjuvant design in the preparation of new conjugate anticancer vaccines. ∎

1. Feizi, T. Demonstration by monoclonal antibodies that carbohydrate structures of glycoproteins and glycolipids are onco-developmental antigens. *Nature* **314,** 53–57 (1985).
2. Livingston, P. O. Approaches to augmenting the immunogenicity of melanoma gangliosides: from whole melanoma cells to ganglioside–KLH conjugate vaccines. *Immunol. Rev.* **145,** 147–166 (1995).
3. Hakomori, S. Tumor malignancy defined by aberrant glycosylation and sphingo(glyco)lipid metabolism. *Cancer Res.* **56,** 5309–5318 (1996).
4. Fung, P. Y., Madej, M., Koganty, R. R. & Longenecker, B. M. Active specific immunotherapy of a murine mammary adenocarcinoma using a synthetic tumor-associated glycoconjugate. *Cancer Res.* **50,** 4308–4314 (1990).
5. Slovin, S. F., Keding, S. J. & Ragupathi, G. Carbohydrate vaccines as immunotherapy for cancer. *Immunol. Cell Biol.* **83,** 418–428 (2005).
6. Ouerfelli, O., Warren, J. D., Wilson, R. M. & Danishefsky, S. J. Synthetic carbohydrate-based antitumor vaccines: challenges and opportunities. *Expert Rev. Vaccines* **4,** 677–685 (2005).
7. Helling, F. *et al.* GD3 vaccines for melanoma: superior immunogenicity of keyhole limpet hemocyanin conjugate vaccines. *Cancer Res.* **54,** 197–203 (1994).
8. Helling, F. *et al.* GM2–KLH conjugate vaccine: increased immunogenicity in melanoma patients after administration with immunological adjuvant QS-21. *Cancer Res.* **55,** 2783–2788 (1995).
9. Kensil, C. R., Patel, U., Lennick, M. & Marciani, D. Separation and characterization of saponins with adjuvant activity from *Quillaja saponaria* Molina cortex. *J. Immunol.* **146,** 431–437 (1991).
10. Ragupathi, G. Carbohydrate antigens as targets for active specific immunotherapy. *Cancer Immunol. Immunother.* **43,** 152–157 (1996).
11. Livingston, P. O. *et al.* Improved survival in stage III melanoma patients with GM2 antibodies: a randomized trial of adjuvant vaccination with GM2 ganglioside. *J. Clin. Oncol.* **12,** 1036–1044 (1994).
12. MacLean, G. D., Reddish, M. A., Koganty, R. R. & Longenecker, B. M. Antibodies against mucin-associated sialyl-Tn epitopes correlate with survival of metastatic adenocarcinoma patients undergoing active specific immunotherapy with synthetic STn vaccine. *J. Immunother. Emph. Tumor Immunol.* **19,** 59–68 (1996).
13. Slovin, S. F. *et al.* Carbohydrate vaccines in cancer: immunogenicity of a fully synthetic globo H hexasaccharide conjugate in man. *Proc. Natl Acad. Sci. USA* **96,** 5710–5715 (1999).
14. Ragupathi, G. *et al.* Induction of antibodies against GD3 ganglioside in melanoma patients by vaccination with GD3-lactone-KLH conjugate plus immunological adjuvant QS-21. *Int. J. Cancer* **85,** 659–666 (2000).
15. Ragupathi, G. *et al.* Consistent antibody response against ganglioside GD2 induced in patients with melanoma by a GD2 lactone-keyhole limpet hemocyanin conjugate vaccine plus immunological adjuvant QS-21. *Clin. Cancer Res.* **9,** 5214–5220 (2003).
16. Slovin, S. F. *et al.* Fully synthetic carbohydrate-based vaccines in biochemically relapsed prostate cancer: clinical trial results with α-*N*-acetylgalactosamine-*O*-serine/threonine conjugate vaccine. *J. Clin. Oncol.* **21,** 4292–4298 (2003).
17. Krug, L. M. *et al.* Vaccination of patients with small-cell lung cancer with synthetic fucosyl GM-1 conjugated to keyhole limpet hemocyanin. *Clin. Cancer Res.* **10,** 6094–6100 (2004).
18. Slovin, S. F. *et al.* Thomsen–Friedreich (TF) antigen as a target for prostate cancer vaccine: clinical trial results with TF cluster (c)-KLH plus QS21 conjugate vaccine in patients with biochemically relapsed prostate cancer. *Cancer Immunol. Immunother.* **54,** 694–702 (2005).
19. Ragupathi, G., Gathuru, J. & Livingston, P. Antibody inducing polyvalent cancer vaccines. *Cancer Treat. Res.* **123,** 157–180 (2005).
20. Livingston, P. O. & Ragupathi, G. Cancer vaccines targeting carbohydrate antigens. *Hum. Vaccin.* **2,** 137–143 (2006).
21. Toyokuni, T. & Singhal, A. K. Synthetic carbohydrate vaccines based on tumour-associated antigens. *Chem. Soc. Rev.* 231–242 (1995).
22. Danishefsky, S. J. & Allen, J. R. From the laboratory to the clinic: a retrospective on fully synthetic carbohydrate-based anticancer vaccines. *Angew. Chem. Int. Ed.* **39,** 836–863 (2000).
23. Brocke, C. & Kunz, H. Synthesis of tumor-associated glycopeptide antigens. *Bioorg. Med. Chem.* **10,** 3085–3112 (2002).
24. Buskas, T., Ingale, S. & Boons, G. J. Glycopeptides as versatile tools for glycobiology. *Glycobiology* **16,** 113R–136R (2006).
25. Paulsen, H. Advances in stereoselective chemical syntheses of complex oligosaccharides. *Angew. Chem. Int. Ed. Engl.* **21,** 155–224 (1982).
26. Sinay, P. Recent advances in glycosylation reactions. *Pure Appl. Chem.* **63,** 519–528 (1991).
27. Toshima, K. & Tatsuta, K. Recent progress in O-glycosylation methods and its application to natural products synthesis. *Chem. Rev.* **93,** 1503–1531 (1993).
28. Schmidt, R. R. & Kinzy, W. Anomeric-oxygen activation for glycoside synthesis: the trichloroacetimidate method. *Adv. Carbohydr. Chem. Biochem.* **50,** 21–123 (1994).
29. Garegg, P. J. Thioglycosides as glycosyl donors in oligosaccharide synthesis. *Adv. Carbohydr. Chem. Biochem.* **52,** 179–205 (1997).
30. Davis, B. G. Recent developments in oligosaccharide synthesis. *J. Chem. Soc. Perkin Trans. I* **1,** 2137–2160 (2000).
31. Hanessian, S. & Lou, B. Stereocontrolled glycosyl transfer reactions with unprotected glycosyl donors. *Chem. Rev.* **100,** 4443–4464 (2000).
32. Nicolaou, K. C. & Mitchell, H. J. Adventures in carbohydrate chemistry: new synthetic technologies, chemical synthesis, molecular design, and chemical biology. *Angew. Chem. Int. Ed.* **40,** 1576–1624 (2001).
33. Danishefsky, S. J. & Bilodeau, M. T. Glycals in organic synthesis: the evolution of comprehensive strategies for the assembly of oligosaccharides and glycoconjugates of biological consequence. *Angew. Chem. Int. Ed.* **35,** 1380–1419 (1996).
34. Di Bussolo, V., Kim, Y.-J. & Gin, D. Y. Direct oxidative glycosylations with glycal donors. *J. Am. Chem. Soc.* **120,** 13515–13516 (1998).
35. Shi, L., Kim, Y. J. & Gin, D. Y. C2-acyloxyglycosylation with glycal donors. *J. Am. Chem. Soc.* **123,** 6939–6940 (2001).
36. Lemieux, R. U. & Ratcliffe, R. M. The azidonitration of tri-*O*-acetyl-D-galactal. *Can. J. Chem.* **57,** 1244–1251 (1979).
37. Kan, C. *et al.* Photo amidoglycosylation of an allal azidoformate. Synthesis of β-2-amido allopyranosides. *Org. Lett.* **3,** 381–384 (2001).

38. Di Bussolo, V., Liu, J., Huffman, L. G. & Gin, D. Y. Acetamidoglycosylation with glycal donors: a one-pot glycosidic coupling with direct installation of the natural C(2)-*N*-acetylamino functionality. *Angew. Chem. Int. Ed.* **39,** 204–207 (2000).
39. Kannagi, R. *et al.* New globoseries glycosphingolipids in human teratocarcinoma reactive with the monoclonal antibody directed to a developmentally regulated antigen, stage-specific embryonic antigen 3. *J. Biol. Chem.* **258,** 8934–8942 (1983).
40. Bremer, E. G. *et al.* Characterization of a glycosphingolipid antigen defined by the monoclonal antibody MBr1 expressed in normal and neoplastic epithelial cells of human mammary gland. *J. Biol. Chem.* **259,** 14773–14777 (1984).
41. Hakomori, S. & Zhang, Y. Glycosphingolipid antigens and cancer therapy. *Chem. Biol.* **4,** 97–104 (1997).
42. Bilodeau, M. T. *et al.* Total synthesis of a human breast tumor associated antigen. *J. Am. Chem. Soc.* **117,** 7840–7841 (1995).
43. Park, T. K. *et al.* Total synthesis and proof of structure of a human breast tumor (globo-H) antigen. *J. Am. Chem. Soc.* **118,** 11488–11500 (1996).
44. Zhu, T. & Boons, G.-J. A two-directional and highly convergent approach for the synthesis of the tumor-associated antigen globo-H. *Angew. Chem. Int. Ed.* **38,** 3495–3497 (1999).
45. Douglas, N. L., Ley, S. V., Lucking, U. & Warriner, S. L. Tuning glycoside reactivity: new tool for efficient oligosaccharide synthesis. *J. Chem. Soc. Perkin Trans. I* **1,** 51-66 (1998).
46. Burkhart, F., Zhang, Z., Wacowich-Sgarbi, S. & Wong, C.-H. Synthesis of the Globo H hexasaccharide using the programmable reactivity-based one-pot strategy. *Angew. Chem. Int. Ed.* **40,** 1274–1277 (2001).
47. Zhang, Z. Programmable one-pot oligosaccharide synthesis. *J. Am. Chem. Soc.* **121,** 734–753 (1999).
48. Huang, C. Y. *et al.* Carbohydrate microarray for profiling the antibodies interacting with Globo H tumor antigens. *Proc. Natl Acad. Sci. USA* **103,** 15–20 (2006).
49. Lassaletta, J. & Schmidt, R. R. Glycosyl imidates. Part 75. Synthesis of the hexasaccharide moiety of globo H (human breast cancer) antigen. *Liebigs Ann.* 1417–1423 (1996).
50. Bosse, F., Marcaurelle, L. A. & Seeberger, P. H. Linear synthesis of the tumor-associated carbohydrate antigens Globo-H, SSEA-3, and Gb3. *J. Org. Chem.* **67,** 6659–6670 (2002).
51. Hakomori, S. Aberrant glycosylation in tumors and tumor-associated carbohydrate antigens. *Adv. Cancer Res.* **52,** 257–331 (1989).
52. Boons, G.-J. & Demchenko, A. V. Recent advances in *O*-sialylation. *Chem. Rev.* **100,** 4539–4565 (2000).
53. Haberman, J. M. & Gin, D. Y. Dehydrative sialylation with C2-hemiketal sialyl donors. *Org. Lett.* **5,** 2539–2541 (2003).
54. Castro-Palomino, J. C. *et al.* Efficient synthesis of ganglioside GM2 for use in cancer vaccines. *Angew. Chem. Int. Ed.* **36,** 1998–2001 (1997).
55. Gordon, J. E. & Turrell, G. C. Observations on *N*-methylacetonitrilium and *N*-phenylbenzonitrilium hexachloroantimonates. *J. Org. Chem.* **24,** 269–271 (1959).
56. Schmidt, R. R. & Ruecker, E. Stereoselective glycosidations of uronic acids. *Tetrahedron Lett.* **21,** 1421–1424 (1980).
57. Hasegawa, A. *et al.* Synthetic studies on sialoglycoconjugates 25: reactivity of glycosyl promoters in α-glycosylation of *N*-acetylneuraminic acid with the primary and secondary hydroxyl groups in the suitably protected galactose and lactose derivatives. *J. Carbohydr. Chem.* **10,** 493–498 (1991).
58. De Meo, C., Demchenko, A. V. & Boons, G. J. A stereoselective approach for the synthesis of α-sialosides. *J. Org. Chem.* **66,** 5490–5497 (2001).
59. Hollingsworth, M. A. & Swanson, B. J. Mucins in cancer: protection and control of the cell surface. *Nature Rev. Cancer* **4,** 45–60 (2004).
60. Hilkens, J., Ligtenberg, M. J., Vos, H. L. & Litvinov, S. V. Cell membrane-associated mucins and their adhesion-modulating property. *Trends Biochem. Sci.* **17,** 359–363 (1992).
61. Müller, S. & Hanisch, F. G. Recombinant MUC1 probe authentically reflects cell-specific *O*-glycosylation profiles of endogenous breast cancer mucin. High density and prevalent core 2-based glycosylation. *J. Biol. Chem.* **277,** 26103–26112 (2002).
62. Marcos, N. T. *et al.* Polypeptide GalNAc-transferases, ST6GalNAc-transferase I, and ST3Gal-transferase I expression in gastric carcinoma cell lines. *J. Histochem. Cytochem.* **51,** 761–771 (2003).
63. Banoub, J., Boullanger, P. & Lafont, D. Synthesis of oligosaccharides of 2-amino-2-deoxy sugars. *Chem. Rev.* **92,** 1167–1195 (1992).
64. Liebe, B. & Kunz, H. Solid-phase synthesis of a tumor-associated sialyl-T$_N$ antigen glycopeptide with a partial sequence of the 'tandem repeat' of the MUC-1 mucin. *Angew. Chem. Int. Ed.* **36,** 618–621 (1997).
65. Kunz, H. & Birnbach, S. Synthesis of tumor associated T$_N$- and T-antigen type *O*-glycopeptides and their conjugation to bovine serum albumin. *Angew. Chem. Int. Ed.* **98,** 360–362 (1986).
66. Sames, D., Chen, X. T. & Danishefsky, S. J. Convergent total synthesis of a tumour-associated mucin motif. *Nature* **389,** 587–591 (1997).
67. Paulsen, H., Rauwald, W. & Weichert, U. Building units of oligosaccharides. LXXXVI. Glycosidation of oligosaccharide thioglycosides to *O*-glycoprotein segments. *Liebigs Ann.* 75–86 (1988).
68. George, S. K. *et al.* Chemoenzymatic synthesis of sialylated glycopeptides derived from mucins and T-cell stimulating peptides. *J. Am. Chem. Soc.* **123,** 11117–11125 (2001).
69. Svarovsky, S. A. & Barchi, J. J. Highly efficient preparation of tumor antigen-containing glycopeptide building blocks from novel pentenyl glycosides. *Carbohydr. Res.* **338,** 1925–1935 (2003).
70. Rademann, J. & Schmidt, R. R. Solid-phase synthesis of a glycosylated hexapeptide of human sialophorin, using the trichloroacetimidate method. *Carbohydr. Res.* **269,** 217–225 (1995).
71. Paulsen, H., Peters, S., Bielfeldt, T., Meldal, M. & Bock, K. Synthesis of the glycosyl amino acids Nᵅ-Fmoc-Ser[Ac$_4$-β-D-Gal *p*-(1 → 3)-Ac$_2$-α-D-GalN$_3$ *p*]-OPfp and Nᵅ-Fmoc-Thr[Ac$_4$-β-D-Gal *p*-(1 → 3)-Ac$_2$-α-D-GalN$_3$ *p*]-OPfp and the application in the solid-phase peptide synthesis of multiply glycosylated mucin peptides with Tⁿ and T antigenic structures. *Carbohydr. Res.* **268,** 17–34 (1995).
72. Nakahara, Y., Nakahara, Y. & Ogawa, T. Solid-phase synthesis of an *O*-linked glycopeptide based on a benzyl-protected glycan approach. *Carbohydr. Res.* **292,** 71–81 (1996).
73. Winterfeld, G. A., Khodair, A. I. & Schmidt, R. R. O-glycosyl amino acids by 2-nitrogalactal concatenation — synthesis of a mucin-type *O*-glycan. *Eur. J. Org. Chem.* 1009–1021 (2003).

74. Elofsson, M., Salvador, L. A. & Kihlberg, J. Preparation of Tn and sialyl Tn building blocks used in Fmoc solid-phase synthesis of glycopeptide fragments from HIV gp120. *Tetrahedron* **53,** 369–390 (1997).

75. Seitz, O. & Kunz, H. A novel allylic anchor for solid-phase synthesis — synthesis of protected and unprotected O-glycosylated mucin-type glycopeptides. *Angew. Chem. Int. Ed.* **34,** 803–805 (1995).

76. Dziadek, S., Hobel, A., Schmitt, E. & Kunz, H. A fully synthetic vaccine consisting of a tumor-associated glycopeptide antigen and a T-cell epitope for the induction of a highly specific humoral immune response. *Angew. Chem. Int. Ed.* **44,** 7630–7635 (2005).

77. Dziadek, S., Brocke, C. & Kunz, H. Biomimetic synthesis of the tumor-associated (2,3)-sialyl-T antigen and its incorporation into glycopeptide antigens from the mucins MUC1 and MUC4. *Chem. Eur. J.* **10,** 4150–4162 (2004).

78. Kensil, C. R. Saponins as vaccine adjuvants. *Crit. Rev. Ther. Drug Carr. Syst.* **13,** 1–55 (1996).

79. Jacobsen, N. E. *et al.* Structure of the saponin adjuvant QS-21 and its base-catalyzed isomerization product by ^{1}H and natural abundance ^{13}C NMR spectroscopy. *Carbohydr. Res.* **280,** 1–14 (1996).

80. Zhu, X., Yu, B., Hui, Y. & Schmidt, R. R. Synthesis of the trisaccharide and tetrasaccharide moieties of the potent immunoadjuvant QS-21. *Eur. J. Org. Chem.* 965–973 (2004).

81. Nukada, T., Berces, A., Zgierski, M. Z. & Whitfield, D. M. Exploring the mechanism of neighboring group assisted glycosylation reactions. *J. Am. Chem. Soc.* **120,** 13291–13295 (1998).

82. Kim, Y.-J. *et al.* Synthetic studies of complex immunostimulants from *Quillaja saponaria*: synthesis of the potent clinical immunoadjuvant QS-21A$_{api}$. *J. Am. Chem. Soc.* **128,** 11906–11915 (2006).

83. Garcia, B. A. & Gin, D. Y. Dehydrative glycosylation with activated diphenyl sulfonium reagents. Scope, mode of C(1)-hemiacetal activation, and detection of reactive glycosyl intermediates. *J. Am. Chem. Soc.* **122,** 4269–4279 (2000).

84. Ishihara, K. & Yamamoto, H. Arylboron compounds as acid catalysts in organic synthetic transformations. *Eur. J. Org. Chem.* 527–538 (1999).

85. Nores, G. A., Dohi, T., Taniguchi, M. & Hakomori, S. Density-dependent recognition of cell surface GM3 by a certain anti-melanoma antibody, and GM3 lactone as a possible immunogen: requirements for tumor-associated antigen and immunogen. *J. Immunol.* **139,** 3171–3176 (1987).

86. Zhang, S. *et al.* Selection of tumor antigens as targets for immune attack using immunohistochemistry: II. Blood group-related antigens. *Int. J. Cancer* **73,** 50–56 (1997).

87. Ragupathi, G. *et al.* Preparation and evaluation of unimolecular pentavalent and hexavalent antigenic constructs targeting prostate and breast cancer: a synthetic route to anticancer vaccine candidates. *J. Am. Chem. Soc.* **128,** 2715–2725 (2006).

88. Lo-Man, R. *et al.* A fully synthetic therapeutic vaccine candidate targeting carcinoma-associated Tn carbohydrate antigen induces tumor-specific antibodies in nonhuman primates. *Cancer Res.* **64,** 4987–4994 (2004).

89. Grigalevicius, S. *et al.* Chemoselective assembly and immunological evaluation of multiepitopic glycoconjugates bearing clustered Tn antigen as synthetic anticancer vaccines. *Bioconjug. Chem.* **16,** 1149–1159 (2005).

90. Svarovsky, S. A., Szekely, Z. & Barchi, J. J. Synthesis of gold nanoparticles bearing the Thomsen–Friedenreich disaccharide: a new multivalent presentation of an important tumor antigen. *Tetrahedron Asymmetry* **16,** 587–598 (2005).

91. Jiang, Z. H. & Koganty, R. R. Synthetic vaccines: the role of adjuvants in immune targeting. *Curr. Med. Chem.* **10,** 1423–1439 (2003).

92. Buskas, T., Ingale, S. & Boons, G. J. Towards a fully synthetic carbohydrate-based anticancer vaccine: synthesis and immunological evaluation of a lipidated glycopeptide containing the tumor-associated tn antigen. *Angew. Chem. Int. Ed.* **44,** 5985–5988 (2005).

93. Marciani, D. J. *et al.* Development of semisynthetic triterpenoid saponin derivatives with immune stimulating activity. *Vaccine* **18,** 3141–3151 (2000).

94. Marciani, D. J. Vaccine adjuvants: role and mechanisms of action in vaccine immunogenicity. *Drug Discov. Today* **8,** 934–943 (2003).

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence should be addressed to the authors (gind@mskcc.org; danica_galonic@hms.harvard.edu).

Unusual sugar biosynthesis and natural product glycodiversification

Christopher J. Thibodeaux[1], Charles E. Melançon[2] & Hung-wen Liu[1,2,3]

The enzymes involved in the biosynthesis of carbohydrates and the attachment of sugar units to biological acceptor molecules catalyse an array of chemical transformations and coupling reactions. In prokaryotes, both common sugar precursors and their enzymatically modified derivatives often become substituents of biologically active natural products through the action of glycosyltransferases. Recently, researchers have begun to harness the power of these biological catalysts to alter the sugar structures and glycosylation patterns of natural products both *in vivo* and *in vitro*. Biochemical and structural studies of sugar biosynthetic enzymes and glycosyltransferases, coupled with advances in bioengineering methodology, have ushered in a new era of drug development.

Given the numerous physiological roles of carbohydrates, which encompass many aspects of normal cellular function and survival, it is not surprising that both monosaccharide and polysaccharide structures vary markedly throughout nature, providing the chemical diversity required to fulfil a vast range of biological tasks. Two general strategies are used by nature to expand the diversity of carbohydrate structure and function, a process referred to as glycodiversification. In one strategy, an unusual sugar structure can be generated through enzymatic modification of the functional groups of a common sugar precursor. This mode of diversification exists in all living organisms but is most elaborate in prokaryotes, which manufacture highly modified sugars as building blocks for cell wall and secondary metabolite biosynthesis. The second strategy is to assemble different glycoforms by altering the attachment of activated sugar substrates to 'acceptor' molecules such as natural product aglycones, proteins or other sugars — a feat performed by glycosyltransferases. In this case, the identity of the sugar, the number of sugar moieties, or the regiochemistry or stereochemistry of sugar attachment to the acceptor molecule can all be varied to generate structural diversity. Thus, the abundance of natural sugar structures is ultimately derived from the activities of the enzymes responsible for sugar biosynthesis and attachment.

In this review we summarize some of the strategies used by sugar biosynthetic enzymes to generate these myriad carbohydrate structures. Special emphasis is given to the enzymes of bacterial deoxysugar biosynthesis, both because these sugars exhibit abundant structural modifications not seen in eukaryotic sugars and because many of these sugars become appendages on biologically active natural products[1,2]. We also discuss the structure and mechanisms of glycosyltransferases, which ultimately determine the feasibility of natural product enzymatic glycodiversification. Finally, we highlight recent progress in the fields of combinatorial biosynthesis and chemoenzymatic/enzymatic synthesis, and demonstrate how these bioengineering methods have been successfully used to alter the sugar components of natural products.

Monosaccharide activation

In order to be used by the biosynthetic enzymes and glycosyltransferases within a cell, monosaccharides must first be activated as either nucleoside monophosphate (NMP) or nucleoside diphosphate (NDP) derivatives. Most NDP-sugars (such as NDP-α-D-glucose, **3**, Fig. 1) originate from primary metabolic intermediates such as fructose-6-phosphate and glucose-6-phosphate (**1**). Activation typically involves the coupling of a sugar-1-phosphate (**2**) to an NMP moiety from the corresponding nucleoside triphosphate (NTP) by a nucleotidylyltransferase[3]. The sugar-1-phosphate is either the product of an anomeric kinase reaction or is generated from a sugar-6-phosphate by an appropriate mutase. Interestingly, only ten NDP-sugars are used as the basic construction units for most biosynthetic glycosylation reactions in eukaryotic cells[4].

In contrast to eukaryotes, the structures of prokaryotic NDP-sugars are much more diverse (Fig. 1). Because most unusual sugars are 6-deoxyhexoses[5], the immediate biosynthetic precursor of most deoxysugars is an NDP-4-keto-6-deoxy-D-hexose (usually NDP-4-keto-6-deoxy-α-D-glucose, **4**). From this common intermediate, different combinations of oxidation, reduction, deoxygenation, epimerization, isomerization, group-transfer and rearrangement reactions yield the vast array of diverse NDP-sugar structures observed across the prokaryotic kingdom. In most cases, the enzymes involved in these transformations have only recently been identified and mechanistically characterized[5–7]. Importantly, there is often a high degree of amino-acid sequence similarity among enzymes catalysing similar reactions in separate deoxysugar biosynthetic pathways. This has facilitated the identification and characterization of deoxysugar biosynthetic loci, and has also enabled the pathway engineering studies discussed below.

Diverse mechanisms for unusual sugar biosynthesis

Although many sugar biosynthetic enzymes have been studied, those involved in deoxygenation are among the most mechanistically diverse and intriguing known[6–9], and thus deserve special attention. The specific C–O bond cleavage strategy used by each type of deoxygenating enzyme in the biosynthesis of deoxyhexoses is determined by the location of the scissile C–O bond (either α or β) relative to an essential 4-keto group on the sugar ring[9]. As depicted in Fig. 2a, C6 deoxygenation is catalysed by an NDP-hexose-4,6-dehydratase (E_{od}), which belongs to

the short-chain dehydrogenase/reductase (SDR) family and relies on a tightly bound NAD^+ cofactor to oxidize the hydroxyl group at C4 of NDP-hexose (such as NDP-α-D-glucose, 3) to give a 4-ketosugar intermediate (5). The C4 oxidation activates the C5 proton for abstraction, allowing elimination of a water molecule across the C5–C6 bond to form 6. Hydride transfer from the transient NADH to C6 of 6 leads to the NDP-4-keto-6-deoxy-α-D-glucose product (4). The installation of a keto group at C4 provides the necessary ring activation for subsequent deoxygenation at C2, C3 and C4, as well as many other transformations in sugar biosynthesis.

A similar reaction sequence (namely, α-H^+ abstraction followed by β-C–O bond cleavage) occurs in the C2 deoxygenation catalysed by TylX3 in the TDP-L-mycarose (7, Fig. 1) biosynthetic pathway of *Streptomyces fradiae* (Fig. 2b). Here, the 4-keto group of 8 helps lower the pK_a of the C3 proton, facilitating deprotonation at C3 and the elimination of the C2 hydroxyl group to afford 10. Zn^{2+}, which is required for catalysis by TylX3, may have a role in C3 deprotonation by activating an active-site-bound water molecule as a base. Alternatively, it may stabilize the enediolate intermediate (9) or act as a general acid in the elimination of the C2 hydroxyl group (9 → 10). Tautomerization of 10 yields the unstable 3,4-diketone 11, which is reduced by the next enzyme in the pathway, the NADPH-dependent 3-ketoreductase, TylC1, to give the 2,6-dideoxyhexose product 12.

The mechanism of C3 deoxygenation in the biosynthesis of 3,6-dideoxysugars (such as 13, Fig. 1) is more complex than C2 and C6 deoxygenation because the scissile C–O bond is adjacent to the 4-keto group. Thus, C–O bond cleavage cannot be triggered by an initial α-H^+ abstraction event. This reaction is catalysed by the combined action of two enzymes, a 3-dehydrase (E_1) and a reductase (E_3), originally isolated from the CDP-L-ascarylose (13) biosynthetic pathway of *Yersinia pseudotuberculosis* (Fig. 2c). Catalysis is initiated by the formation of a Schiff base between the 4-keto group of 14 and the pyridoxamine 5′-phosphate (PMP) cofactor of E_1, followed by deprotonation at C4′ and 1,4-dehydration to give 15 as the intermediate. Subsequent reduction catalysed by E_3 through two one-electron transfers from NADH to generate a radical intermediate 16 is mediated by a relay system consisting of the FAD and [2Fe–2S] centre in E_3 and a [2Fe–2S] centre in E_1. Hydrolysis of the reduced species 17 gives the 3,6-dideoxyhexose product 18 and regenerates PMP. The elaborate and complex reaction performed by E_1/E_3 and its homologues in other unusual sugar pathways is unprecedented in enzyme catalysis.

Recently, the mechanism of C4 deoxygenation catalysed by the enzymes DesI and DesII during the biosynthesis of TDP-D-desosamine (19, Fig. 1) from *Streptomyces venezuelae* has been explored. The aminotransferase DesI catalyses the conversion of 8 to TDP-4-amino-4,6-α-D-dideoxyglucose, whereas DesII, a member of the radical *S*-adenosyl-L-methionine (SAM) superfamily, is responsible for the subsequent oxidative deamination reaction. Although the DesII reaction is probably initiated by C3 hydrogen abstraction, its mechanistic details remain obscure[10]. Cumulatively, these examples demonstrate the elegance and versatility of nature's sugar biosynthetic machinery. The fact that several mechanisms have evolved for C–O bond cleavage, using enzymes that accept similar substrates but require distinct sets of cofactors, underscores the importance of these modifications in the natural amplification of carbohydrate structural diversity.

Evolution in unusual sugar biosynthesis

We have noticed that many of the enzymes involved in unusual sugar biosynthesis (including E_{od}, Fig. 2a) are members of the short-chain dehydrogenase/reductase (SDR) family[11]. Enzymes in this family use a conserved protein fold to bind a nicotinamide adenine dinucleotide cofactor (either NAD^+ or $NADP^+$, denoted $NAD(P)^+$) that has an essential role in catalysis. Interestingly, the nicotinamide cofactor is most often used as a true prosthetic group by the sugar biosynthetic enzymes of the SDR family, rather than as a co-substrate as is typically seen in most NAD(P)-dependent enzymes. Catalysis by these enzymes is always initiated by the oxidation of one of the sugar hydroxyl groups by the $NAD(P)^+$ prosthetic group. After the requisite reaction chemistry, the oxidized nicotinamide coenzyme is regenerated at the end of each catalytic cycle (Fig. 3).

UDP-D-galactose-4-epimerase from *Escherichia coli*[12,13], CDP-D-tyvelose-2-epimerase from *Y. pseudotuberculosis*[14] and ADP-β-L-*glycero*-D-*manno*-heptose-6-epimerase from *E. coli*[15] each catalyse epimerization of a hydroxyl group at an unactivated position of a hexose, and all three are members of the SDR superfamily (Fig. 3a). After oxidation of the target hydroxyl
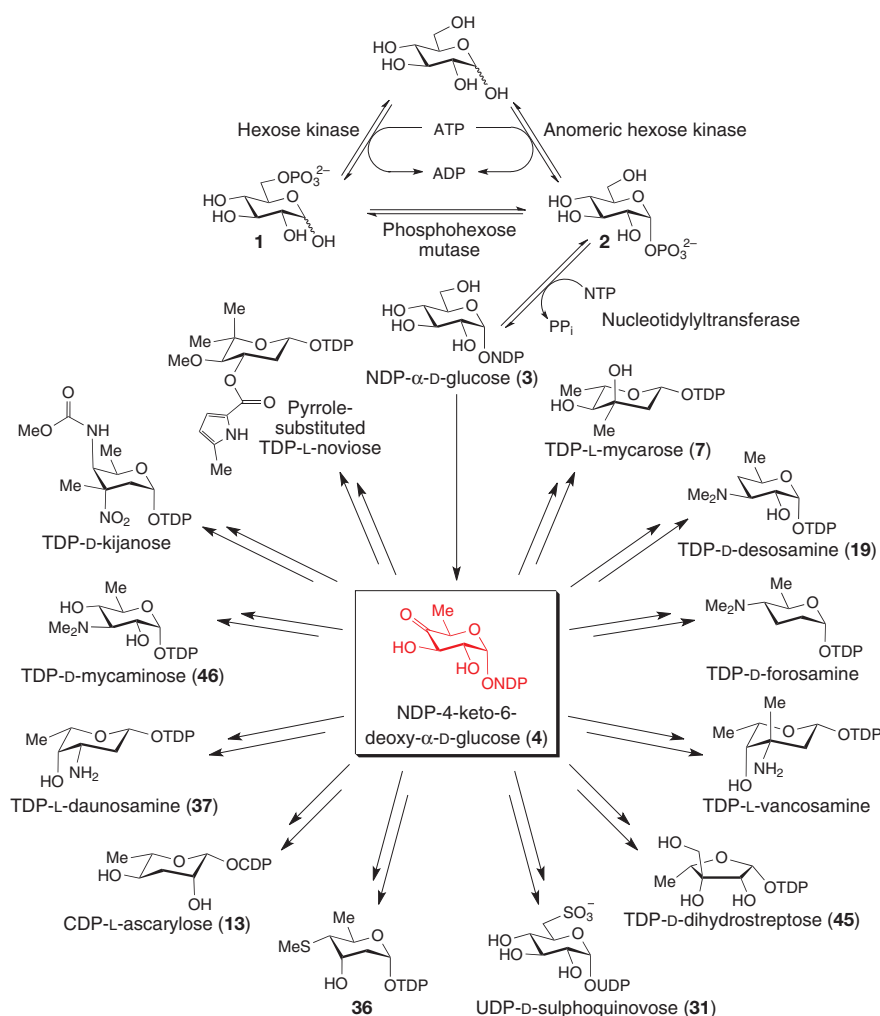

Figure 1 | Structural diversity of prokaryotic NDP-sugars derived from NDP-4-keto-6-deoxy-α-D-glucose. NDP-α-D-glucose (3) is generated from common metabolites through the action of hexose kinases, phosphohexose mutases and nucleotidylyltransferases. Most prokaryotic deoxysugars are 6-deoxyhexoses and are synthesized from 3 via an NDP-4-keto-6-deoxy-α-D-glucose intermediate (4), which is the product of an NDP-glucose-4,6-dehydratase reaction — the committed step in 6-deoxysugar biosynthesis. PP_i, pyrophosphate.
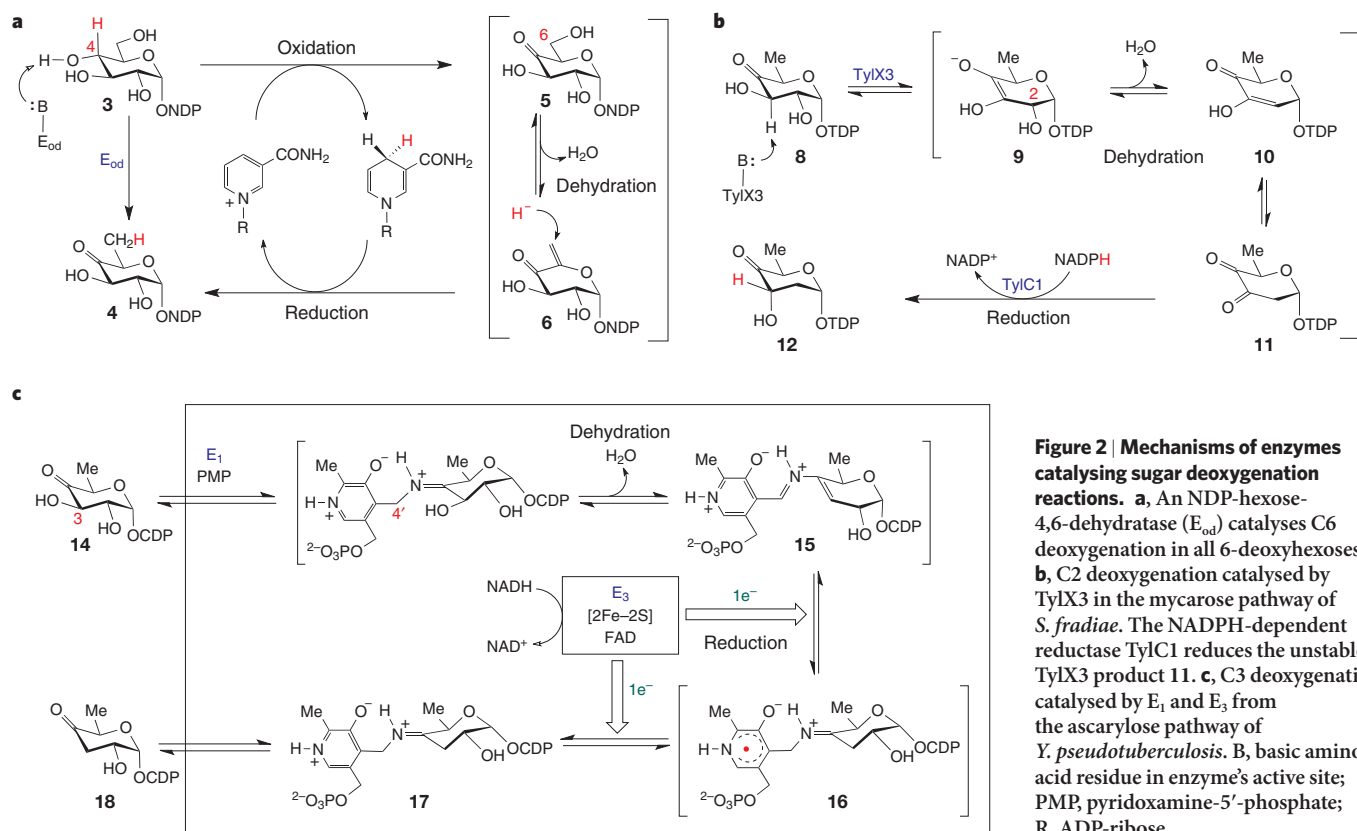
Figure 2 | Mechanisms of enzymes catalysing sugar deoxygenation reactions. a, An NDP-hexose-4,6-dehydratase (E_{od}) catalyses C6 deoxygenation in all 6-deoxyhexoses. **b**, C2 deoxygenation catalysed by TylX3 in the mycarose pathway of *S. fradiae*. The NADPH-dependent reductase TylC1 reduces the unstable TylX3 product **11**. **c**, C3 deoxygenation catalysed by E_1 and E_3 from the ascarylose pathway of *Y. pseudotuberculosis*. B, basic amino-acid residue in enzyme's active site; PMP, pyridoxamine-5′-phosphate; R, ADP-ribose.

group by $NAD(P)^+$, the ketosugar intermediate of each respective reaction (**20**, **21** or **22**) rotates in the enzyme's active site such that the opposite face of the sugar keto group becomes exposed to the transient $NAD(P)H$. Internal hydride return from $NAD(P)H$ reduces the keto group and completes the isomerization. In the reaction catalysed by GDP-D-mannose-3,5-epimerase/reductase from *Arabidopsis thaliana*, oxidation of the C4 hydroxyl facilitates C5 epimerization through active site acid/base chemistry[16] (**23** → **24** → **25**, Fig. 3b). The epimerized sugar can then be reduced at C4 to give GDP-β-L-gulose (**26**), or it can be further epimerized at C3 (**25** → **27** → **28** → **29**, Fig. 3b) before C4 reduction to give GDP-β-L-galactose (**30**).

The reaction catalysed by another SDR enzyme, UDP-α-D-sulpho-quinovose synthase (SQD1) from *A. thaliana*[17], involves the formation of an α,β-unsaturated intermediate similar to **6** in the reaction catalysed by E_{od} (Fig. 2a). However, instead of a nucleophilic hydride attack on C6 of **6**, *in vitro* studies have shown that SQD1 uses sulphite as the nucleophile[18]. After sulphite addition, NADH reduction of the 4-keto group completes the catalytic cycle to give **31** (Fig. 1). As a final demonstration of the versatility of this class of sugar biosynthetic enzyme, UDP-glucuronic acid decarboxylase has been shown to catalyse the decarboxylation of UDP-glucuronic acid (**32**) to give UDP-xylose (**33**, Fig. 3c). Once again, oxidation of the C4 hydroxyl group precedes decarboxylation and facilitates the formation of a transient enolate (**34**). Stereospecific protonation at C5 (**34** → **35**) followed by internal hydride return at C4 gives the final product (**33**)[19].

Cumulatively, the SDR family enzymes illustrate a powerful evolutionary strategy conceived by nature for the synthesis of diverse carbohydrate structures. These enzymes use a common protein fold and have similar cofactor requirements to perform various chemical transformations. Because of the central role of SDR enzymes in carbohydrate biosynthesis, and the availability of X-ray crystal structures of several of these enzymes[11,13,16,17], this family of catalysts is an attractive target for structure-based protein-engineering efforts to construct sugar-modifying enzymes with customized activities.

Structures and mechanisms of glycosyltransferases

Despite the impressive variety of carbohydrate structures found in nature, most of the glycosyltransferases responsible for the attachment of NDP-sugars to their acceptor molecules fall into one of two structurally distinct families[20–22]. In the glycosyltransferase-A (GT-A) family, the overall enzyme topology consists of separate, but closely associated, NDP-sugar-binding and acceptor-binding domains[23–25]. The NDP-sugar-binding domain generally contains a conserved Asp-X-Asp amino-acid motif that binds a divalent metal ion (either Mg^{2+} or Mn^{2+}); however, some GT-A enzymes do not require this metal ion for activity[26,27]. When present, the metal ion coordinates the diphosphate moiety of the NDP-sugar and electrostatically stabilizes the developing negative charge on the NDP leaving group during catalysis[28]. The GT-A family includes most of the eukaryotic glycosyltransferases responsible for glycan biosynthesis in the Golgi and endoplasmic reticulum, as well as many prokaryotic glycosyltransferases[29]. Members of the GT-B family contain two Rossmann-type domains separated by a deep cleft[25,29] and, unlike their counterparts in the GT-A family, the GT-B enzymes characterized so far do not require a metal ion for NDP-sugar binding and catalysis[29]. Members of the GT-B family are also dispersed throughout nature, but include most prokaryotic glycosyltransferases that glycosylate secondary metabolites and the O-linked β-*N*-acetylglucosamine (O-GlcNAc) transferase (OGT), which is involved in the glycosylation of cytosolic and nuclear proteins in eukaryotes[29] (see page 1017).

Mechanistically, glycosyltransferases can be classified as either inverting or retaining, depending on the overall stereochemical course of the reaction they catalyse[30] (Fig. 4). The stereochemistry of the transfer reaction does not correlate directly with a specific glycosyltransferase structural family, because both the GT-A and GT-B families include inverting and retaining glycosyltransferases[25,31]. Structural[32–34] and mechanistic[28,35] data, combined with *ab initio* molecular orbital calculations[36], support an in-line S_N2 attack mechanism for inverting glycosyltransferases (Fig. 4a).

The mechanism of retaining glycosyltransferases is still ambiguous. It was originally thought to involve a double displacement reaction analogous to that of retaining glycosidases[37,38] (Fig. 4b). However, structural[39–42] and biochemical[43,44] studies of several retaining glycosyltransferases in the presence of substrate analogues failed to identify an active-site nucleophile that could be involved in forming the enzyme–sugar adduct. This led to the proposal of a front-side S_N2 (or S_Ni) attack mechanism involving a late oxocarbenium-like transition state[39], in which the new glycosidic linkage between the acceptor and donor is formed on the same side of the sugar ring as the scissile glycosidic linkage to the NDP leaving group. Recently, however, Lairson et al. identified a covalent attachment between sugar and enzyme in a mutant α1,4-galactosyltransferase C (LgtC) from Neisseria meningitidis[37], rekindling the mechanistic debate about this important class of enzyme.

In addition to the canonical glycosyltransferase-catalysed reaction involving sugar transfer from an NDP-sugar to the hydroxyl group of an aglycone acceptor, several unusual modes of glycosylation exist in nature. It has recently been demonstrated that the glycosyltransferases of certain macrolide and anthracycline antibiotics require an auxiliary or accessory protein for efficient glycosyl transfer[45–48]. Although the exact biochemical function of this protein is not yet clear, it seems to be involved in activating its partner glycosyltransferase for catalysis, perhaps either by helping the glycosyltransferase fold into an active conformation or by inducing a conformational change that results in activation. In addition to the ubiquitous O-glycosides, N-, C- and S-glycosides have also been found in nature. Perhaps the best-known example of N-glycosylation is the reaction catalysed by the heteromeric, multisubunit oligosaccharyltransferase (OST) complex that initiates N-glycan biosynthesis in eukaryotes. In this case, an oligosaccharide core is transferred from a dolichol-pyrophosphate donor to the asparagine residue of an Asn-X-Ser/Thr-X motif (where X is not Pro) of the target protein substrate in the endoplasmic reticulum[49]. Despite its ubiquity in eukaryotic cells, the catalytic mechanism and the contributions of the individual OST subunits to substrate recognition and catalysis remain largely unexplored[49].

Although N-glycosides are much less common in prokaryotes, several N-glycosyltransferases involved in the biosynthesis of indolocarbozoles[50–52] and the single-subunit OST from Campylobacter jejuni[53] have recently been identified and characterized. C-glycosides have been found in several natural aromatic products from prokaryotes and plants[54], and the mechanism of C-glycosylation has been proposed to involve nucleophilic attack by an activated aromatic ring on the anomeric carbon of the NDP-sugar[55] (Fig. 4c). Although several prokaryotic natural products contain thiosugars (such as 36, Fig. 1), authentic
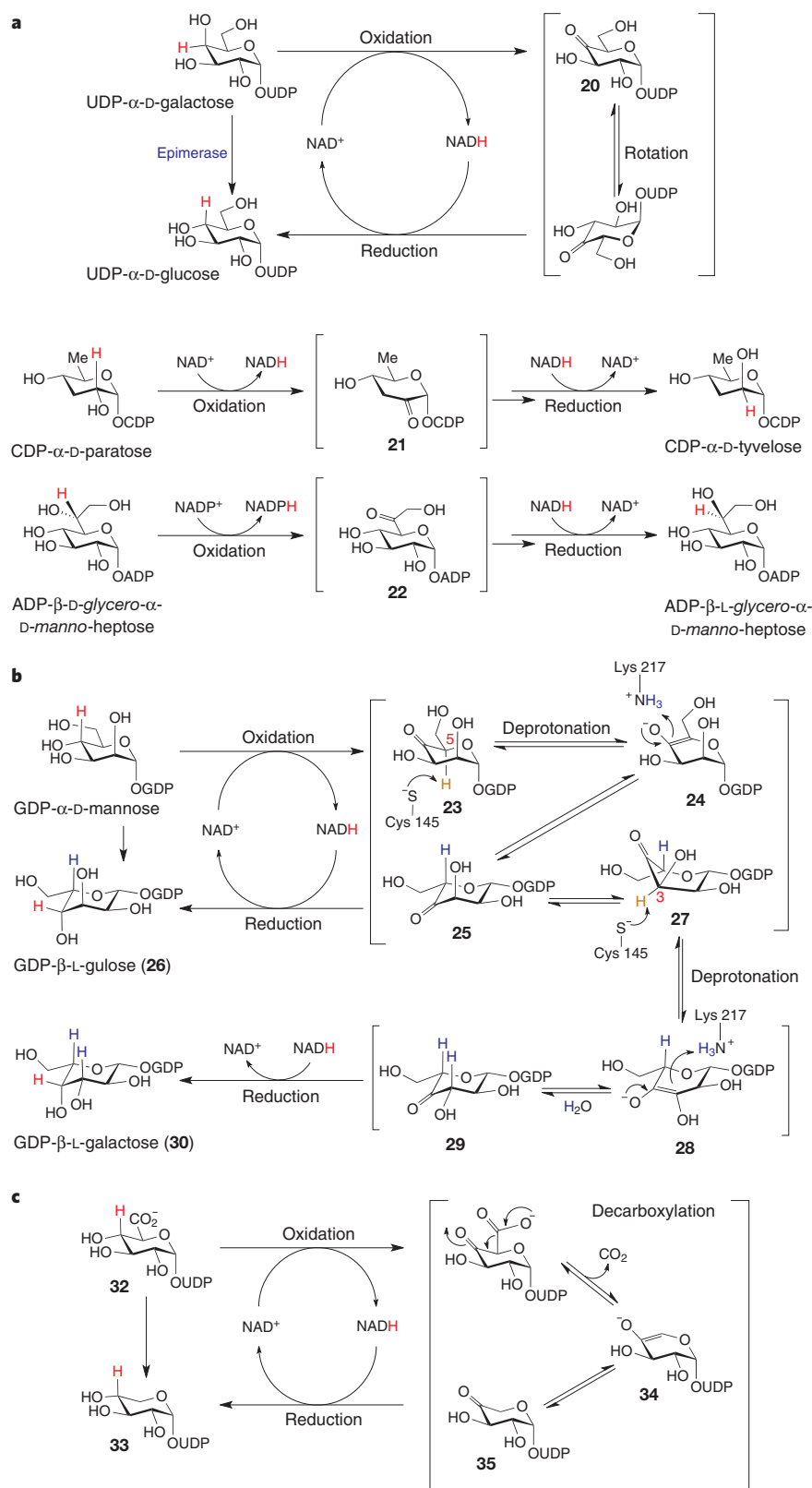
Figure 3 | Examples of sugar biosynthetic enzymes from the short-chain dehydrogenase/reductase enzyme family. a, Epimerization at unactivated carbon centres catalysed by UDP-α-D-galactose-4-epimerase, CDP-α-D-tyvelose-2-epimerase and ADP-β-L-glycero-D-manno-heptose-6-epimerase. Each involves oxidation of the hydroxyl group at the chiral centre to be epimerized, followed by rotation of the keto intermediate in the active site and reduction from the opposite face to complete the epimerization. **b,** In the mechanism of GDP-α-D-mannose-3,5-epimerase/reductase, oxidation to the keto intermediate followed by C5 epimerization and C4 reduction yields GDP-β-L-gulose (23 → 24 → 25 → 26). Intermediate 25 can also be epimerized at C3 and reduced at C4 to yield GDP-β-L-galactose (25 → 27→ 28 → 29 → 30). **c,** Reaction catalysed by UDP-glucuronic acid decarboxylase.
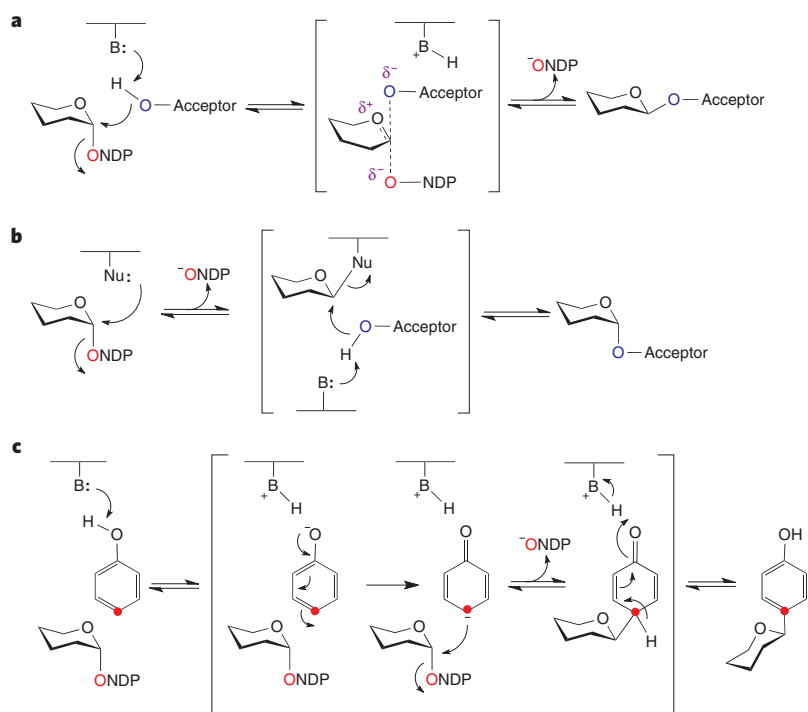
Figure 4 | **Mechanism of glycosyltransferases. a**, For inverting glycosyltransferases, general base (B)-catalysed in-line S_N2 attack at C1 of the NDP-sugar donor by the acceptor leads to inversion of stereochemistry at the anomeric carbon. **b**, The double displacement mechanism for retaining glycosyltransferases involves the S_N2 attack of an active site nucleophile (Nu) at C1 of the NDP-sugar to form a covalent sugar–enzyme adduct, followed by an S_N2 attack of the acceptor at C1 to form the glycosidic linkage with retention of anomeric configuration. **c**, The putative mechanism for C-glycosylation in aryl-*C*-glycosides involves aromatic substitution facilitated by an *ortho*/*para*-directing hydroxyl group.

S-glycosyltransferases have not yet been identified in these organisms. To the best of our knowledge, UGT74B1, which is involved in glucosinolate biosynthesis in *A. thaliana*[56], is the only *S*-glycosyltransferase to have been characterized *in vitro*[57].

Bioengineering glycosylated natural products

Not all biologically active natural products are suitable for clinical use, owing to problems with solubility, bioavailability, toxicity or ineffectiveness against resistant organisms, making preparation of their analogues important in drug-discovery efforts. As a result of their structural complexity, many natural products are challenging to prepare or modify through purely synthetic methods. Because of the difficulties often encountered in total synthesis, re-engineering of the existing biosynthetic machinery offers an alternative approach for the creation of natural product analogues with several distinct advantages: access to chemical derivative structures not easily obtained by synthetic means, ease of compound production, reduced production of toxic by-products, scalability of the process, and the potential for performing combinatorial biosynthesis.

The sequences of more than 100 gene clusters involved in the production of glycosylated natural products are known, and the activities of many sugar biosynthetic enzymes and glycosyltransferases have been characterized. This means that these gene/enzyme components can potentially be used in their natural cellular contexts or in artificially assembled *in vivo* or *in vitro* systems for the synthesis of novel glycosylated products. All successful bioengineering applications ultimately depend on the tolerance of these biosynthetic enzymes for alternative substrates. Particularly important is the availability of 'substrate-flexible' glycosyltransferases, which, as highlighted below, have recently been used for the glycoengineering of secondary metabolites.

Glycodiversification of natural products *in vivo*

An overview of the common strategies for biological glycodiversification of natural products is outlined in Fig. 5. Precursor-directed biosynthesis involves feeding analogues of a natural product precursor to cell cultures of the producing organism to effect structural changes in a natural product[58]. Mutational biosynthesis (or mutasynthesis) is a variation on precursor-directed biosynthesis that involves blocking the formation of an intermediate that is being replaced by the exogenously fed analogue through gene disruption[59]. Metabolic pathway

engineering, or combinatorial biosynthesis, is another emerging technique used to alter the structure of a given natural product *in vivo*, and has recently been used in the glycodiversification of several classes of natural product aglycone. In one early glycodiversification study, expression of a glucosyltransferase from the vancomycin biosynthetic pathway of *Amycolatopsis orientalis* in *Streptomyces toyocaensis*, the producer of the structurally similar antibiotic A47934, resulted in the formation of a glucosylated A47934 derivative[60].

In a separate study, the daunorubicin biosynthetic pathway of *Streptomyces peucetius* was altered to produce epirubicin, a semi-synthetic antitumour drug with improved therapeutic properties. Structurally, daunorubicin and epirubicin differ only in the configuration of the C4 hydroxyl group of the attached sugar, which is L-daunosamine (**37**, Fig. 1) in daunorubicin. In this experiment, the gene encoding the proposed sugar 4-ketoreductase of the daunorubicin pathway, *dnrV*, was disrupted and replaced with either *avrB* or *eryBIV*, two *dnrV* homologues from the avermectin and erythromycin biosynthetic pathways, respectively. These two genes were predicted to encode enzymes that could catalyse the reduction of 4-ketosugars with the opposite stereochemistry to DnrV. Indeed, both replacement mutants made significant quantities of epirubicin[61]. Interestingly, the epirubicin sugar generated by these studies, 4-epi-L-daunosamine, does not occur naturally, illustrating the potential of bioengineering methods to construct novel sugar structures using combinations of natural biosynthetic enzymes.

Although the relaxed substrate tolerance of natural product glycosyltransferases was implicated by these early glycoengineering results, a more detailed investigation of the substrate specificity of a natural product glycosyltransferase (DesVII, a desosaminyltransferase) was only conducted recently in studies of TDP-D-desosamine (**19**, Fig. 1) biosynthesis in the methymycin/pikromycin pathway of *S. venezuelae*. As shown in Fig. 6a, individual disruption of four separate D-desosamine biosynthetic genes, *desI*, *desII*, *desV* and *desVI*, led to the formation of new macrolide analogues (**41**–**44**) carrying sugars corresponding to the accumulated D-desosamine pathway intermediates[62–65] (**8** and **38**–**40**). Expression of gene cassettes encoding TDP-D-dihydrostreptose (**45**, Fig. 1) biosynthetic genes (*strM* and *strL*) from *Streptomyces griseus*[66], or TDP-D-mycaminose (**46**, Fig. 1) biosynthetic genes (*tyl1a* and *tylB*) from *S. fradiae*[67] in an *S. venezuelae* mutant lacking a functional *desI* gene (*KdesI*) led to novel methymycin/pikromycin derivatives bearing L-rhamnose (**47**) or D-mycaminose (**48**), respectively. In a separate

study, a library of 14-membered ring macrolactones was also shown to be glycosylated by DesVII (ref. 68). Together, these results clearly demonstrated that DesVII is promiscuous towards both its sugar and aglycone substrates.

A second example of a substrate-promiscuous glycosyltransferase that has been used to create an array of glycosylated natural product analogues *in vivo* is ElmGT from the elloramycin pathway of *Streptomyces olivaceus*. Using a combination of gene-knockout, heterologous-expression and feeding experiments in several producing and non-producing *Streptomyces* strains, ElmGT was used to generate a total of 12 new elloramycin derivatives, each bearing a unique sugar structure[69–74]. Cumulatively, these and other[51,75–80] examples firmly established the feasibility of using pathway engineering to produce novel natural product glycoforms through the assembly of custom-designed sugar biosynthetic pathways and substrate-promiscuous glycosyltransferases.

Although many natural product glycosyltransferases exhibit inherent substrate flexibility, others may be amenable to protein engineering to further alter or expand their donor and/or acceptor substrate specificities. In a remarkable set of experiments, Hoffmeister *et al.* constructed several chimaeric variants of two urdamycin A glycosyltransferases,

UrdGT1b and UrdGT1c, which have 91% sequence identity yet show different specificities for both sugar and acceptor substrates. *In vivo* analysis of the activities of a series of UrdGT1b/UrdGT1c chimaeras allowed the identification of a 30-amino-acid region near the amino terminus of both proteins that is responsible for both donor and acceptor substrate recognition. Through mutagenesis, 10 of the 18 non-identical amino acids in this region were shown to be important for specificity. This allowed the construction of a library of 2,048 variants harbouring different combinations of parental amino-acid sequences, and these variants were subsequently screened for new activities *in vivo*[81,82]. Several of these chimaeric enzymes exclusively catalysed the production of a novel derivative of urdamycin that contained a branched sugar appendage. Although this feat of engineering has not yet been duplicated in other systems, recent progress in the development of high-throughput screening methods for glycosyltransferase activity[83,84] will facilitate future efforts to construct glycosyltransferases with unnatural activities and substrate specificities. Another significant limitation to successful glycosyltransferase engineering is the dearth of structural data on these enzymes. So far, the crystal structures of only a handful of natural product glycosyltransferases have been reported[85–88].
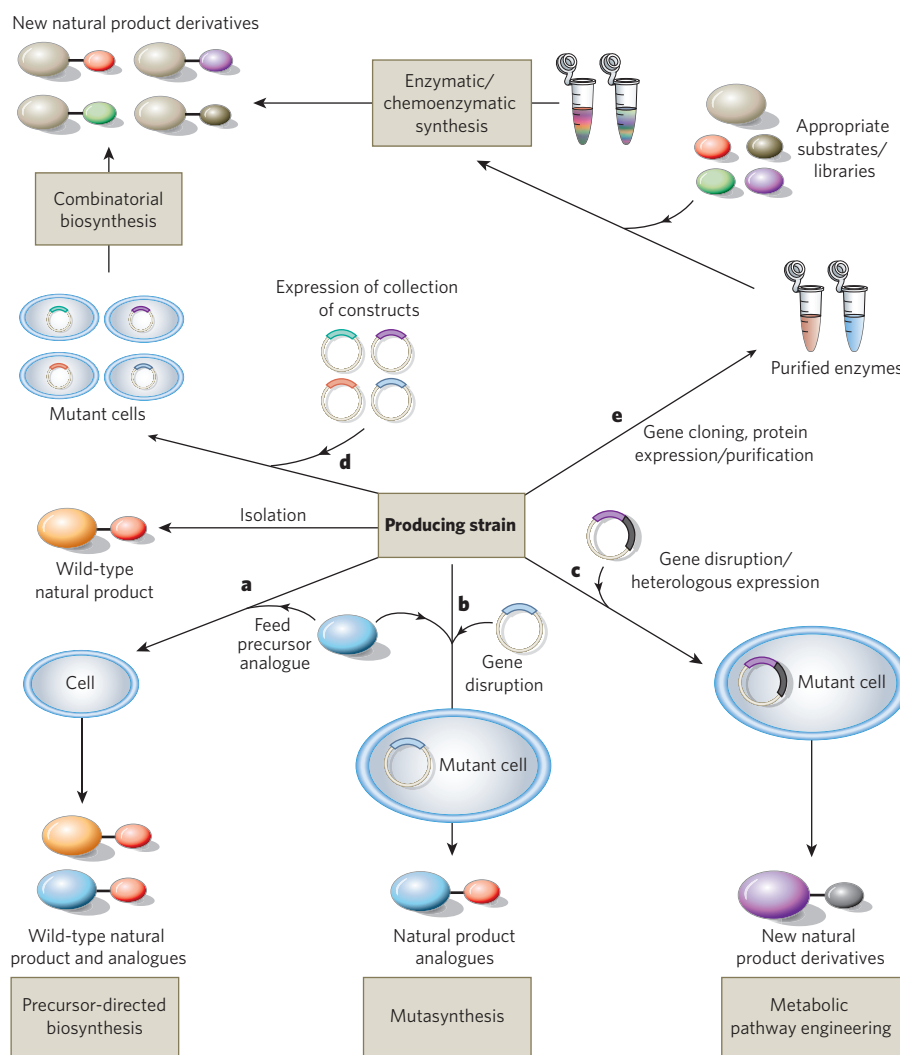
Figure 5 | **Overview of bioengineering strategies.** A producing strain can be manipulated to generate new natural products through a combination of feeding experiments (precursor-directed biosynthesis (**a**) or mutasynthesis (**b**)) and/or genetic manipulation (gene disruption or gene disruption/heterologous expression). In metabolic pathway engineering (**c**), a sugar biosynthetic intermediate is rerouted to generate a new product through gene disruption and heterologous gene expression. In combinatorial biosynthesis (**d**), sugar biosynthetic pathways from separate organisms can be re-engineered and 'combined' in a single organism to give hybrid or novel natural products. In enzymatic/chemoenzymatic synthesis (**e**), sugar biosynthetic enzymes with natural or engineered substrate flexibility are isolated and used for the construction of natural product libraries *in vitro*, using purified aglycone and NDP-sugar substrates.

1013

As the three-dimensional structures of more glycosyltransferases become available, it may become possible to identify key amino acids that could be mutated to alter the substrate specificity or catalytic activity of these glycosyltransferases.

Glycodiversification of natural products *in vitro*

The *in vitro* demonstration of natural product glycosyltransferase activity has facilitated both enzymatic and chemoenzymatic synthesis of diverse natural product glycoforms. The availability of various NDP-sugar and aglycone substrates is crucial to the success of this approach. Recently, improvements in chemical methods for NDP-sugar synthesis and the development of *in vitro* glycorandomization (IVG, also referred to as glycodiversification or glyco-optimization) have made small libraries of NDP-sugars accessible for *in vitro* glycosylation assays. IVG is a chemoenzymatic method that uses engineered substrate-flexible anomeric sugar kinases and nucleotidylyltransferases for the conversion of diverse sugars to their NDP-activated forms. Application of IVG to the GtfE-catalysed glycosylation of the vancomycin aglycone yielded more than 30 vancomycin analogues with altered sugar structures, including three with sugars containing an azide functional group that permitted further combinatorial functionalization using the Huisgen 1,3-dipolar

cycloaddition ('click' chemistry)[89]. IVG has also been used to assess glycosyltransferase substrate flexibility and to generate glycosylated natural product analogues of methymycin[45] and novobiocin[90].

Recently, the Eguchi and Thorson research groups demonstrated that several natural product glycosyltransferase-catalysed reactions are reversible *in vitro*[91,92]. In the presence of a glycosylated natural product and NDP, these glycosyltransferases can synthesize the corresponding NDP-sugar and aglycone. Thus, the reversibility of the glycosyltransferase-catalysed reaction could provide easy access to various NDP-sugars for *in vitro* studies. Next, they exploited this reversibility to perform one-pot sugar and aglycone exchange reactions. This methodology, combined with IVG, led to the synthesis of more than 70 derivatives of the antitumour compound calicheamicin[91] (Fig. 6b). These demonstrations of glycosyltransferase utility are likely to affect many areas of glycobiology and glycochemistry research[93].

Conclusions and future directions

The enzymes involved in the biosynthesis of the unusual sugars in bacteria are excellent targets for drug development, because most of these sugars are absent in humans. Many of them are important constituents of the bacterial cell wall, and their biosyntheses involve
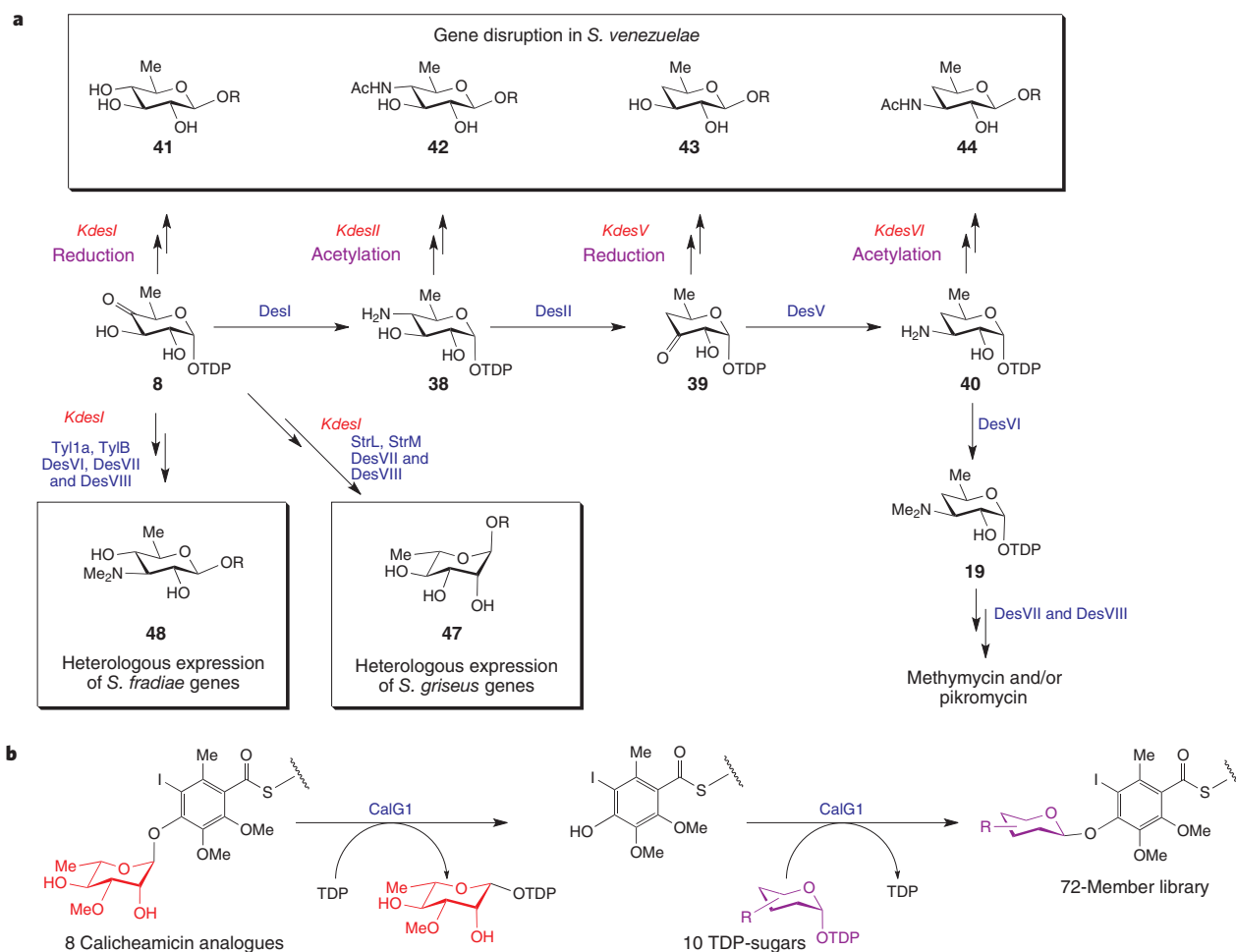
Figure 6 | Glycodiversification of natural products. a, *In vivo* glycodiversification in *S. venezuelae*. The D-desosamine biosynthetic pathway (8 → 38 → 39 → 40 → 19) was altered either through gene disruption to give 41–44 or through a combination of gene disruption and heterologous gene expression to give 47 and 48. In all cases, the endogenous glycosyltransferase–auxiliary-protein pair DesVII and DesVIII coupled the 'unnatural' sugars to the native *S. venezuelae* aglycones (in which R represents methynolide or pikronolide). The particular *S. venezuelae* mutant that was used in each study is shown in red (for example, in *KdesI*, the *desI* gene is disrupted). The names of the enzymes involved

in the chemical transformations are shown in blue. Other enzymatic modifications that are required to generate the observed products are shown in purple. The enzymes responsible for the N-acetylation in 42 and 44 and the keto-group reduction in 41 and 43 are unknown. **b**, Applications of *in vitro* glycorandomization. After removal of the 3-*O*-methyl-rhamnose moiety of eight different calicheamicin analogues by means of a reverse-glycosyltransferase reaction, the calicheamicin glycosyltransferase, CalG1, was found to couple 10 different TDP-sugars to 8 calicheamicin aglycones, yielding a 72-member glycorandomized library. (Eight of the possible 80 combinations were not obtained.)

mechanistically novel enzymatic steps that have been actively targeted for inhibitor development[94–97]. Thus, future mechanistic and structural studies of bacterial carbohydrate biosynthetic enzymes will undoubtedly continue to present researchers with alternative drug targets and inspiration for the development of new antibiotics. As an alternative approach to drug design, engineering natural product sugar structures through the manipulation of sugar biosynthetic machinery is an effective means by which to generate glycoconjugates with new or enhanced biological activities. The viability of *in vivo* bioengineering strategies, such as combinatorial biosynthesis, and *in vitro* chemoenzymatic synthesis methods have infused new excitement into glycodiversification efforts[93]. Rapid advances in the identification and sequencing of natural product biosynthetic gene clusters, coupled with the isolation and the structural and functional characterization of the encoded enzymes, will undoubtedly augment our ability to customize the glycosylation patterns of natural products through these bioengineering approaches. The power, elegance and utility of nature's biosynthetic machinery has clearly ushered in a new era of carbohydrate research, and has ensured a promising future for natural product glycodiversification in the search for new and improved drugs. ∎

Note added in proof: Lovering *et al.* recently demonstrated that the glycosyltransferase domain of a bifunctional glycosyltransferase/transpeptidase enzyme involved in peptidoglycan biosynthesis in *Staphylococcus aureus* adopts a structural fold that is distinct from that of the GT-A and GT-B glycosyltransferase families, making it the first member of a new family of glycosyltransferases to be structurally characterized[98].

1. Weymouth-Wilson, A. C. The role of carbohydrates in biologically active natural products. *Nat. Prod. Rep.* **14,** 99–110 (1997).
2. Thorson, J. S., Hosted, T. J., Jiang, J., Biggins, J. B. & Ahlert, J. Nature's carbohydrate chemists: the enzymatic glycosylation of bioactive bacterial metabolites. *Curr. Org. Chem.* **5,** 139–167 (2001).
3. Barton, W. A. *et al.* Structure, mechanism and engineering of a nucleotidylyltransferase as a first step toward glycorandomization. *Nature Struct. Biol.* **8,** 545–551 (2001).
4. Varki, A. *et al. Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).
5. Johnson, D. A. & Liu, H.-w. in *Comprehensive Natural Product Chemistry* (eds Barton, D. H. R., Meth-Cohn, O. & Nakanishi, K.) 311–365 (Elsevier, Amsterdam, 1999).
6. He, X., Agnihotri, G. & Liu, H.-w. Novel enzymatic mechanisms in carbohydrate metabolism. *Chem. Rev.* **100,** 4615–4661 (2000).
7. He, X. M. & Liu, H.-w. Formation of unusual sugars: mechanistic studies and biosynthetic applications. *Annu. Rev. Biochem.* **71,** 701–754 (2002).
8. Hallis, T. M. & Liu, H.-w. Learning Nature's strategies for making deoxy sugars: pathways, mechanisms, and combinatorial applications. *Acc. Chem. Res.* **32,** 579–588 (1999).
9. He, X. & Liu, H.-w. Mechanisms of enzymatic C–O bond cleavages in deoxyhexose biosynthesis. *Curr. Opin. Chem. Biol.* **6,** 590–597 (2002).
10. Szu, P.-h., He, X., Zhao, L. & Liu, H.-w. Biosynthesis of TDP-ᴅ-desosamine: identification of a strategy for C4 deoxygenation. *Angew. Chem. Int. Ed.* **44,** 6742–6746 (2005).
11. Oppermann, U. *et al.* Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chem. Biol. Interact.* **143–144,** 247–253 (2003).
12. Frey, P. A. The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. *FASEB J.* **10,** 461–470 (1996).
13. Thoden, J. B., Frey, P. A. & Holden, H. M. Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose-4-epimerase from *Escherichia coli*: implications for the catalytic mechanism. *Biochemistry* **35,** 5137–5144 (1996).
14. Hallis, T. M., Zhao, Z. & Liu, H.-w. New insights into the mechanism of CDP-ᴅ-tyvelose 2-epimerase: an enzyme catalyzing epimerization at an unactivated stereocenter. *J. Am. Chem. Soc.* **122,** 10493–10503 (2000).
15. Morrison, J. P., Read, J. A., Coleman, W. G. & Tanner, M. E. Dismutase activity of ADP-ʟ-*glycero*-ᴅ-*manno*-heptose 6-epimerase: evidence for a direct oxidation/reduction mechanism. *Biochemistry* **44,** 5907–5915 (2005).
16. Major, L. L., Wolucka, B. A. & Naismith, J. H. Structure and function of GDP-mannose-3′,5′-epimerase: an enzyme which performs three chemical reactions at the same active site. *J. Am. Chem. Soc.* **127,** 18309–18320 (2005).
17. Mulichak, A. M., Theisen, M. J., Essigmann, B., Benning, C. & Garavito, R. M. Crystal structure of SQD1, an enzyme involved in the biosynthesis of the plant sulfolipid headgroup donor UDP-sulfoquinovose. *Proc. Natl Acad. Sci. USA* **96,** 13097–13102 (1999).
18. Sanda, S., Leustek, T., Theisen, M. J., Garavito, R. M. & Benning, C. Recombinant *Arabidopsis* SQD1 converts UDP-glucose and sulfite to the sulfolipid head group precursor UDP-sulfoquinovose *in vitro. J. Biol. Chem.* **276,** 3941–3946 (2001).
19. Schutzbach, J. S. & Feingold, D. S. Biosynthesis of uridine diphosphate ᴅ-xylose IV. Mechanism of action of uridine diphosphoglucuronate carboxy-lyase. *J. Biol. Chem.* **245,** 2476–2482 (1970).
20. Davies, G. J., Gloster, T. M., & Henrissat, B. Recent structural insights into the expanding world of carbohydrate-active enzymes. *Curr. Opin. Struct. Biol.* **15,** 637–645 (2005).
21. Campbell, J. A., Davies, G. J., Bulone, V. & Henrissat, B. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.* **326,** 929–942 (1997).
22. Coutinho, P. M., Deleury, E., Davies, G. J. & Henrissat, B. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **328,** 307–317 (2003).
23. Ünligil, U. M. & Rini, J. M. Glycosyltransferase structure and mechanism. *Curr. Opin. Struct. Biol.* **10,** 510–517 (2000).
24. Bourne, Y. & Henrissat, B. Glycoside hydrolases and glycosyltransferases: families and functional modules. *Curr. Opin. Struct. Biol.* **11,** 593–600 (2001).
25. Breton, C., Mucha, J. & Jeanneau, C. Structural and functional features of glycosyltransferases. *Biochimie* **83,** 713–718 (2001).
26. Pak, J. E. *et al.* X-ray crystal structure of leukocyte type core 2 β1,6-*N*-acetylglucosaminyltransferase. Evidence for a convergence of metal ion-independent glycosyltransferase mechanism. *J. Biol. Chem.* **281,** 26693–26701 (2006).
27. Chiu, C. P. *et al.* Structural analysis of the sialyltransferase CstII from *Campylobacter jejuni* in complex with a substrate analogue. *Nature Struct. Mol. Biol.* **11,** 163–170 (2004).
28. Murray, B. W., Takayama, S., Schultz, J. & Wong, C.-H. Mechanism and specificity of human α-1,3-fucosyltransferase V. *Biochemistry* **35,** 11183–11195 (1996).
29. Hu, Y. & Walker, S. Remarkable structural similarities between diverse glycosyltransferases. *Chem. Biol.* **9,** 1287–1296 (2002).
30. Sinnott, M. L. Catalytic mechanisms of enzymic glycosyl transfer. *Chem. Rev.* **90,** 1171–1202 (1990).
31. Gibson, R. P., Turkenburg, J. P., Charnock, S. J., Lloyd, R. & Davies, G. J. Insights into trehalose synthesis provided by the structure of the retaining glucosyltransferase OtsA. *Chem. Biol.* **9,** 1337–1346 (2002).
32. Pedersen, L. C., Darden, T. A. & Negishi, M. Crystal structure of β1,3-glucuronyltransferase I in complex with active donor substrate UDP-GlcUA. *J. Biol. Chem.* **277,** 21869–21873 (2002).
33. Charnock, S. J. & Davies, G. J. Structure of the nucleotide-diphospho-sugar transferase, SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* **38,** 6380–6385 (1999).
34. Tarbouriech, N., Charnock, S. J. & Davies, G. J. Three-dimensional structures of the Mn and Mg dTDP complexes of the family GT-2 glycosyltransferase SpsA: a comparison with related NDP-sugar glycosyltransferases. *J. Mol. Biol.* **314,** 655–661 (2001).
35. Qiao, L., Murray, B. W., Shimazaki, M., Schultz, J. & Wong, C.-H. Synergistic inhibition of human α-1,3-fucosyltransferase V. *J. Am. Chem. Soc.* **118,** 7653–7662 (1996).
36. Tvaroska, I., Andre, I. & Carver, J. P. *Ab initio* molecular orbital study of the catalytic mechanism of glycosyltransferases: description of reaction pathways and determination of transition-state structures for inverting *N*-acetylglucosaminyltransferases. *J. Am. Chem. Soc.* **122,** 8762–8776 (2000).
37. Lairson, L. L. *et al.* Intermediate trapping on a mutant retaining α-galactosyltransferase identifies an unexpected aspartate residue. *J. Biol. Chem.* **279,** 28339–28344 (2004).
38. Zechel, D. L. & Withers, S. G. Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* **33,** 11–18 (2000).
39. Persson, K. *et al.* Crystal structure of the retaining galactosyltransferase LgtC from *Neisseria meningitidis* in complex with donor and acceptor sugar analogs. *Nature Struct. Biol.* **8,** 166–175 (2001).
40. Pedersen, L. C. *et al.* Crystal structure of an α-1,4-*N*-acetylhexosaminyltransferase (EXTL2), a member of the exostosin gene family involved in heparan sulfate biosynthesis. *J. Biol. Chem.* **278,** 14420–14428 (2003).
41. Boix, E. *et al.* Structure of UDP complex of UDP-galactose:β-galactoside-α-1,3-galactosyltransferase at 1.53-Å resolution reveals a conformational change in the catalytically important C terminus. *J. Biol. Chem.* **276,** 48608–48614 (2001).
42. Boix, E., Zhang, Y., Swaminathan, G. J., Brew, K. & Acharya, K. R. Structural basis of ordered binding of donor and acceptor substrates to the retaining glycosyltransferase, α-1,3-galactosyltransferase. *J. Biol. Chem.* **277,** 28310–28318 (2002).
43. Ly, H. D., Lougheed, B., Wakarchuk, W. W. & Withers, S. G. Mechanistic studies of a retaining α-galactosyltransferase from *Neisseria meningitidis. Biochemistry* **41,** 5075–5085 (2002).
44. Zhang, Y. *et al.* Roles of individual enzyme-substrate interactions by α-1,3-galactosyltransferase in catalysis and specificity. *Biochemistry* **42,** 13512–13521 (2003).
45. Borisova, S. A. *et al.* Substrate specificity of the macrolide-glycosylating enzyme pair DesVII/DesVIII: opportunities, limitations, and mechanistic hypotheses. *Angew. Chem. Int. Ed. Engl.* **45,** 2748–2753 (2006).
46. Kao, C.-L., Borisova, S. A., Kim, H. J., & Liu, H.-w. Linear aglycones are the substrates for glycosyltransferase DesVII in methymycin biosynthesis: analysis and implications. *J. Am. Chem. Soc.* **128,** 5606–5607 (2006).
47. Yuan, Y. *et al. In vitro* reconstitution of EryCIII activity for the preparation of unnatural macrolides. *J. Am. Chem. Soc.* **127,** 14128–14129 (2005).
48. Lu, W. *et al.* AknT is an activating protein for the glycosyltransferase AknS in ʟ-aminodeoxysugar transfer to the aglycone of aclacinomycin A. *Chem. Biol.* **12,** 527–534 (2005).
49. Imperiali, B. & Tai, V. W. -F. in *Carbohydrate-based Drug Discovery* Vol. 1 (ed. Wong, C. -H.) 281–303 (Wiley-VCH, Weinheim, 2003).
50. Onaka, H., Taniguchi, S.-i., Igarashi, Y. & Furumai, T. Characterization of the biosynthetic gene cluster of rebeccamycin from *Lechevalieria aerocolonigenes* ATCC 39243. *Biosci. Biotechnol. Biochem.* **67,** 127–138 (2003).
51. Salas, A. P. *et al.* Deciphering the late steps in the biosynthesis of the anti-tumour indolocarbazole staurosporine: sugar donor substrate flexibility of the StaG glycosyltransferase. *Mol. Microbiol.* **58,** 17–27 (2005).
52. Gao, Q., Zhang, C., Blanchard, S. & Thorson, J. S. Deciphering indolocarbazole and enediyne aminodideoxypentose biosynthesis through comparative genomics: insights from the AT2433 biosynthetic locus. *Chem. Biol.* **13,** 733–743 (2006).
53. Wacker, M. *et al. N*-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli. Science* **298,** 1790–1793 (2002).
54. Hultin, P. G. Bioactive *C*-glycosides from bacterial secondary metabolism. *Curr. Topics Med. Chem.* **5,** 1299–1331 (2005).
55. Bililign, T., Hyun, C.-G., Williams, J. S., Czisny, A. M. & Thorson, J. S. The hedamycin locus implicates a novel aromatic PKS priming mechanism. *Chem. Biol.* **11,** 959–969 (2004).
56. Halkier, B. A. & Gershenzon, J. Biology and biochemistry of glucosinolates. *Annu. Rev. Plant Biol.* **57,** 303–333 (2006).

57. Grubb, C. D. *et al.* *Arabidopsis* glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. *Plant J.* **40,** 893–908 (2004).

58. Thiericke, R. & Rohr, J. Biological variation of microbial metabolites by precursor-directed biosynthesis. *Nat. Prod. Rep.* **10,** 265–289 (1993).

59. Weist, S. & Sussmuth, R. D. Mutational biosynthesis — a tool for the generation of structural diversity in the biosynthesis of antibiotics. *Appl. Microbiol. Biotech.* **68,** 141–150 (2005).

60. Solenberg, P. J. *et al.* Production of hybrid glycopeptide antibiotics *in vitro* and in *Streptomyces toyocaensis. Chem. Biol.* **4,** 195–202 (1997).

61. Madduri, K. *et al.* Production of the antitumor drug epirubicin (4′-epidoxorubicin) and its precursor by a genetically engineered strain of *Streptomyces peucetius. Nature Biotechnol.* **16,** 69–74 (1998).

62. Zhao, L., Sherman, D. H. & Liu, H.-w. Biosynthesis of desosamine: construction of a new methymycin/neomethymycin analogue by deletion of a desoamine biosynthetic gene. *J. Am. Chem. Soc.* **120,** 10256–10257 (1998).

63. Zhao, L., Que, N. L. S., Xue, Y., Sherman, D. H. & Liu, H.-w. Mechansitc studies of desosamine biosynthesis: C-4 deoxygenation precedes C-3 transamination. *J. Am. Chem. Soc.* **120,** 12159–12160 (1998).

64. Zhao, L., Borisova, S., Yeung, S. M. & Liu, H. Study of C-4 deoxygenation in the biosynthesis of desosamine: evidence implicating a novel mechanism. *J. Am. Chem. Soc.* **123,** 7909–7910 (2001).

65. Borisova, S. A., Zhao, L., Sherman, D. H. & Liu, H. W. Biosynthesis of desosamine: construction of a new macrolide carrying a genetically designed sugar moiety. *Org. Lett.* **1,** 133–136 (1999).

66. Yamase, H., Zhao, L. & Liu, H.-w. Engineering a hybrid sugar biosynthetic pathway: production of L-rhamnose and its implication on dihydrostreptose biosynthesis. *J. Am. Chem. Soc.* **122,** 12397–12398 (2000).

67. Melancon, C. E., Yu, W. L. & Liu, H. W. TDP-mycaminose biosynthetic pathway revised and conversion of desosamine pathway to mycaminose pathway with one gene. *J. Am. Chem. Soc.* **127,** 12240–12241 (2005).

68. Tang, L. & McDaniel, R. Construction of desosamine containing polyketide libraries using a glycosyltransferase with broad substrate specificity. *Chem. Biol.* **8,** 547–555 (2001).

69. Decker, H., Haag, S., Udvarnoki, G. & Rohr, J. Novel genetically engineered tetracenomycins. *Angew. Chem. Int. Ed. Engl.* **34,** 1107–1110 (1995).

70. Wohlert, S.-E. *et al.* Novel hybrid tetracenomycins through combinatorial biosynthesis using a glycosyltransferase encoded by the *elm* genes in cosmid 16F4 and which shows broad sugar substrate specificity. *J. Am. Chem. Soc.* **120,** 10596–10601 (1998).

71. Rodriguez, L. *et al.* Engineering deoxysugar biosynthetic pathways from antibiotic-producing microorganisms: a tool to produce novel glycosylated bioactive compounds. *Chem. Biol.* **9,** 721–729 (2002).

72. Fischer, C. *et al.* Digitoxosyltetracenomycin C and glucosyltetracenomycin C, two novel elloramycin analogues obtained by exploring the sugar donor substrate flexibility of glycosyltransferase ElmGT. *J. Nat. Prod.* **65,** 1685–1689 (2002).

73. Lombo, F. *et al.* Engineering biosynthetic pathways for deoxysugars: branched-chain sugar pathways and derivatives from the antitumor tetracenomycin. *Chem. Biol.* **11,** 1709–1718 (2004).

74. Perez, M. *et al.* Combining sugar biosynthesis genes for the generation of L- and D-amicetose and formation of two novel antitumor tetracenomycins. *Chem. Commun. (Camb.)* **12,** 1604–1606 (2005).

75. Zhang, C. *et al.* RebG- and RebM-catalyzed indolocarbazole diversification. *ChemBioChem* **7,** 795–804 (2006).

76. Sanchez, C. *et al.* Combinatorial biosynthesis of antitumor indolocarbazole compounds. *Proc. Natl Acad. Sci. USA* **102,** 461–466 (2005).

77. Trefzer, A. *et al.* Elucidation of the function of two glycosyltransferase genes (*lanGT1* and *lanGT4*) involved in landomycin biosynthesis and generation of new oligosaccharide antibiotics. *Chem. Biol.* **8,** 1239–1252 (2001).

78. Torkkell, S. *et al.* The entire nogalamycin biosynthetic gene cluster of *Streptomyces nogalater*: characterization of a 20-kb DNA region and generation of hybrid structures. *Mol. Gen. Genet.* **266,** 276–288 (2001).

79. Wohlert, S.-E. *et al.* Insights about the biosynthesis of the avermectin deoxysugar L-oleandrose through heterologous expression of *Streptomyces avermitilis* deoxysugar genes in *Streptomyces lividans. Chem. Biol.* **8,** 681–700 (2001).

80. Gaisser, S. *et al.* A defined system for hybrid macrolide biosynthesis in *Saccharopolyspora erythraea. Mol. Microbiol.* **36,** 391–401 (2000).

81. Hoffmeister, D. *et al.* Engineered urdamycin glycosyltransferases are broadened and altered in substrate specificity. *Chem. Biol.* **9,** 287–295 (2002).

82. Hoffmeister, D., Ichinose, K. & Bechthold, A. Two sequence elements of glycosyltransferases involved in urdamycin biosynthesis are responsible for substrate specificity and enzymatic activity. *Chem. Biol.* **8,** 557–567 (2001).

83. Love, K. R., Swoboda, J. G., Noren, C. J. & Walker, S. Enabling glycosyltransferase evolution: a facile substrate-attachment strategy for phage-display enzyme evolution. *ChemBioChem* **7,** 753–756 (2006).

84. Aharoni, A. *et al.* High-throughput screening methodology for the directed evolution of glycosyltransferases. *Nature Methods* **3,** 609–614 (2006).

85. Mulichak, A. M., Losey, H. C., Walsh, C. T. & Garavito, R. M. Structure of the UDP-glucosyltransferase GtfB that modifies the heptapeptide aglycone in the biosynthesis of vancomycin group antibiotics. *Structure* **9,** 547–557 (2001).

86. Mulichak, A. M. *et al.* Structure of the TDP-epi-vancosaminyltransferase GtfA from the chloroeremomycin biosynthetic pathway. *Proc. Natl Acad. Sci. USA* **100,** 9238–9243 (2003).

87. Mulichak, A. M., Lu, W., Losey, H. C., Walsh, C. T. & Garavito, R. M. Crystal structure of vancosaminyltransferase GtfD from the vancomycin biosynthetic pathway: interactions with acceptor and nucleotide ligands. *Biochemistry* **43,** 5170–5180 (2004).

88. Offen, W. *et al.* Structure of a flavonoid glucosyltransferase reveals the basis for plant natural product modification. *EMBO J.* **25,** 1396–1405 (2006).

89. Fu, X. *et al.* Antibiotic optimization via *in vitro* glycorandomization. *Nature Biotechnol.* **21,** 1467–1469 (2003).

90. Albermann, C. *et al.* Substrate specificity of NovM: implications for novobiocin biosynthesis and glycorandomization. *Org. Lett.* **5,** 933–936 (2003).

91. Zhang, C. *et al.* Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions. *Science* **313,** 1291–1294 (2006).

92. Minami, A., Kakinuma, K. & Eguchi, T. Aglycon switch approach toward unnatural glycosides from natural glycoside with glycosyltransferase VinC. *Tetrahedr. Lett.* **46,** 6187–6190 (2005).

93. Melancon, C. E., Thibodeaux, C. J. & Liu, H.-w. Glyco-stripping and glyco-swapping. *ACS Chem. Biol.* **1,** 499–504 (2006).

94. Grison, C., Petek, S., Finance, C. & Coutrot, P. Synthesis and antibacterial activity of mechanism-based inhibitors of KDO8P synthase and DAH7P synthase. *Carbohydrate Res.* **340,** 529–537 (2005).

95. Carlson, E. E., May, J. F. & Kiessling, L. L. Chemical probes of UDP-galactopyranose mutase. *Chem. Biol.* **13,** 825–837 (2006).

96. van Heijenoort, J. Recent advances in the formation of the bacterial peptidoglycan monomer unit. *Nat. Prod. Rep.* **18,** 503–519 (2001).

97. Zoeiby, A. E., Sanschagrin, F. & Levesque, R. C. Structure and function of the Mur enzymes development of novel inhibitors. *Mol. Microbiol.* **47,** 1–12 (2003).

98. Lovering, A. L., de Castro, L. H., Lim, D. & Strynadka, N. C. J. Structural insight into the transglycosylation step of bacterial cell-wall biosynthesis. *Science* **315,** 1402–1405 (2007).

Cycling of O-linked β-*N*-acetylglucosamine on nucleocytoplasmic proteins

Gerald W. Hart[1], Michael P. Housley[1] & Chad Slawson[1]

All animals and plants dynamically attach and remove O-linked β-*N*-acetylglucosamine (*O*-GlcNAc) at serine and threonine residues on myriad nuclear and cytoplasmic proteins. *O*-GlcNAc cycling, which is tightly regulated by the concerted actions of two highly conserved enzymes, serves as a nutrient and stress sensor. On some proteins, *O*-GlcNAc competes directly with phosphate for serine/threonine residues. Glycosylation with *O*-GlcNAc modulates signalling, and influences protein expression, degradation and trafficking. Emerging data indicate that *O*-GlcNAc glycosylation has a role in the aetiology of diabetes and neurodegeneration.

The discovery of O-linked β-*N*-acetylglucosamine (*O*-GlcNAc) more than 20 years ago[1] disproved the long-held dogma that protein glycosylation is restricted to the luminal compartments of the secretory machinery and to the cell surface and extracellular matrix (see ref. 2 for a review). Early studies of *O*-GlcNAc's subcellular localization in rat hepatocytes[3] established that it is highly concentrated at the nuclear envelope, particularly at the nuclear pore complex, but is also abundant and widespread within chromatin. However, several cytosolic and cytoskeletal proteins were also found to be glycosylated with *O*-GlcNAc (O-GlcNAcylated). Later studies of *Drosophila* polytene chromosomes established that O-GlcNAcylated proteins are abundant throughout chromosomes. The carboxy-terminal domain of a subpopulation of RNA polymerase II is extensively O-GlcNAcylated, and almost all RNA polymerase II transcription factors are modified by the sugar (see ref. 4 for a review of early studies). So far, more than 500 proteins have been identified to be O-GlcNAcylated, and these proteins are involved in almost all aspects of cellular metabolism[5–7].

O-GlcNAcylation is in many ways distinct from 'classical' protein glycosylation. First, it is found mostly within the cytoplasm or nucleoplasm. Second, unlike the extraordinarily complex array of glycans found on extracellular glycoproteins, *O*-GlcNAc is not elongated or further modified.

Third, *O*-GlcNAc cycles (Fig. 1) by means of mechanisms and on a timescale similar to those of phosphorylation and quite different from the cycling of complex extracellular glycans (see ref. 8 for a review). O-GlcNAcylation is one of the most common post-translational modifications. In terms of high-energy compounds, the intracellular concentration of the direct donor for O-GlcNAcylation, UDP-GlcNAc, is second only to that of ATP, with 2–5% of all glucose being used to generate this sugar nucleotide[9].

If *O*-GlcNAc is so abundant, why was it not detected until 1983, and why is it still overlooked by most of those studying signalling and transcription? First and foremost, because *O*-GlcNAc is uncharged and small, it does not usually alter the migration of proteins in gel electrophoresis, even on high-resolution two-dimensional gels. Second, all cells contain high concentrations of hydrolases that rapidly remove *O*-GlcNAc from intracellular proteins upon cellular damage. Third, like phosphorylation, O-GlcNAcylation is often substoichiometric at any site on a protein, making it particularly difficult to detect by physical methods such as mass spectrometry. Most importantly, *O*-GlcNAc is very labile upon ionization in a mass spectrometer, and much of it is often lost at the source. *O*-GlcNAc is also almost entirely lost during standard CAD (collision-activated dissociation) methods[10]. In MALDI-MS (matrix-assisted laser-desorption/ionization mass spectrometry) the unmodified peptides are preferentially ionized, often completely suppressing the signal from *O*-GlcNAc–peptide ions (which are less surface active), when both are present in a mixture. Fortunately, recently developed potent hexosaminidase inhibitors[11,12] and panspecific *O*-GlcNAc antibodies[13] have greatly improved the ease of detection of *O*-GlcNAc. Furthermore, when new mass-spectrometric methods — such as TMS-ECD (Fourier transform mass spectrometry with electron-capture dissociation) or ETD (ion-trap mass spectrometry with electron-transfer dissociation)[14] — have been combined with newly developed enzymatic/chemical tagging and enrichment methods[15,16], they have facilitated detection and site mapping of *O*-GlcNAc. Finally, analyses of *O*-GlcNAc function are complicated by the lack of a recognizable consensus motif.

O-GlcNAc is ubiquitous and essential

O-GlcNAc has been found in all metazoans studied so far. However, genes encoding known *O*-GlcNAc transferases or *O*-GlcNAcases (Fig. 1) have not yet been found in yeast. Numerous viruses that infect metazoans contain proteins modified by *O*-GlcNAc[4]. The O-GlcNAcylated
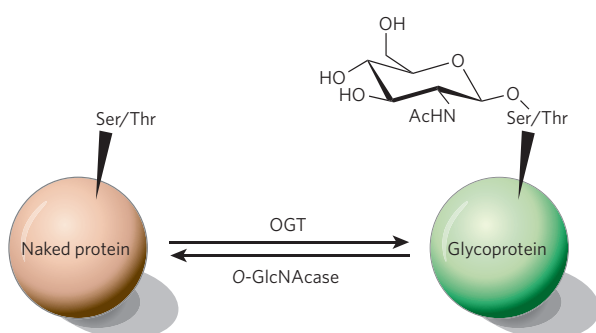

Figure 1 | O-GlcNAcylation cycles on a timescale and in a manner similar to phosphorylation. Cycling of *O*-GlcNAc on serine or threonine residues of nuclear and cytoplasmic proteins is controlled by two highly conserved enzymes, OGT and *O*-GlcNAcase. *O*-GlcNAc cycles at rates similar to that of *O*-phosphate in response to various cellular stimuli.

[1]Department of Biological Chemistry, Johns Hopkins University, School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205-2185, USA.

proteins in these viruses are generally in complexes surrounding their nucleic acids and are not attached to the proteins of the capsid, which is the usual location for classical complex glycans.

O-GlcNAcylation is directly involved in growth hormone (gibberellic acid) signalling in plants, and both *SPY* and *SECRET AGENT* (*SEC*) encode *O*-GlcNAc transferases[17]. Mutations in either *SPY* or *SEC* cause severe growth defects; simultaneous mutation of both genes is lethal. Unlike plants, mammals and insects seem to have only a single gene encoding the catalytic subunit of the *O*-GlcNAc transferase (OGT)[18,19]. Gene disruption in mice established that OGT is required for embryonic-stem-cell viability[20]. Tissue-targeted disruption in mice showed that O-GlcNAcylation is essential to several cell types[21]. OGT deletion causes hyperphosphorylation of tau in neurons, which is followed by cell death, induces T-cell apoptosis and causes growth arrest in fibroblasts. Cre–*lox*-mediated deletion of OGT in cultured fibroblasts results in death as pre-existing OGT protein levels diminish. The *in vivo* tissue-targeted deletions of OGT not only showed that mammalian cells require a functional *Ogt* allele, but also demonstrated that O-GlcNAcylation modulates both protein phosphorylation and expression in essential signalling pathways[21]. By contrast, *ogt*-knockout *Caenorhabditis elegans* are viable, but are defective in dauer larva formation in *daf-2*-knockout animals and in the regulation of macronutrient storage[22]. Interestingly, similar phenotypes were seen in *C. elegans* when *O*-GlcNAcase was deleted, suggesting that *O*-GlcNAc cycling, not just the amount of site occupancy, is important to *O*-GlcNAc's functions[23]. *O*-GlcNAcase deletion has not yet been examined in mammals.
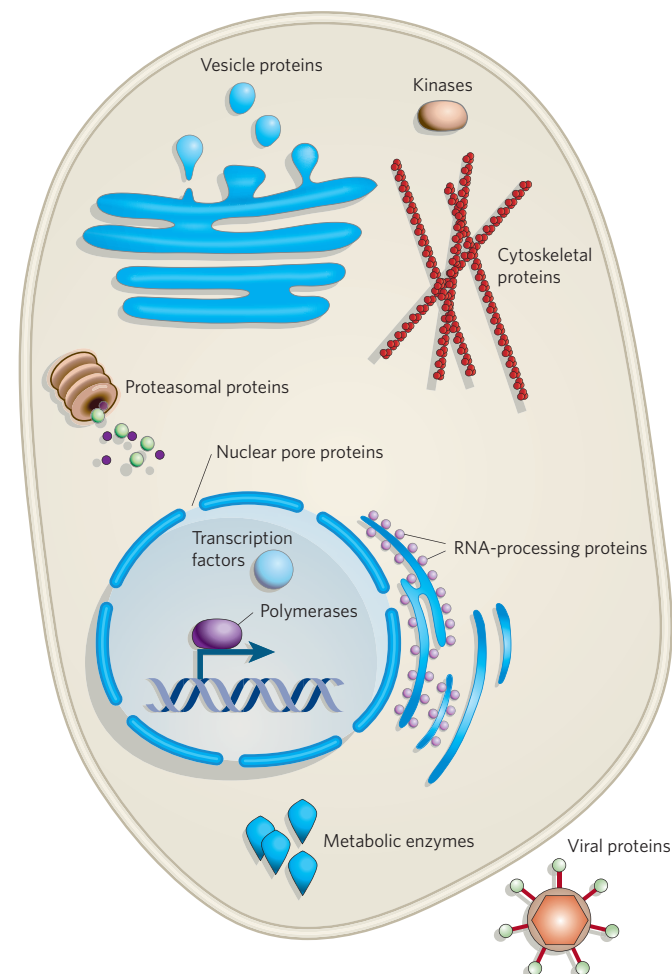
Figure 2 | Many proteins from almost all functional classes are O-GlcNAcylated. In metazoan cells, *O*-GlcNAc cycling occurs on many proteins across almost all functional classes within both the nucleus and the cytoplasm. Many viral proteins are also O-GlcNAcylated.

O-GlcNAc is found on many nuclear and cytoplasmic proteins

O-GlcNAcylated proteins can be found in almost every intracellular compartment, and there are proteins in almost every functional class that are subject to O-GlcNAcylation[5–7] (Fig. 2). About one-quarter of identified O-GlcNAcylated proteins are involved in transcription or translation. O-GlcNAcylation is abundant on proteins involved in stress responses and energy metabolism. Many proteins involved in cytoskeletal regulation are O-GlcNAcylated, particularly those involved in bridging actin and those regulating microtubule assembly[7]. Even tubulin itself is O-GlcNAcylated, but the stoichiometry seems to be quite low. Intermediate filament proteins such as cytokeratins and neurofilaments are extensively *O*-GlcNAc modified[7]. So far, only about 50 sites of *O*-GlcNAc addition have been mapped. These have revealed that there is no obvious consensus sequence. However, about 50% of the known sites have a Pro-Val-Ser motif similar to that recognized by proline-directed kinases. Many of the known *O*-GlcNAc sites have high 'PEST' scores — PEST (Pro-Glu-Ser-Thr) being a sequence that is associated with rapid degradation — which suggests that O-GlcNAcylation at these sites might slow or prevent degradation[24].

O-GlcNAc's complex interplay with O-phosphate

O-GlcNAc and *O*-phosphate site-mapping studies suggest that there are at least four different types of dynamic interplay between *O*-GlcNAc and *O*-phosphate (Fig. 3). First, there is competitive occupancy at the same site, for example that which occurs in the transcription factor c-Myc[25] and oestrogen receptor-β[26], and on the oncoprotein SV-40 large T-antigen[27] and endothelial nitric oxide synthase[28]. Second, competitive and alternative occupancy occur at adjacent sites, such as that observed in the tumour suppressor p53 (ref. 29) and synapsin I (ref. 30). Third, there is a complex interplay whereby some *O*-phosphate attachment sites on a given protein are the same as some *O*-GlcNAc sites, whereas others are adjacent to, or even distant from, each other, such as on the C-terminal domain of RNA polymerase II (ref. 31) and on cytokeratins[32]. The final type of interplay involves proteins in which this relationship has yet to be clearly defined. The interplay between *O*-GlcNAc and *O*-phosphate is also underscored by the recent finding that OGT transiently forms complexes containing the catalytic subunit of protein phosphatase 1 (PP1c)[33].

The main *O*-GlcNAc site on the c-Myc oncoprotein was mapped to threonine 58 (ref. 34), a site located within the protein's transactivation domain. Thr 58 is the mutation site responsible for most cases of lymphoma, and is also a glycogen synthase kinase 3 (GSK3) domain site. Inhibition of GSK3 activity reduces phosphorylation at Thr 58, and concomitantly increases O-GlcNAcylation at this site. Serum stimulation of growth-arrested HL60 cells rapidly increases phosphorylation and concomitantly decreases O-GlcNAcylation at Thr 58 (ref. 25). Similarly, studies of oestrogen receptor-β suggest that the alternative O-GlcNAcylation/O-phosphorylation at Ser 16 might act reciprocally to regulate both its activity and its turnover[35].

Site-directed mutagenesis is unable to distinguish between functions mediated by *O*-GlcNAc and *O*-phosphate when they occur at the same, or perhaps even adjacent, amino-acid residues. Conversion of serine or threonine to a non-hydroxyamino acid, such as alanine, simultaneously prevents both modifications. Thus, any biological effect observed cannot be attributed to one or other modification. One way to address this problem is through the *in vitro* synthesis of full-length proteins that are stoichiometrically modified to contain only one or other modification at a given site, followed by the use of protein transfection to identify the functions of each isoform. Recent advances enabling the synthesis of these site-specific protein isoforms[36] should allow more rapid progress in the future.

Adjacent occupancy of a protein region by either *O*-GlcNAc or *O*-phosphate can affect protein function by each modification influencing the cycling of the other, or by their affect on other post-translational modifications. For example, *O*-GlcNAc at Ser 149 of p53 reduces phosphorylation at Thr 155, and also reduces ubiquitin-mediated degradation of p53 (ref. 29). Phosphorylation can prevent O-GlcNAcylation
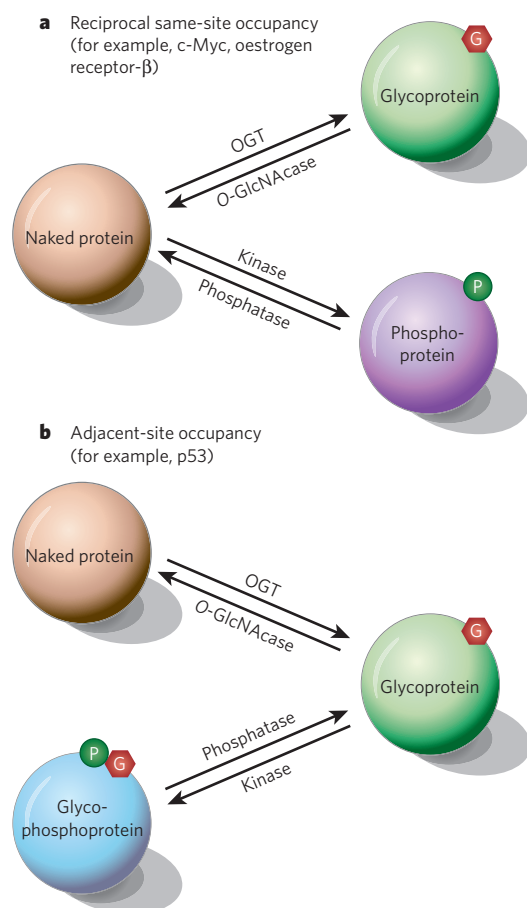
a Reciprocal same-site occupancy (for example, c-Myc, oestrogen receptor-β)

b Adjacent-site occupancy (for example, p53)

Figure 3 | There is a complex and dynamic interplay between _O_-GlcNAc and _O_-phosphate. **a**, Site-mapping studies have shown that on some proteins _O_-GlcNAc and _O_-phosphate compete dynamically for the same serine or threonine hydroxyl moiety. This reciprocal occupancy seems to produce different activities or stability in the proteins. Two examples are Thr 58 of the c-Myc oncoprotein and Ser 16 of oestrogen receptor-β. **b**, In certain proteins, _O_-GlcNAc and _O_-phosphate can also occur next to each other. Adjacent occupancy by each modification reciprocally influences the functions or turnover of proteins, as has recently been reported for the tumour suppressor p53 (ref. 29).

at an adjacent site. For example, a serine to alanine mutation at position 62 of the ERK kinase site of c-Myc increases O-GlcNAcylation at Thr 58 (ref. 25). An even more complex situation occurs when _O_-GlcNAc or _O_-phosphate, or both, competitively occupy two adjacent sites, as occurs in the C-terminal domain of RNA polymerase II (ref. 37). _In vitro_, the presence of a single phosphate residue on the C-terminal domain (10 repeats) prevents it from being a substrate for OGT. Similarly, a single _O_-GlcNAc on the C-terminal domain greatly reduces its acceptor activity for C-terminal-domain kinase, even though numerous potential sites remain unmodified in either case[31]. The dynamic and complex interplay between _O_-GlcNAc and _O_-phosphate requires modification of the current dogma with respect to cellular signalling. Clearly, the system is not binary with an 'on' or 'off' state for each signalling molecule controlled by phosphate occupancy alone. Rather, the combination of modifications, which are differentially regulated, creates enormous molecular diversity (Fig. 4). This diversity could permit almost continuous 'analogue' control of a protein's functions. For example, the extent of occupancy at multiple sites by either sort of modification can rapidly provide a continuous threshold for distinct molecular interactions.

Enzymes regulating _O_-GlcNAc cycling

OGT (uridine diphospho-_N_-acetylglucosamine:polypeptide β-_N_-acetyl-glucosaminyltransferase) catalyses the addition of _O_-GlcNAc to

proteins (Fig. 1). OGT was identified in and first purified from rat liver[38], and has since been cloned in rats, humans, _C. elegans_[18,19] and other organisms. It is highly conserved from _C. elegans_ to humans. In humans, OGT resides near the centromere at Xq13, a chromosomal region linked to several neurological diseases, including Parkinson's dystonia. Rat liver OGT is a heterotrimer with two 110-kDa subunits and one 78-kDa subunit, but in most tissues, including the brain and the pancreas, the enzyme does not require the 78-kDa subunit to be active. A smaller mitochondrial form of OGT has also been described[39].

OGT is present in all cells, although it is most abundant in the glucose-sensing cells of the pancreas and in the brain. The catalytic site of OGT is within its C-terminal domain, which seems to have evolved from the glycogen phosphorylase superfamily[40]. OGT's amino terminus contains several tetratricopeptide repeats (TPRs), which function as protein–protein interaction domains. The crystal structure of OGT's TPR domain has remarkable structural similarity to the armadillo repeats of importin-α, a nuclear transport protein, suggesting that these repeats might have a similar binding/targeting mechanism in both proteins[41].

OGT regulation is complex, and many questions still need to be answered. It differs markedly from the regulation of kinases. Mammals have only a single catalytic subunit for OGT. By contrast, mammalian genomes encode hundreds of site-specific kinases. Although _in vitro_ studies with small peptide substrates have shown that OGT is highly specific for peptide sequence, OGT's specificity _in vivo_ seems to be largely determined by accessory proteins, which target its catalytic subunits. The limited number of targeting proteins studied so far seem to associate dynamically with OGT's TPR domains. Thus OGT's protein specificity seems to be regulated similarly to that of phosphatases, such as PP1c[42]. For example, binding to the transcriptional repressor Sin3A (ref. 43), or to the RNA-polymerase-II-targeting protein OIP106 (ref. 44), is important for OGT's activity on transcription proteins. OGT's overall catalytic activity is controlled by the concentrations of its donor substrate, UDP-GlcNAc, over a broad range. Concentrations of UDP-GlcNAc are highly sensitive to fluxes in nutrients, energy and metabolic nitrogen, making it an ideal metabolic sensor. Rapid changes
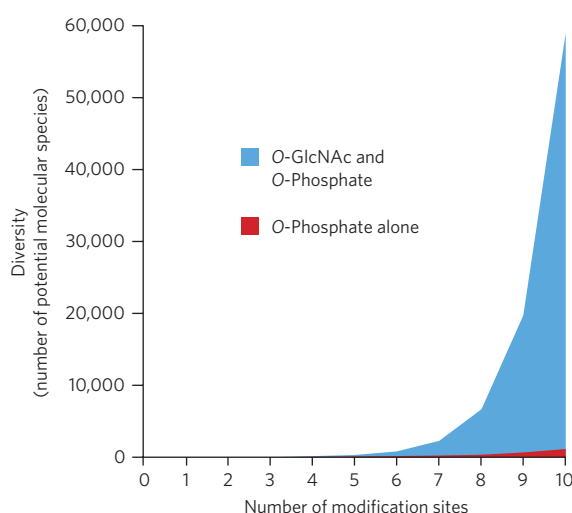

Figure 4 | The dynamic interplay between _O_-GlcNAc and _O_-phosphate enables the rapid generation of enormous molecular diversity in response to cellular stimuli. The cycling of two post-translational modifications that are independently regulated, such as that of O-GlcNAcylation and phosphorylation, on the same polypeptide, provides the cell with a rapid-response mechanism for generating enormous molecular diversity to fine-tune protein interactions and functions. As the number (n) of potential modification sites increases, the potential molecular diversity increases drastically. This example illustrates the maximal theoretical molecular diversity if the two modifications independently and reciprocally compete for the same sites. Red, molecular diversity = 2^n; blue, molecular diversity = 3^n.

1019

in UDP-GlcNAc concentration serve as a sensor by directly affecting the extent of O-GlcNAcylation of regulatory proteins. In addition, OGT is both O-GlcNAcylated and tyrosine phosphorylated.

Nucleocytoplasmic β-*N*-acetylglucosaminidase (O-GlcNAcase) was first identified as 'hexosaminidase C'[45]. O-GlcNAcase was purified from the rat kidney and bovine brain, and the human gene was cloned using peptide sequencing[46]. The cloned sequence was found to be identical to that of *MGEA5* (meningioma-expressed antigen 5)[47], which had been identified genetically in human meningiomas. *MGEA5* maps to a chromosomal region associated with late-onset Alzheimer's disease in humans (10q24.1). O-GlcNAcase seems to be a bifunctional enzyme. The N terminus encodes the glycosidase domain, whereas the C terminus shows homology to the GCN5 histone acetyltransferases (HATs). O-GlcNAcase is reported to have both HAT and O-GlcNAcase activities *in vitro*[48]. *In vitro*, caspase-3 cleaves O-GlcNAcase to separate the HAT and O-GlcNAcase domains, and the latter domain remains catalytically active[49]. O-GlcNAcase also interacts with a vast number of cellular proteins, but the location of interaction sites on the enzyme and the functional significance of these interactions have not yet been determined. O-GlcNAcase and OGT sometimes occur within the same functional complex. In these complexes, both enzymes must be tightly regulated to prevent futile cycling of O-GlcNAc.

Biological functions of *O*-GlcNAc

O-GlcNAcylation not only has an important role in many fundamental cellular processes, but also its dysregulation contributes to the aetiology of important human diseases, particularly diabetes (Box 1) and neurological disorders (Table 1).

Regulation of transcription

RNA polymerase II and basal or specific transcription factors are extensively modified by O-GlcNAc. O-GlcNAcylation can either suppress or enhance transcription, depending on the promoter involved and other associated proteins. O-GlcNAc directly regulates the activities of various transcription factors (see refs 6, 7 for reviews).

Hyper-O-GlcNAcylation of the transcription factor SP1, which occurs in diabetes, enhances the transcription of genes characteristic of this diseased state, but suppresses transcription of other genes. For example, hyper-O-GlcNAcylated SP1 drives transcription of plasminogen activator and extracellular matrix proteins that have an important role in diabetic cardiovascular disease. O-GlcNAcylation of YY1, a developmentally important transcription factor, blocks binding to the retinoblastoma protein[6,7]. ELF-1, which regulates the transcription of several haematopoietic genes, is both O-GlcNAcylated and phosphorylated by protein kinase-Cθ before translocation to the nucleus, where it promotes transcription of T-cell-receptor-ζ. The transcription factor STAT5A requires O-GlcNAcylation to bind to the co-activator of transcription, CREB-binding protein (CBP). Increased O-GlcNAcylation of the pancreatic β-cell transcription factor PDX-1, which regulates insulin transcription, increases its affinity for DNA. O-GlcNAcylation of CREB (cyclic-AMP-responsive element-binding protein), a transcription factor that has a role in long-term memory, diminishes its interaction with TAF$_{II}$130, a component of the basal transcription machinery, thus repressing its transcriptional activity[7]. O-GlcNAcylation of oestrogen receptors and of SP1 may prevent their rapid degradation[7]. Therefore, although much remains unknown about the functions of O-GlcNAcylation in transcription, this modification is clearly important, justifying more detailed studies.

Protein trafficking and turnover

O-GlcNAcylation regulates the trafficking of β-catenin and E-cadherin to the cell surface of epithelial cells[50]. In neurons, the sites of O-GlcNAc attachment on synapsin I (ref. 30) suggest a role for O-GlcNAcylation in modulating synapsin's interactions with the cytoskeleton. Several studies have shown that both direct and indirect O-GlcNAcylation of proteins slows their proteolytic degradation[51]. Proteomic analyses have demonstrated that many proteins of both the catalytic and regulatory cores of the 26S proteasome are modified by O-GlcNAc[52]. Increased O-GlcNAcylation of the RPT2 ATPase, a component of the 19S cap, blocks its ATPase activity, reducing proteasome-catalysed degradation[53].

Box 1 | Dysregulation of *O*-GlcNAc underlies diabetes and glucose toxicity

O-GlcNAc has a key role in the regulation of insulin signalling and as a mediator of glucose toxicity (see refs 62, 63 for reviews; Table 1). Increasing O-GlcNAcylation in adipocytes or muscle blocks insulin signalling at several points. Elevated O-GlcNAcylation on insulin receptor substrate seems to reduce its interactions with phosphatidylinositol-3-OH kinase, thus blocking insulin signalling at an early stage[64,65]. In transgenic mouse models, OGT overexpression in muscle or adipose tissue causes overt diabetes[66]. Many aspects of 'glucose toxicity' require its metabolism into glucosamine, which, in turn, elevates UDP-GlcNAc and increases O-GlcNAcylation. Hyperglycaemia, hyperlipidaemia and/or hyperinsulinaemia all result in abnormal increases in O-GlcNAcylation, which disturbs the normal dynamic balance between O-GlcNAcylation and O-phosphorylation that controls signalling, transcription and other cellular functions (Table 1). It is proposed that, when prolonged, this imbalance contributes to the chronic and pleiotropic toxicity associated with diabetes.

Since the 1950s more than 1,200 papers have linked hexosamine metabolism with diabetes. However, Marshall and colleagues were the first to directly link glucosamine metabolism to insulin resistance (the hallmark of type 2 diabetes)[67]. In cultured adipocytes,

glucosamine is much more potent than glucose in inducing insulin-resistance. Glucose's ability to induce insulin resistance is prevented by deoxynorleucine (DON), an inhibitor of glucose: fructose amidotransferase (GFAT), the key enzyme that converts fructose-6-phosphate to glucosamine-6-phosphate[66]. Many later studies led to the current models for the roles of both hexosamine metabolism and O-GlcNAc in the aetiology of diabetes. Glucosamine metabolism is required for fat-induced insulin resistance *in vivo*. O-GlcNAcylation of glycogen synthase reduces its activation by insulin[68]. A single-nucleotide polymorphism commonly occurs in the *MGEA5* gene (which encodes O-GlcNAcase) within a human population with type 2 diabetes[69]. In rat skeletal muscle, hyperinsulinaemia, as seen in type 2 diabetes, markedly increases the levels of O-GlcNAc on many proteins[70].

Elevated glucose levels increase the O-GlcNAcylation of proteins in the insulin-secreting pancreatic β-cells[71]. Hyperglycaemia qualitatively and quantitatively alters the expression of many O-GlcNAcylated proteins in rat aorta and in cultured rat aortic smooth muscle, suggesting a role for O-GlcNAc in glucose toxicity to vascular tissues[62,63]. Transgenic mice overexpressing the glucose transporter protein GLUT1 in muscle are insulin

resistant and have increased O-GlcNAcylation of GLUT4 vesicle-associated proteins[63]. Non-esterified fatty acids increase both UDP-GlcNAc levels and the O-GlcNAc-dependent DNA-binding activity of the SP1 transcription factor in cultured human myotubes (see ref. 63 for a recent review). Cultured myotubes grown in high glucose and/or insulin have enhanced O-GlcNAcylation of numerous proteins[62,63,70]. Glycoproteomic analyses of these myotubes show that hyper-O-GlcNAcylated proteins include HSP70, α-tubulin and SP1 (ref. 72). Rat myocytes exposed to high glucose levels exhibit altered O-GlcNAcylation, which contributes to impaired calcium cycling in the heart and to diabetic cardiomyopathy by reducing the transcription of a sarcoplasmic reticulum calcium ATPase, SERCA2a (ref. 73). When diabetic cardiomyocytes are transfected with an adenovirus encoding O-GlcNAcase, cardiomyocyte functions improve markedly. Thus, it is now clear that hyperglycaemia, hyperlipidaemia and insulin all regulate O-GlcNAcylation, which, in turn, modulates signalling and transcription. Evidence is mounting that dysregulation of O-GlcNAcylation is a major mechanism underlying the molecular basis of glucose toxicity and insulin resistance, the two main hallmarks of type 2 diabetes.

Table 1 | Pathologies associated with aberrant O-GlcNAcylation

Tissue	Proposed mechanism	Pathology
Muscle, fat	Decreased GLUT4 trafficking Decreased insulin signalling Decreased glycogen synthesis	Insulin resistance
Vasculature	Decreased eNOS activity Increased SP1 activation	Vascular disease Erectile dysfunction Retinopathy
Kidney	Increased TGF-β expression	Glomerular sclerosis
Cardiac muscle	Decreased SERCA2a expression	Cardiomyopathy
Brain	Increased tau phosporylation Decreased tau O-GlcNAcylation	Alzheimer's disease

Increased flux through the hexosamine biosynthetic pathway due to nutrient excess and/or stress can result in elevated O-GlcNAcylation of proteins contributing to various disease states. For example, altered O-GlcNAcylation has been associated with glucose toxicity and contributes to insulin resistance in diabetes. Diabetes-associated pathologies can result from abnormal transcriptional control and the inability of kinase cascades to activate O-GlcNAcylated enzymes. In the case of Alzheimer's disease, it has been proposed that decreased glucose availability to the brain causes a decrease in tau O-GlcNAcylation, allowing tau to become abnormally hyperphosphorylated. eNOS, endothelial nitric oxide synthase; GLUT4, glucose transporter 4; SERCA2a, sarcoplasmic reticulum Ca^{2+}-transport ATPase; TGF-β, transforming growth factor-β.

Neurodegenerative disease

As stated above, the genes encoding both OGT and O-GlcNAcase map to chromosomal regions associated with neurodegenerative disease (see ref. 7 for a review). Tau is required for microtubule polymerization and stability in neurons. In neurons affected by Alzheimer's disease, tau becomes hyperphosphorylated, which, in turn, causes it to aggregate into the paired helical filaments that constitute the visible neurofibrillary tangles characteristic of the disease. In the healthy adult brain, tau is extensively O-GlcNAcylated at more than 12 sites. O-GlcNAcylation of human tau negatively regulates its O-phosphorylation in a site-specific manner both *in vitro* and *in vivo*[54]. Neuron glucose metabolism declines with age, and the decline is more marked in Alzheimer's disease. One normal function of O-GlcNAc seems to be the prevention of excessive phosphorylation at sites at which this would be deleterious. Most proteins involved in the pathology of Alzheimer's disease are O-GlcNAcylated. Targeted deletion of OGT in mouse neurons results in tau hyperphosphorylation similar to that found in Alzheimer's disease[21]. An animal model of mice subjected to starvation also mimicked the hypo-O-GlcNAcylation and hyper-O-phosphorylation of tau seen in the brains of humans with Alzheimer's disease[54]. Comparisons between brain tissue from healthy individuals and those with Alzheimer's disease have shown that global O-GlcNAcylation is reduced in Alzheimer's disease.

The amyloid-β precursor protein (APP) is O-GlcNAcylated[55] and its cytoplasmic tail is phosphorylated, which is known to affect its proteolytic processing. Abnormal proteolysis of APP gives rise to the toxic β1–42 peptide fragment, which forms amyloid plaques that are found in neurons in Alzheimer's disease. Neurofilaments are hypo-O-GlcNAcylated in neurons in a rat model of amyotrophic lateral sclerosis (Lou Gehrig's disease)[56]. O-GlcNAcylation of the clathrin assembly proteins AP-3 and AP-180 declines in Alzheimer's disease, suggesting that reduced O-GlcNAcylation contributes to the loss of synaptic vesicle recycling. Neuron-targeted deletion of OGT in mice showed that O-GlcNAc is essential for neuronal function and survival[21]. Current data suggest that O-GlcNAc has a significant role in normal neuronal function, and in the pathology of neurodegenerative disease.

Stress survival and the cell cycle

In every mammalian cell type examined so far, one of the earliest responses to cellular stress is a rapid and global increase in O-GlcNAcylation on many proteins[57]. Stress-induced increases in O-GlcNAc are dynamic, returning to normal after 24–48 hours. Experimentally decreasing OGT and/or O-GlcNAc levels results in cells that are less stress tolerant, whereas increasing the levels of OGT and/or O-GlcNAc results in cells that are more viable after severe stress. Thus, short-term increase in O-GlcNAcylation seems to be an important survival mechanism[7,57]. O-GlcNAc mediates stress tolerance in part by altering

the levels of heat-shock proteins (HSPs). However, O-GlcNAc might protect cells through several other mechanisms. It might do this by stabilizing protein structure, or by preventing protein–protein aggregation, possibly directly or through HSP70's O-GlcNAc-binding activity[58]. It may modulate signal-transduction pathways previously implicated in glucose-dependent protection, such as JNK activation. Or it may alter the activity or localization of HSPs, many of which are known to be O-GlcNAcylated. Whatever the underlying mechanisms, increased O-GlcNAcylation on many proteins is one of the most rapid cellular responses to stress from a wide variety of unrelated sources.

O-GlcNAcylation also acts as a cellular regulator of growth and proliferation. OGT deletion in mouse embryonic fibroblasts causes growth delays, increased cyclin inhibitor p27 levels, and cell death[21]. *Xenopus* oocyte maturation is delayed by galactose capping of O-GlcNAc (through the action of microinjected β-galactosyltransferase) or by incubation with an O-GlcNAcase inhibitor[59,60]. Overexpression of OGT or O-GlcNAcase causes mitotic exit defects in HeLa cells, and disrupts the normal timing of cyclin protein expression and mitotic phosphorylation[61]. OGT localizes with the mitotic spindle in early metaphase but relocates to the midbody during telephase. OGT overexpression blocks cytokinesis, causing polyploidy, suggesting that O-GlcNAcylation has a direct role in the late stages of cell division[61].

Future directions and key unanswered questions

The most important impediment to research on O-GlcNAc is the lack of sensitive and easy-to-use tools for its detection, quantification and site localization. Future developments in mass-spectrometric instrumentation and in the generation of O-GlcNAc-dependent, site-specific antibodies[25,56] will greatly facilitate efforts in this area of research. Some major unanswered questions remain, including: what are the specific relationships and interactions between O-GlcNAc and O-phosphate in complex biological systems? What are the molecular mechanistic roles of O-GlcNAc in specific signalling cascades and in transcriptional regulation? How are the O-GlcNAc-cycling enzymes regulated and targeted? How does increased O-GlcNAcylation lead to insulin resistance, and what are O-GlcNAc's specific roles in glucose toxicity? Elucidation of O-GlcNAc's specific roles in transcription, signalling, glucose toxicity and insulin resistance should lead to new avenues for the diagnosis and treatment of diabetes and other diseases. The interplay between O-GlcNAc and O-phosphate seems to be critical to the aetiology of neurodegenerative disease, as well as to the functions and turnover of oncoproteins or tumour suppressors. ∎

1. Torres, C. R. & Hart, G. W. Topography and polypeptide distribution of terminal N-acetylglucosamine residues on the surfaces of intact lymphocytes. Evidence for O-linked GlcNAc. *J. Biol. Chem.* **259,** 3308–3317 (1984).
2. Hart, G. W., Haltiwanger, R. S., Holt, G. D. & Kelly, W. G. Glycosylation in the nucleus and cytoplasm. *Annu. Rev. Biochem.* **58,** 841–874 (1989).
3. Holt, G. D. & Hart, G. W. The subcellular distribution of terminal N-acetylglucosamine moieties. Localization of a novel protein–saccharide linkage, O-linked GlcNAc. *J. Biol. Chem.* **261,** 8049–8057 (1986).
4. Hart, G. W. Dynamic O-linked glycosylation of nuclear and cytoskeletal proteins. *Annu. Rev. Biochem.* **66,** 315–335 (1997).
5. Wells, L., Vosseller, K. & Hart, G. W. Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science* **291,** 2376–2378 (2001).
6. Love, D. C. & Hanover, J. A. The hexosamine signaling pathway: deciphering the 'O-GlcNAc code'. *Sci. STKE* **2005,** re13 (2005).
7. Zachara, N. E. & Hart, G. W. Cell signaling, the essential role of O-GlcNAc! *Biochim. Biophys. Acta* **1761,** 599–617 (2006).
8. Slawson, C., Housley, M. P. & Hart, G. W. O-GlcNAc cycling: how a single sugar post-translational modification is changing the way we think about signaling networks. *J. Cell. Biochem.* **97,** 71–83 (2006).
9. McClain, D. A. & Crook, E. D. Hexosamines and insulin resistance. *Diabetes* **45,** 1003–1009 (1996).
10. Greis, K. D. & Hart, G. W. Analytical methods for the study of O-GlcNAc glycoproteins and glycopeptides. *Methods Mol. Biol.* **76,** 19–33 (1998).
11. Stubbs, K. A., Zhang, N. & Vocadlo, D. J. A divergent synthesis of 2-acyl derivatives of PUGNAc yields selective inhibitors of O-GlcNAcase. *Org. Biomol. Chem.* **4,** 839–845 (2006).
12. Horsch, M., Hoesch, L., Vasella, A. & Rast, D. M. N-acetylglucosamino-1,5-lactone oxime and the corresponding (phenylcarbamoyl)oxime. Novel and potent inhibitors of beta-N-acetylglucosaminidase. *Eur. J. Biochem.* **197,** 815–818 (1991).
13. Comer, F. I., Vosseller, K., Wells, L., Accavitti, M. A. & Hart, G. W. Characterization of a mouse monoclonal antibody specific for O-linked N-acetylglucosamine. *Anal. Biochem.* **293,** 169–177 (2001).

14. Syka, J. E., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl Acad. Sci. USA* **101,** 9528–9533 (2004).

15. Vocadlo, D. J., Hang, H. C., Kim, E. J., Hanover, J. A. & Bertozzi, C. R. A chemical approach for identifying O-GlcNAc-modified proteins in cells. *Proc. Natl Acad. Sci. USA* **100,** 9116–9121 (2003).

16. Wells, L. *et al.* Mapping sites of O-GlcNAc modification using affinity tags for serine and threonine post-translational modifications. *Mol. Cell. Proteomics* **1,** 791–804 (2002).

17. Hartweck, L. M., Genger, R. K., Grey, W. M. & Olszewski, N. E. SECRET AGENT and SPINDLY have overlapping roles in the development of *Arabidopsis thaliana* L. Heyn. *J. Exp. Bot.* **57,** 865–875 (2006).

18. Kreppel, L. K., Blomberg, M. A. & Hart, G. W. Dynamic glycosylation of nuclear and cytosolic proteins. Cloning and characterization of a unique O-GlcNAc transferase with multiple tetratricopeptide repeats. *J. Biol. Chem.* **272,** 9308–9315 (1997).

19. Lubas, W. A., Frank, D. W., Krause, M. & Hanover, J. A. O-Linked GlcNAc transferase is a conserved nucleocytoplasmic protein containing tetratricopeptide repeats. *J. Biol. Chem.* **272,** 9316–9324 (1997).

20. Shafi, R. *et al.* The O-GlcNAc transferase gene resides on the X chromosome and is essential for embryonic stem cell viability and mouse ontogeny. *Proc. Natl Acad. Sci. USA* **97,** 5735–5739 (2000).

21. O'Donnell, N., Zachara, N. E., Hart, G. W. & Marth, J. D. *Ogt*-dependent X-chromosome-linked protein glycosylation is a requisite modification in somatic cell function and embryo viability. *Mol. Cell. Biol.* **24,** 1680–1690 (2004).

22. Hanover, J. A. *et al.* A *Caenorhabditis elegans* model of insulin resistance: altered macronutrient storage and dauer formation in an OGT-1 knockout. *Proc. Natl Acad. Sci. USA* **102,** 11266–11271 (2005).

23. Forsythe, M. E. *et al. Caenorhabditis elegans* ortholog of a diabetes susceptibility locus: oga-1 (O-GlcNAcase) knockout impacts O-GlcNAc cycling, metabolism, and dauer. *Proc. Natl Acad. Sci. USA* **103,** 11952–11957 (2006).

24. Rechsteiner, M. & Rogers, S. W. PEST sequences and regulation by proteolysis. *Trends Biochem. Sci.* **21,** 267–271 (1996).

25. Kamemura, K., Hayes, B. K., Comer, F. I. & Hart, G. W. Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: alternative glycosylation/phosphorylation of Thr-58, a known mutational hot spot of c-Myc in lymphomas, is regulated by mitogens. *J. Biol. Chem.* **277,** 19229–19235 (2002).

26. Cheng, X., Cole, R. N., Zaia, J. & Hart, G. W. Alternative O-glycosylation/O-phosphorylation of the murine estrogen receptor beta. *Biochemistry* **39,** 11609–11620 (2000).

27. Medina, L., Grove, K. & Haltiwanger, R. S. SV40 large T antigen is modified with O-linked N-acetylglucosamine but not with other forms of glycosylation. *Glycobiology* **8,** 383–391 (1998).

28. Du, X. L. *et al.* Hyperglycemia inhibits endothelial nitric oxide synthase activity by posttranslational modification at the Akt site. *J. Clin. Invest.* **108,** 1341–1348 (2001).

29. Yang, W. H. *et al.* Modification of p53 with O-linked N-acetylglucosamine regulates p53 activity and stability. *Nature Cell Biol.* **8,** 1074–1083 (2006).

30. Cole, R. N. & Hart, G. W. Glycosylation sites flank phosphorylation sites on synapsin I: O-linked N-acetylglucosamine residues are localized within domains mediating synapsin I interactions. *J. Neurochem.* **73,** 418–428 (1999).

31. Comer, F. I. & Hart, G. W. Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry* **40,** 7845–7852 (2001).

32. Chou, C. F., Smith, A. J. & Omary, M. B. Characterization and dynamics of O-linked glycosylation of human cytokeratin 8 and 18. *J. Biol. Chem.* **267,** 3901–3906 (1992).

33. Wells, L., Kreppel, L. K., Comer, F. I., Wadzinski, B. E. & Hart, G. W. O-GlcNAc transferase is in a functional complex with protein phosphatase 1 catalytic subunit. *J. Biol. Chem.* **279,** 38466–38470 (2004).

34. Chou, T.-Y., Hart, G. W. & Dang, C. V. c-Myc is glycosylated at threonine 58, a known phosphorylation site and a mutational hot spot in lymphomas. *J. Biol. Chem.* **270,** 18961–18965 (1995).

35. Cheng, X. G. & Hart, G. W. Alternative O-glycosylation/O-phosphorylation of serine-16 in murine estrogen receptor β. Post-translational regulation of turnover and transactivation activity. *J. Biol. Chem.* **276,** 10570–10575 (2001).

36. Zhang, Z. *et al.* A new strategy for the synthesis of glycoproteins. *Science* **303,** 371–373 (2004).

37. Kelly, W. G., Dahmus, M. E. & Hart, G. W. RNA polymerase II is a glycoprotein. Modification of the COOH–terminal domain by O-GlcNAc. *J. Biol. Chem.* **268,** 10416–10424 (1993).

38. Haltiwanger, R. S., Blomberg, M. A. & Hart, G. W. Glycosylation of nuclear and cytoplasmic proteins. Purification and characterization of a uridine diphospho-N-acetylglucosamine:polypeptide β-N-acetylglucosaminyltransferase. *J. Biol. Chem.* **267,** 9005–9013 (1992).

39. Hanover, J. A. *et al.* Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc transferase encoded by a single mammalian gene. *Arch. Biochem. Biophys.* **409,** 287–297 (2003).

40. Wrabl, J. O. & Grishin, N. V. Homology between O-linked GlcNAc transferases and proteins of the glycogen phosphorylase superfamily. *J. Mol. Biol.* **314,** 365–374 (2001).

41. Jinek, M. *et al.* The superhelical TPR-repeat domain of O-linked GlcNAc transferase exhibits structural similarities to importin α. *Nature Struct. Mol. Biol.* **11,** 1001–1008 (2004).

42. Cohen, P. T. W. Protein phosphatase 1 — targeted in many directions. *J. Cell Sci.* **115,** 241–256 (2002).

43. Yang, X., Zhang, F. & Kudlow, J. E. Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: coupling protein O-GlcNAcylation to transcriptional repression. *Cell* **110,** 69–80 (2002).

44. Iyer, S. P., Akimoto, Y. & Hart, G. W. Identification and cloning of a novel family of coiled-coil domain proteins that interact with O-GlcNAc transferase. *J. Biol. Chem.* **278,** 5399–5409 (2003).

45. Braidman, I. *et al.* Characterisation of human N-acetyl-beta-hexosaminidase C. *FEBS Lett.* **41,** 181–184 (1974).

46. Gao, Y., Wells, L., Comer, F. I., Parker, G. J. & Hart, G. W. Dynamic O-glycosylation of nuclear and cytosolic proteins: cloning and characterization of a neutral, cytosolic β-N-acetylglucosaminidase from human brain. *J. Biol. Chem.* **276,** 9838–9845 (2001).

47. Heckel, D. *et al.* Novel immunogenic antigen homologous to hyaluronidase in meningioma. *Hum. Mol. Genet.* **7,** 1859–1872 (1998).

48. Toleman, C., Paterson, A. J., Whisenhunt, T. R. & Kudlow, J. E. Characterization of the histone acetyltransferase (HAT) domain of a bifunctional protein with activable O-GlcNAcase and HAT activities. *J. Biol. Chem.* **279,** 53665–53673 (2004).

49. Wells, L. *et al.* Dynamic O-glycosylation of nuclear and cytosolic proteins: further characterization of the nucleocytoplasmic β-N-acetylglucosaminidase, O-GlcNAcase. *J. Biol. Chem.* **277,** 1755–1761 (2002).

50. Zhu, W., Leber, B. & Andrews, D. W. Cytoplasmic O-glycosylation prevents cell surface transport of E-cadherin during apoptosis. *EMBO J.* **20,** 5999–6007 (2001).

51. Han, I. & Kudlow, J. E. Reduced O glycosylation of Sp1 is associated with increased proteasome susceptibility. *Mol. Cell. Biol.* **17,** 2550–2558 (1997).

52. Sumegi, M., Hunyadi-Gulyas, E., Medzihradszky, K. F. & Udvardy, A. 26S proteasome subunits are O-linked N-acetylglucosamine-modified in *Drosophila melanogaster*. *Biochem. Biophys. Res. Commun.* **312,** 1284–1289 (2003).

53. Zhang, F. *et al.* O-GlcNAc modification is an endogenous inhibitor of the proteasome. *Cell* **115,** 715–725 (2003).

54. Liu, F., Iqbal, K., Grundke-Iqbal, I., Hart, G. W. & Gong, C. X. O-GlcNAcylation regulates phosphorylation of tau: a mechanism involved in Alzheimer's disease. *Proc. Natl Acad. Sci. USA* **101,** 10804–10809 (2004).

55. Griffith, L. S., Mathes, M. & Schmitz, B. Beta-amyloid precursor protein is modified with O-linked N-acetylglucosamine. *J. Neurosci. Res.* **41,** 270–278 (1995).

56. Ludemann, N. *et al.* O-glycosylation of the tail domain of neurofilament protein M in human neurons and in spinal cord tissue of a rat model of amyotrophic lateral sclerosis (ALS). *J. Biol. Chem.* **280,** 31648–31658 (2005).

57. Zachara, N. E. & Hart, G. W. O-GlcNAc a sensor of cellular state: the role of nucleocytoplasmic glycosylation in modulating cellular function in response to nutrition and stress. *Biochim. Biophys. Acta* **1673,** 13–28 (2004).

58. Guinez, C., Lemoine, J., Michalski, J. C. & Lefebvre, T. 70-kDa-heat shock protein presents an adjustable lectinic activity towards O-linked N-acetylglucosamine. *Biochem. Biophys. Res. Commun.* **319,** 21–26 (2004).

59. Fang, B. & Miller, M. W. Use of galactosyltransferase to assess the biological function of O-linked N-acetyl-D-glucosamine: a potential role for O-GlcNAc during cell division. *Exp. Cell Res.* **263,** 243–253 (2001).

60. Slawson, C., Shafii, S., Amburgey, J. & Potter, R. Characterization of the O-GlcNAc protein modification in *Xenopus laevis* oocyte during oogenesis and progesterone-stimulated maturation. *Biochim. Biophys. Acta* **1573,** 121–129 (2002).

61. Slawson, C. *et al.* Perturbations in O-linked β-N-acetylglucosamine protein modification cause severe defects in mitotic progression and cytokinesis. *J. Biol. Chem.* **280,** 32944–32956 (2005).

62. Wells, L., Vosseller, K. & Hart, G. W. A role for N-acetylglucosamine as a nutrient sensor and mediator of insulin resistance. *Cell. Mol. Life Sci.* **60,** 222–228 (2003).

63. Buse, M. G. Hexosamines, insulin resistance, and the complications of diabetes: current status. *Am. J. Physiol. Endocrinol. Metab.* **290,** E1–E8 (2006).

64. Vosseller, K., Wells, L., Lane, M. D. & Hart, G. W. Elevated nucleocytoplasmic glycosylation by O-GlcNAc results in insulin resistance associated with defects in Akt activation in 3T3-L1 adipocytes. *Proc. Natl Acad. Sci. USA* **99,** 5313–5318 (2002).

65. Federici, M. *et al.* Insulin-dependent activation of endothelial nitric oxide synthase is impaired by O-linked glycosylation modification of signaling proteins in human coronary endothelial cells. *Circulation* **106,** 466–472 (2002).

66. McClain, D. A. *et al.* Altered glycan-dependent signaling induces insulin resistance and hyperleptinemia. *Proc. Natl Acad. Sci. USA* **99,** 10695–10699 (2002).

67. Marshall, S., Bacote, V. & Traxinger, R. R. Discovery of a metabolic pathway mediating glucose-induced desensitization of the glucose transport system. Role of hexosamine biosynthesis in the induction of insulin resistance. *J. Biol. Chem.* **266,** 4706–4712 (1991).

68. Parker, G., Taylor, R., Jones, D. & McClain, D. Hyperglycemia and inhibition of glycogen synthase in streptozotocin-treated mice: role of O-linked N-acetylglucosamine. *J. Biol. Chem.* **279,** 20636–20642 (2004).

69. Lehman, D. M. *et al.* A single nucleotide polymorphism in MGEA5 encoding O-GlcNAc-selective N-acetyl-beta-D glucosaminidase is associated with type 2 diabetes in Mexican Americans. *Diabetes* **54,** 1214–1221 (2005).

70. Yki-Jarvinen, H., Virkamaki, A., Daniels, M. C., McClain, D. & Gottschalk, W. K. Insulin and glucosamine infusions increase O-linked N-acetyl-glucosamine in skeletal muscle proteins *in vivo. Metabolism* **47,** 449–455 (1998).

71. Liu, K., Paterson, A. J., Chin, E. & Kudlow, J. E. Glucose stimulates protein modification by O-linked GlcNAc in pancreatic β cells: linkage of O-linked GlcNAc to β cell death. *Proc. Natl Acad. Sci. USA* **97,** 2820–2825 (2000).

72. Walgren, J. L., Vincent, T. S., Schey, K. L. & Buse, M. G. High glucose and insulin promote O-GlcNAc modification of proteins, including α-tubulin. *Am. J. Physiol. Endocrinol. Metab.* **284,** E424–E434 (2003).

73. Clark, R. J. *et al.* Diabetes and the accompanying hyperglycemia impairs cardiomyocyte calcium cycling through increased nuclear O-GlcNAcylation. *J. Biol. Chem.* **278,** 44230–44237 (2003).

Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins

Ajit Varki[1]

All cells in nature are covered by a dense and complex array of carbohydrates. Given their prominence on cell surfaces, it is not surprising that these glycans mediate and/or modulate many cellular interactions. Proteins that bind sialic acid, a sugar that is found on the surface of the cell and on secreted proteins in vertebrates, are involved in a broad range of biological processes, including intercellular adhesion, signalling and microbial attachment. Studying the roles of such proteins in vertebrates has improved our understanding of normal physiology, disease and human evolution.

Among the few seemingly universal biological findings is the fact that all the cells of every species studied so far are covered with a dense coating of glycans (oligosaccharides or polysaccharides), which was originally revealed by electron microscopists and named the 'glycocalyx'[1]. These sugar chains are often covalently attached to underlying proteins or lipids, and form various structures that are cell-type specific and developmentally regulated, typically changing in response to environmental cues[2–4], and causing disease when genetically deficient[5]. In eukaryotic cells, the main mechanisms by which cell surfaces become glycan-coated occur in the endoplasmic reticulum–Golgi biosynthetic pathway[3,4,6,7]. Other complex machineries fulfil similar functions in prokaryotes[8]. The probable evolutionary reasons for this prominence of cell-surface glycosylation are discussed elsewhere[9,10]. Given that glycans have dominated cell surfaces for billions of years, it is reasonable to predict that many cellular interactions are mediated or modulated by these molecules. It should also not be surprising that these biological roles range widely, from highly specific 'lock-and-key' type mechanisms all the way to relatively non-specific effects such as negative charge repulsion — and everything in between.

This review discusses principles involved in understanding and classifying the cellular interactions that are mediated by glycans, and then focuses on sialic acids and the proteins that recognize them. Particular attention is paid to vertebrate sialic-acid-recognizing proteins, such as the selectins and the Siglecs. Such molecules are presented as examples of different models of glycan-mediated cellular interactions, an area of increasing biological interest and importance.

General principles of glycan-mediated cellular interactions

There are various classes of cell-surface glycan, including N- and O-linked glycans, glycosphingolipids, glycosaminoglycans, glycophospholipid anchors and lipo-oligo/polysaccharides. Each of these has characteristic core or linkage regions, and they often present varying types of terminal or internal sequence[2–4]. The structural complexity of glycans is much greater than that of proteins and lipids, and the details of their structural features are not reviewed here. Another unusual feature is that in aqueous solution most glycans present an ensemble of many different conformations, only one of which might be relevant to a given interaction[11]. Interactions involving glycans can include recognition by other glycans (carbohydrate–carbohydrate interactions)[12], as well as recognition by certain proteins[13]. This review focuses on the latter class of interaction.

Most glycan–protein interactions involve the recognition and locking into place of one of a specific glycan's many conformations in the protein's binding pocket. This is, in part, why most glycan-based interactions require multivalency to achieve the avidity sufficient for biologically meaningful interactions[13]. There are also many intermediate situations in which cell-surface glycans have relative specificity for certain surface proteins, modulating rather than dictating their activity[12]. Glycan-binding proteins can themselves be glycosylated, potentially affecting any stage in their life cycle, from their initial folding, their delivery to the cell surface and their ability to interact with other molecules at the surface, to their turnover.

Glycan-binding proteins have traditionally been classified according to the type of glycan they recognize, or on the basis of their biological roles. As discussed elsewhere[14], they may be better classified on the basis of their structural and evolutionary relationships into two broad groups: lectins[13,15], which can be further subclassified into evolutionarily related subgroups with common ancestors, and glycosaminoglycan-binding proteins, which seem to have evolved by convergent evolution to bind the acidic sulphated sequences found in glycosaminoglycans (see page 1030). This evolutionary approach is useful and most logical when seeking to understand orthologous functions between organisms[15]. Glycan-binding proteins comprise a very large group of molecules that cannot be reviewed in the available space. The rest of this review therefore focuses on one class of glycan (the sialic acids) and the proteins in nature that recognize them. Particular attention is then given to vertebrate sialic-acid-recognizing lectins — that is, those intrinsic to the multicellular organisms that synthesize sialic acids.

The sialic acids

The sialic acids are a family of sugars with a shared nine-carbon backbone that are typically found attached to the terminal positions of several classes of cell-surface and secreted glycan molecules (Fig. 1). Although they are probably evolutionarily ancient, sialic acids are most prominent in the deuterostome lineage (which comprises vertebrates and 'higher' invertebrates)[16,17]. Interestingly, several of the rare non-deuterostome organisms that express sialic acids are microbial pathogens that invade vertebrates, apparently because this is a successful form of 'molecular mimicry'[18]. Having emerged as major cell-surface molecules of deuterostomes during the Cambrian expansion about 530 million years ago, the sialic acids have become greatly diversified, such that there are more than 50 structural variations in nature[16,17] (see Fig. 2

[1]Departments of Medicine and Cellular & Molecular Medicine, Glycobiology Research and Training Center, University of California at San Diego, La Jolla, California 92093-0687, USA.
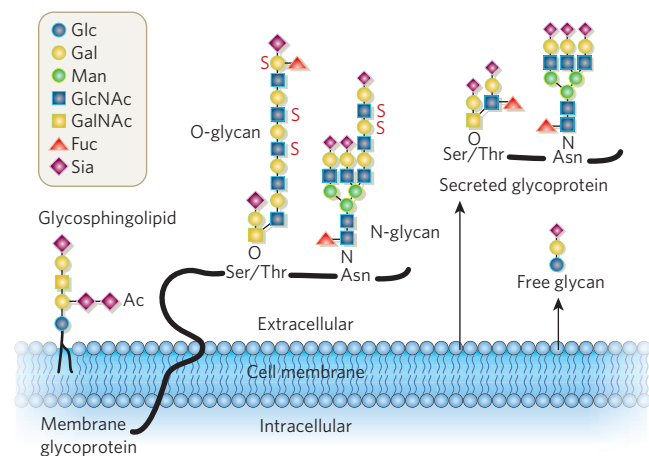
Figure 1 | Sialic acids on cell-surface and secreted molecules. Sialic acids (Sia) are typically found at the terminal position of N- and O-linked glycans attached to the cell surface and to secreted glycoproteins, as well as on glycosphingolipids expressed at the cell surface. Ac, O-acetyl ester; Fuc, fucose; Gal, galactose; GalNAc, N-acetylgalactosamine; Glc, glucose; GlcNAc, N-acetylglucosamine; Man, mannose; S, sulphate ester.

for two examples), each of which can be attached to underlying glycans by means of various α-linkages from the C2 position[19]. This remarkable diversity is expressed in a cell-type and developmentally regulated manner, and often changes in response to environmental cues. Thus, sialic acids are in a position to mediate specific interactions governed by the binding of proteins that are intrinsic to the same species. However, the relative lack of conservation of sialylation patterns even between related species[20] suggests that diversity may also be driven by the pathogens that exploit sialic acids as targets for recognition[9,10,14]. Of special note is the fact that the evolutionary lineage from which humans descended suffered an unusual genetic defect[21] in the production of one of the common Old World primate sialic acids, N-glycolylneuraminic acid (Neu5Gc) — secondarily causing accumulation of the precursor sialic acid N-acetylneuraminic acid (Neu5Ac) (Fig. 2).

Details about sialic acids[16,17] and the usefulness of metabolic incorporation of unnatural sialic acid precursors into mammalian cells[22] are discussed elsewhere. Included in this review are examples of interactions involving the three main classes of vertebrate glycan that carry sialic acids: N-glycans, O-glycans and glycosphingolipids (Fig. 1) — as well as certain bacterial polysaccharides that can display them[18].

Sialic-acid-binding proteins in organisms without sialic acids
Given their terminal location and diversity, it is not surprising that many pathogens have evolved binding proteins or toxins specific for certain sialic acids, often with additional specificity for the linkage and/or the underlying sugar chain (see Table 1 for examples). Mysteriously, many multicellular sialic-acid-deficient organisms (such as plants, arthopods and molluscs) also express sialic-acid-binding proteins, which have a remarkable degree of specificity for certain types of sialic acid and their linkages (Table 1). These lectins might serve to recognize sialic acids on pathogens, or on other species that display them. It is also possible that some of these lectins recognize other related acidic sugars (such as keto-deoxyoctulosonic acids), and have sufficient cross-specificity to be identified and classified as sialic-acid-binding lectins. Regardless of the reasons, such proteins can be extremely useful tools with which to study sialic acids[20,23].

Sialic-acid-binding proteins in organisms with sialic acids
Until the 1980s it was thought that sialic acids functioned primarily to provide negative charge and hydrophilicity to vertebrate cell surfaces, to mask subterminal galactose residues from recognition by certain

receptors, and to act as receptors for pathogens and toxins. Although examples of the latter role are widespread and show elegant structural specificity, they obviously cannot explain the persistence of sialic acids in all deuterostomes for more than 500 million years. Indeed, sialic acids are essential for early embryonic development in mice[24], but not for embryogenesis in *Drosophila* or *Caenorhabditis elegans*. Thus, sialic acids must have intrinsic functions in vertebrates that go beyond providing negative charge and hydrophilicity, and masking underlying residues. In fact, during the past few decades several vertebrate proteins have been found that mediate specific recognition events involving sialic acids[25,26]. The rest of this review focuses on these vertebrate sialic-acid-recognizing proteins, as examples of glycan–protein interactions. These lectins can be either attached to the cell surface or secreted into the extracellular space, and recognition can involve sialic acids on either the same cell or an opposing cell surface.

Factor H
The first vertebrate sialic-acid-recognizing protein discovered was complement factor H, which negatively regulates the alternative complement pathway. This pathway is constantly being activated ('ticking over') in the circulating blood plasma, but amplification is continually held in check by factor H, a protein present at high levels in blood plasma[27]. When plasma is exposed to a complement-activating surface (such as a foreign bacterium) the alternative pathway evades fluid-phase factor H and amplifies itself on the surface to be attacked, initiating a cascade of events in the innate immune system. Given the promiscuous nature of targets for the pathway, the question arose as to why amplification does not also occur on 'self' surfaces. One explanation is the presence of endogenous species-specific complement-controlling proteins on the plasma membranes of host self surfaces. A second mechanism involves soluble factor H being recruited to such surfaces to continue its role of dissociating the alternative pathway convertase enzyme. This recruitment is supported by anion-binding domains of factor H that specifically recognize sialic acids and/or certain sulphated glycosaminoglycans on vertebrate cell surfaces[27–29]. Further studies must define the precise nature of the factor-H interaction with cell-surface sialic acids that mediate this regulatory function. Of course, several successful microbial pathogens also cover their own surfaces with sialic acids[18]. These can recruit host factor H, providing the same protection that host cells enjoy.
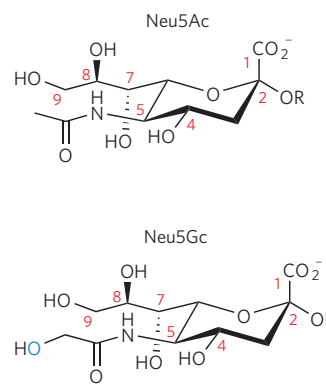

Figure 2 | Two major sialic acids in mammalian cells. Sialic acids share a nine-carbon backbone, a carboxylic acid at the C1 position, and various α-glycosidic linkages to the underlying sugar chain (R) from the C2 position. Various substitutions at the C4, C5, C7, C8 and C9 positions combine with linkage variation to generate the diversity of sialic acids in nature. The structures shown here are the most prevalent sialic acids found in mammalian cells. The only difference between the two is the additional oxygen atom in the N-glycolyl group of Neu5Gc (blue). This modification of the precursor sialic acid Neu5Ac cannot occur in humans because of a genetic mutation that occurred during human evolution.

Selectins

The second group of sialic-acid-binding proteins discovered was the selectins, which are involved in many cell–cell interactions in immunity, haemostasis and inflammation[30,31]. In classic studies, the homing of lymphocytes into lymph nodes through high endothelial venules was shown to be dependent on sialic acids[32]. This turned out to involve recognition of sialylated ligands by leukocyte (L)-selectin, which is constitutively expressed on many types of leukocyte. Independent evidence showed that leukocyte adhesion to activated endothelial cells mediated by endothelial (E)-selectin also involved sialic acids[33]. However, E-selectin is not constitutively expressed, and requires transcriptional induction by various inflammatory stimuli. Parallel studies led to the discovery of platelet (P)-selectin[34], which is stored in the alpha granules of platelets and in Weibel–Palade granules of endothelial cells. This is rapidly mobilized to the cell surface in response to various stimuli, again mediating adhesion to sialic-acid-containing ligands on leukocytes[35,36]. Cloning of the three selectin molecules revealed them to be products of a small cluster of homologous genes[37].
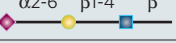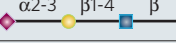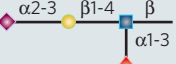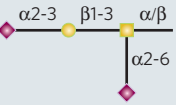
All three types of selectin require sialic acids and nearby α1-3(4)-linked fucose residues for most recognition processes[30,31,33,35,36,38]. This combination is typically found in a terminal glycan motif known as 'sialyl-LewisX' (SLeX; Table 1), which was initially thought to be the shared ligand for all selectins. In fact, SLeX is necessary but not sufficient to generate optimal binding by selectins[38]. As with the Arg-Gly-Asp-Ser motif for integrin recognition, placing SLeX in the context of other structures is important for biologically relevant selectin recognition. Thus, certain leukocyte glycoproteins carrying SLeX in specific arrangements seem to be optimal ligands for E-selectin, and functionally significant recognition by L- or P-selectin requires additional sulphate esters[30,31]. With L-selectin, optimal recognition involves certain sulphated forms of SLeX, with the sulphate typically at the C6 position of N-acetylglucosamine (GlcNAc) residues. This form is constitutively expressed by sialomucin-like proteins of the high endothelial venules of lymph nodes[31,39]. Such ligands are also induced on endothelium at other sites during inflammation.
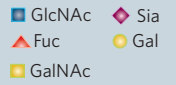
A crucial high-affinity molecule for P-selectin recognition has also been discovered[40], cloned and named P-selectin glycoprotein ligand-1 (PSGL-1)[41]. This sialomucin polypeptide is expressed on all leukocytes, but becomes specialized for selectin recognition on neutrophils and monocytes. In this case, the sulphate component is not on the sugar chain but on adjacent tyrosine residues. Thus, optimal PSGL-1-binding sites for P-selectin (and also for L-selectin) comprise a short O-linked sugar chain bearing an SLeX motif adjacent to two or three sulphated tyrosine residues, all contained within a short sequence at the amino-terminal region of PSGL-1 (refs 42–44). Conclusive evidence for the importance of sialic acid, fucose, PSGL-l and sulphated GlcNAc or tyrosine residues in generating these complex recognition motifs has been obtained by biochemical synthesis[45] and by genetic ablation of genes for the corresponding biosynthetic enzymes in the mouse[39,46]. Many further complexities have since been discovered in the biosynthesis and specialization of the selectin-recognition motifs, and there is now substantial evidence that combinatorial biosynthesis of such recognition motifs is crucial for the formation of biologically relevant selectin ligands.

Meanwhile, studies of selectin-deficient mice[46,47] further confirmed the roles of selectins in the initial interactions among leukocytes, endothelium and platelets, which are involved in mediating various processes in acute and chronic inflammation, in reperfusion injury, and in haemostasis. Owing to space limitations, detailed descriptions are not feasible here. A further fact of biophysical interest is that interactions of P-selectin with PSGL-1 have been shown to involve novel 'catch' bonds, which help to explain the 'rolling' behaviour shown by leukocytes when they encounter endothelial surfaces[48].

A serendipitous finding was that an unrelated glycan (a glycosamino-glycan known as heparan sulphate) can also act as a ligand for L- and P-selectin[49,50], and that a highly sulphated version of this molecule known as heparin is a potent inhibitor of their interactions[50–53]. Although the dominant system of natural selectin recognition seems to be dependent

Table 1 | Selective recognition of sialic acids by sialic-acid-binding proteins

Cognate glycan	Intrinsic	Extrinsic (pathogen-binding)	Extrinsic (unknown function)
α2-6 β1-4 β —R	CD22 (Siglec-2)	Human influenza A	Sambucus nigra agglutinin
α2-3 β1-4 β —R	Sialoadhesin (Siglec-1)	Avian influenza A	Maackia amurensis leukoagglutinin
α2-3 β1-4 β —R, α1-3	E-selectin	Pseudomonas aeruginosa mucoid strain 8830, Anaplasma phagocytophilium	None known
α2-3 β1-3 α/β —R, α2-6	Myelin-associated glycoprotein (Siglec-4)	Plasmodium falciparum Merozoite EBA-175	Maackia amurensis haemagglutinin

■ GlcNAc ◆ Sia ▲ Fuc ● Gal ■ GalNAc

Only a few examples of sialic-acid-bearing glycan termini are shown here, along with examples of cognate sialic-acid-binding proteins. 'Intrinsic' refers to binding proteins made by the same organism that synthesizes the sialylated structures. The third structure is SLeX.

on sialic acid and fucose residues, there is some evidence that heparan-sulphate-mediated interactions can be physiologically significant[53]. How can two such disparate glycan classes be recognized by the same selectin-binding site? A likely explanation is that heparan sulphate and heparin display a high density of carboxylic acid groups, as is found in sialic acids, and a high density of 6-O-sulphated GlcNAc residues, as is found in L-selectin ligands. Thus, some combination of carboxylic acid and sulphate ester groups seems to be presented in 'clustered saccharide patches' by these glycosaminoglycans, mimicking the natural ligand that is presented by PSGL-1 or by some of the sulphated sialomucins recognized by L-selectin[38]. Regardless of whether this is physiologically significant, it has direct relevance to medical practice, as heparin is a well-recognized therapeutic agent and has been used for more than 50 years as an anticoagulant (see page 1046). Because inhibition of L- and P-selectin can be achieved at clinically acceptable heparin concentrations[54,55], these findings have direct therapeutic significance[56], which is currently being pursued. Indeed, it is possible that these and other biological effects of heparin might explain its usefulness as an anti-inflammatory agent[53], in addition to its value as an anticoagulant.

Selectins also have an important role in haematogenous carcinoma metastasis (the spread of epithelial cancers through the bloodstream). SLeX was originally discovered as an antigen prominently expressed on some malignant tumours, particularly tumours of epithelial origin (carcinomas)[57] — a finding associated with poor prognosis[58]. Several investigators subsequently found that E-selectin on activated endothelial cells could bind carcinoma cells that expressed SLeX[58]. However, E-selectin is not constitutively expressed on endothelium. So others pursued the involvement of P- and L-selectin. It seems that the sulphated sialylated mucins of circulating carcinoma cells (and possibly other glycans such as heparan sulphate and/or sulphatides) mimic natural selectin ligands, initiating interactions with platelets and leukocytes[59,60]. This was reminiscent of classic studies showing that platelets and leukocytes facilitate the survival and spread of malignant cells after their entry into the bloodstream[56]. Studies in selectin-deficient mice confirmed a pivotal role of these glycan-binding proteins in various models of cancer metastasis[56,59,60]. Furthermore, when heparin was used as an inhibitor of P- and L-selectin, it markedly decreased metastasis in these model systems. This helps to explain the classic finding that heparin inhibits carcinoma metastasis[61] in a manner apparently independent of its anticoagulant effects[56].

Selectins are also involved in biomedically important processes other than cancer and inflammation, particularly reperfusion injury (which occurs when blood flow is interrupted and then restored in situations such as stroke, heart attack and traumatic shock)[62]. Therefore, selectins are a promising target for the prevention of such injuries. It is unfortunate that early attempts at clinical trials failed, for various scientific, practical and/or

Table 2 | Differences in sialic acid biology in humans compared with other hominids

Gene	Human-specific changes	Possible consequences for humans	Refs
CMAH	Human-specific *Alu* transposon-mediated exon deletion/ frame-shift/inactivation. Loss of cytosine 5′-phosphate (CMP)–Neu5Gc production	Loss of cell-surface Neu5Gc expression, excess of Neu5Ac on cell surfaces. Humans have anti-Neu5Gc antibodies but metabolically incorporate Neu5Gc *in vivo* (from dietary sources), and cultured human cells incorporate Neu5Gc from animal serum or feeder layers (*in vitro*)	21
SIGLEC1	Increased ligand density in humans, enhanced frequency and altered expression pattern in human macrophages	Altered responses to sialic-acid-expressing pathogens?	92
SIGLEC5/14	Restoration of 'essential arginine residue' for sialic acid recognition; expression suppressed on T cells	Hyper-responsive phenotype of human T cells — a possible role in disease propensity?	74, 93
SIGLEC6	Placental expression?	Unknown	94
SIGLEC7/9	Amino-acid changes in the sialic-acid-recognizing domain, allowing Neu5Ac recognition	Altered control of innate immune response by neutrophils? Enhanced susceptibility to sialic-acid-bearing pathogens?	95
SIGLEC11	Human-specific gene conversion, altered binding, new expression in microglia	Altered response of microglia to infections? Altered interactions of microglia with neural cells?	96
SIGLEC12	Human-specific mutation of 'essential arginine residue' decreasing sialic acid recognition	Unknown	97
SIGLEC13	Human-specific *Alu* transposon-mediated gene deletion	Unknown	98
ST6GAL1	Altered expression of Siaα2-6Galβ1-4GlcNAcβ1 termini in certain cell types	Protection from avian influenza virus. Susceptibility to human influenza virus	99

Of the fewer than 60 genes known to be involved in sialic acid biology[20,86], a surprisingly large number show human-specific changes compared with the closely related great apes (chimpanzees, gorillas and orangutans). Most of the differences involve changes in the Siglec family of intrinsic sialic-acid-binding lectins. The biological significance and relevance to human evolution of these changes are being investigated. Some possible consequences are given above.

commercial reasons. However, the basic principles remain unchanged, and further translational studies in this area are definitely warranted.

Siglecs

The negative charge of the sialic acids is broadly sufficient to explain their role in recognition involving factor H and selectins. Indeed, in both instances sulphate esters can replace sialic acids. Thus, until the 1990s there was no evidence for vertebrate sialic-acid-binding lectins dependent on multiple structural features of sialic acids for recognition. Two independent lines of work resulted in the discovery of such molecules. First, sheep erythrocyte rosetting by macrophages was traced to the presence of a cell-surface protein known as sialoadhesin[63]. While the sialic-acid-binding properties of this molecule were being characterized, independent evidence emerged that CD22 on B cells mediated adhesion specifically through the recognition of sialic acids in α2-6 linkage[64]. Subsequent cloning of sialoadhesin[65] revealed that CD22 and sialoadhesin belonged to a subfamily of immunoglobulin (Ig)-related proteins that shared homology at their N-terminal V-set and adjacent C2-set Ig domains (which were named because of their homology to the variable and constant domains of Igs), along with other known molecules termed CD33 and myelin-associated glycoprotein (MAG). These were then shown to have sialic-acid-dependent binding properties[66]. On the basis of these common structural and functional features the family was eventually termed the Siglecs (for sialic-acid-recognizing Ig-superfamily lectins). Homology searches of mouse and human genomes revealed the presence of an extended family of sialic-acid-recognizing molecules that could all be defined as Siglecs (see refs 14, 67, 68 for reviews).

Siglec binding requires recognition of not only the sialic acid's negative charge but also the side chain that comprises the C7, C8 and C9 positions, and its 5-*N*-acyl group. Furthermore, Siglecs often interact specifically with the sialic acid linkage to the underlying sugar chain (Table 1). In some cases, such as MAG, an extended specificity has been found to exist that depends on various aspects of the underlying glycan structure[69,70]. Several reviews on the expression and functions of Siglecs have recently been published[14,67,68].

In a nutshell, Siglecs fall into two evolutionarily distinct categories. The first comprises molecules conserved between mammalian species, including sialoadhesin (Siglec-1), CD22 (Siglec-2), MAG (Siglec-4), and the newly discovered Siglec-15 (T. Angata, personal communication). The second category is the CD33-related Siglecs (CD33rSiglecs), members of which are characterized by general homology to one another, the presence of tyrosine-based signalling motifs in the cytosolic tail, and rapid evolution by various mechanisms[14,67,68]. It has been suggested that

CD33rSiglecs are involved in recognizing sialic acids as self, and thus in inhibiting responses of the immune system. They may also be involved in the recognition of pathogens that express sialic acids[71,72]. A few CD33rSiglecs lack these cytosolic signalling motifs. Instead, they display a charged residue within the transmembrane domain, which they use to recruit the adaptors DAP-10 and DAP-12, proteins that contain potentially activating motifs[73,74]. Siglec-5 and -14 are paired CD33rSiglecs with negative and positive signalling potential, respectively, that are undergoing a fascinating form of concerted genomic evolution[74].

MAG is the most highly conserved Siglec; it is specifically expressed in glial cells and can recognize gangliosides (glycolipids that contain sialic acids and/or sulphate esters at specific positions). This recognition is crucial for the long-term stability of myelin, as is shown by defects in mutant mice that are null for either MAG or its glycan targets[69,70]. There is also evidence to suggest that MAG is partly responsible for the inhibition of neuronal sprouting and reconnection after injury. Thus, interfering appropriately with MAG could potentially improve nerve regeneration after events such as spinal cord injury[75]. Extensive work on CD22 has shown that it acts a negative modulator of signalling through the surface antigen receptor of B cells (see refs 76–78 for examples). This requires recognition of α2-6-linked sialic acids on CD22 itself, and probably on other molecules. The role of sialoadhesin in macrophages is still being investigated, but it seems to both modulate adaptive immune responses[79] and assist in recognition and/or phagocytosis of sialic-acid-expressing pathogens[71,72]. The functions of the evolutionarily conserved Siglec-15, recently discovered by Angata and colleagues, are unknown.

Thus, Siglecs are the largest family of sialic-acid-binding molecules, and much further work is required to define their functions. One unusual feature is that naturally occurring Siglec binding sites are typically 'masked' by the large pool of sialic acids present on the same cell surface[80]. Although such '*cis*' recognition may be of biological significance and may change during activation, '*trans*'-recognition can also occur between cells, as long as the density of sialic acid on the opposing cell is high enough[81]. In this regard, it is interesting that many successful vertebrate pathogens (particularly human pathogens) coat themselves with a high density of sialic acids[18] and can be recognized by CD33rSiglecs[71,72]. It has been suggested (but not proved) that this is a form of molecular mimicry in which the pathogens are mimicking self in order to downregulate the responses of innate immune cells[14].

L1CAM

Another Ig-superfamily member that has been reported to recognize sialic acids is L1 cell-adhesion molecule (L1CAM), which is widely

expressed in the nervous system, and is involved in many aspects of neural development and function[82]. Various human genetic disorders with prominent nervous system defects are caused by mutations in L1CAM. Sialic-acid-recognizing properties of L1CAM were demonstrated by one research group, which proposed that it has specificity for α2-3-linked sialic acids presented on CD24, a heavily glycosylated cell-surface molecule[82,83]. Unlike in Siglecs, this sialic-acid-binding property is not attributable to one of the Ig domains, but seems to derive from a fibronectin type-III repeat in another region of L1CAM[82,83]. Further studies of this recognition system are needed.

Uterine agglutinins

Soluble extracts of the endometrial lining of rat and human uterus contain a haemagglutinin (a protein capable of agglutinating erythrocytes), whose binding is mediated by recognition of sialic acids[84]. Erythrocytes are probably not the natural ligands for this agglutinin, which presumably recognizes other local sialic-acid-bearing ligands, mediating some as yet unknown biological functions. Interestingly, the agglutinin has also been shown to bind to sperm in a sialic-acid-dependent manner[85]. An unexpected finding is that the human uterine agglutinin preferentially recognizes Neu5Gc[85], the sialic acid absent in humans because of a genetic mutation[21]. It remains to be seen whether this apparent evolutionary loss of natural ligands can explain unusual aspects of human female reproductive biology[86]. This agglutinin needs to be cloned and further characterized.

Sperm sialic-acid-binding proteins

It has been reported that the binding of bovine sperm to egg zona pellucida glycoproteins is mediated by α2-3-linked sialic acids[87]. Studies in this area indicate the presence of a sperm plasma membrane sialic-acid-recognizing protein. Use of a specific enzyme inhibitor suggested that a neuraminidase released from cortical granules might also participate in the block to polyspermy by removing sialic acid from the zona

pellucida. Further studies are needed to define the role of sialic acid recognition in such interactions.

Laminin G domains

Abundant evidence indicates that interactions of the G domains of laminins with glycans displayed on α-dystroglycan are important in mediating interactions in the nervous system, at muscle–nerve junctions and in skeletal muscle[88]. One unusual glycan found on α-dystroglycan is a sialic-acid-containing sequence attached to the protein through an O-linked mannose residue[89]. Some evidence suggests that these sialic acids are involved in recognition[89]. However, this does not seem to be universal to all laminin G-domain interactions, and further studies are needed.

CD83

One study showed indirect evidence for sialic acid binding by the Ig-superfamily member CD83 (ref. 90), a marker for mature dendritic cells that is also essential for the generation of CD4+ T cells. The sequences of CD83 do not show homology to Siglecs, and this finding needs further investigation.

How evolution changed human sialic acid recognition

The hominid evolutionary lineage that gave rise to humans suffered a genetic defect in the production of the major Old World primate sialic acid, Neu5Gc[21]. As an isolated incident, this could simply have been due to a random drift of a mutation that occurred in a small population of ancestral hominids[21]. However, multiple differences in sialic acid biology have since been found between humans and our closest evolutionary cousins, the great apes[86] (Table 2). Although it is always difficult to reconstruct evolutionary events, a reasonable scenario is depicted in Fig. 3. Whatever the true sequence of events, it is clear that human evolution was associated with a 'sialoquake' in which a number of the genes associated with sialic acid biology underwent significant
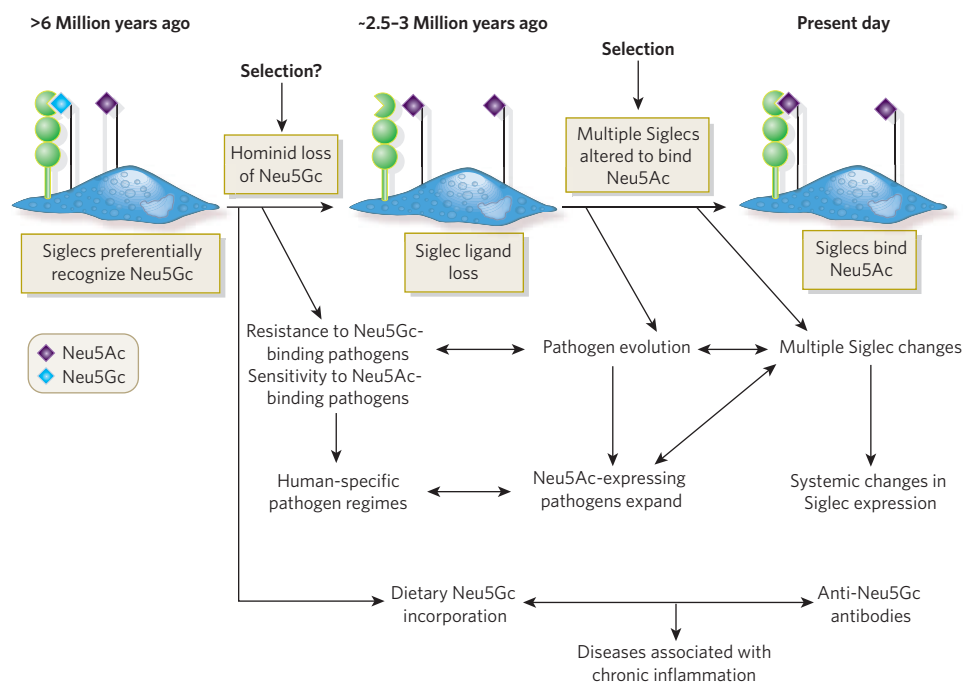
Figure 3 | Proposed scenario for occurrence of multiple changes in sialobiology during human evolution. Initial loss of Neu5Gc in human evolution could have occurred randomly or through selection as the result of a pathogen preferentially recognizing Neu5Gc. This would have led to sudden loss of Neu5Gc-binding sites for several CD33rSiglecs (green), probably generating unusual immune activation. All of the other changes in sialic acid biology could have resulted from evolutionary adjustments to these events (see Table 1 for details of human-specific Siglec changes). The Siglecs are depicted as binding sialic acids on the same cell surface.

Binding could also occur to high densities of sialic acids on other cell surfaces, including those of Neu5Ac-expressing pathogens. Also noted are additional possibilities for evolutionary changes in the range of pathogens that humans suffer from (pathogen regimes), as well as the unusual phenomenon of metabolic incorporation of dietary Neu5Gc into human endothelia and epithelia in the face of an anti-Neu5Gc antibody response. The latter process potentially facilitates diseases associated with chronic inflammation, such as atherosclerosis and cancer. Note that there have been no reports of pathogens that synthesize Neu5Gc.

changes in expression or function (see Table 2). This occurred in a biological system in which some rapid evolution was expected[14,67,68,86]. It remains to be seen what the consequences of all of these changes are for human populations today. Some possibilities are included in Table 2.

Future prospects

Glycan–protein interactions span a wide variety of biological processes, and only a very small subset (the sialic-acid-binding vertebrate proteins) have been considered here. The biology of such interactions remains incompletely understood, partly because of the small number of laboratories that are currently focused on their study. This is not surprising, because the present generation of established cellular and molecular biologists were not exposed to information about glycan biology and biochemistry during their training. This makes it a fertile area for the new generation of young scientists interested in gaining a more complete understanding of biology and disease at a molecular level.

Apart from learning more about known vertebrate sialic-acid-recognizing proteins, it is interesting to speculate as to whether there are any as yet undiscovered ones. In this regard, all known sialic-acid-binding features of vertebrate proteins (other than those of sialoadhesin and CD33rSiglecs) were discovered only after the protein had already been identified and characterized for other reasons. In other words, discovery of sialic acid binding was serendipitous, and related to observations such as unexpected loss of binding upon sialidase treatment. Thus, it is difficult to predict how many sialic-acid-binding proteins are yet to be discovered. After all, sialic acids have been consistently expressed at a high density on the cell surfaces of the deuterostome lineage for more than 500 million years. It is therefore reasonable to predict that they were co-opted for a variety of different recognition functions over evolutionary time. Rather than awaiting discovery of new sialic-acid-binding motifs and proteins by further chance observations, one could take a systematic approach to identifying sialic-acid-binding properties in tissue extracts, or in libraries of existing cDNAs and proteins. The use of glycan arrays[91] could also facilitate such an exploration. Meanwhile, *in vivo* manipulation of sialic-acid-binding protein functions is becoming increasingly feasible as a result of genetic modifications in mice[4,46] and by directed modification of sialic acids using various chemically synthesized precursors[22]. All of the above suggestions can also be broadly applied to glycan-binding proteins in general. It is probable that we have only scratched the surface of this class of biological recognition processes. A new generation of young scientists trained in understanding glycan structure and function must exploit these exciting opportunities. ■

Note added in proof: Ravetch and colleagues recently reported that the small fraction of IgG that binds to *Sambucus nigra* agglutinin (that is, that carries α2-6-linked sialic acids) mediates the inhibition of autoimmune responses, possibly through a sialic-acid-recognizing receptor on macrophages[100].

1. Rambourg, A. & Leblond, C. P. Electron microscope observations on the carbohydrate-rich cell coat present at the surface of cells in the rat. *J. Cell Biol.* **32,** 27–53 (1967).
2. Varki, A. *et al. Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, Plainview, New York, 1999).
3. Drickamer, K. & Taylor, M. *Introduction to Glycobiology* 2nd edn (Oxford Univ. Press, Oxford, 2006).
4. Ohtsubo, K. & Marth, J. D. Glycosylation in cellular mechanisms of health and disease. *Cell* **126,** 855–867 (2006).
5. Freeze, H. H. Genetic defects in the human glycome. *Nature Rev. Genet.* **7,** 537–551 (2006).
6. Hirschberg, C. B., Robbins, P. W. & Abeijon, C. Transporters of nucleotide sugars, ATP, and nucleotide sulfate in the endoplasmic reticulum and Golgi apparatus. *Annu. Rev. Biochem.* **67,** 49–69 (1998).
7. Varki, A. Factors controlling the glycosylation potential of the Golgi apparatus. *Trends Cell Biol.* **8,** 34–40 (1998).
8. Raetz, C. R. H. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71,** 635–700 (2002).
9. Varki, A. Nothing in glycobiology makes sense, except in the light of evolution. *Cell* **126,** 841–845 (2006).
10. Bishop, J. & Gagneux, P. Evolution of carbohydrate antigens — microbial forces shaping host glycomes? *Glycobiology* advance online publication, 19 January 2007 (doi:10.1093/glycob/cwm00).
11. Carver, J. P., Michnick, S. W., Imberty, A. & Cumming, D. A. Oligosaccharide–protein interactions: a three-dimensional view. *Ciba Found. Symp.* **145,** 6–18 (1989).
12. Hakomori, S. Carbohydrate-to-carbohydrate interaction, through glycosynapse, as a basis of cell recognition and membrane organization. *Glycoconj. J.* **21,** 125–137 (2004).
13. Sharon, N. & Lis, H. History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology* **14,** 53R–62R (2004).
14. Varki, A. & Angata, T. Siglecs — the major subfamily of I-type lectins. *Glycobiology* **16,** 1R–27R (2006).
15. Drickamer, K. & Taylor, M. E. Identification of lectins from genomic sequence data. *Methods Enzymol.* **362,** 560–567 (2003).
16. Schauer, R. Achievements and challenges of sialic acid research. *Glycoconjugate J.* **17,** 485–499 (2000).
17. Angata, T. & Varki, A. Chemical diversity in the sialic acids and related α-keto acids: an evolutionary perspective. *Chem. Rev.* **102,** 439–470 (2002).
18. Vimr, E. R., Kalivoda, K. A., Deszo, E. L. & Steenbergen, S. M. Diversity of microbial sialic acid metabolism. *Microbiol. Mol. Biol. Rev.* **68,** 132–153 (2004).
19. Harduin-Lepers, A., Mollicone, R., Delannoy, P. & Oriol, R. The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach. *Glycobiology* **15,** 805–817 (2005).
20. Altheide, T. K. *et al.* System-wide genomic and biochemical comparisons of sialic acid biology among primates and rodents: evidence for two modes of rapid evolution. *J. Biol. Chem.* **281,** 25689–25702 (2006).
21. Varki, A. Loss of N-glycolylneuraminic acid in humans: mechanisms, consequences and implications for hominid evolution. *Am. J. Phys. Anthropol.* **116** (suppl. 33), 54–69 (2001).
22. Prescher, J. A. & Bertozzi, C. R. Chemical technologies for probing glycans. *Cell* **126,** 851–854 (2006).
23. Wearne, K. A., Winter, H. C., O'Shea, K. & Goldstein, I. J. Use of lectins for probing differentiated human embryonic stem cells for carbohydrates. *Glycobiology* **16,** 981–990 (2006).
24. Schwarzkopf, M. *et al.* Sialylation is essential for early development in mice. *Proc. Natl Acad. Sci. USA* **99,** 5267–5270 (2002).
25. Mandal, C. Sialic acid binding lectins. *Experientia* **46,** 433–441 (1990).
26. Lehmann, F., Tiralongo, E. & Tiralongo, J. Sialic acid-specific lectins: occurrence, specificity and function. *Cell. Mol. Life Sci.* **63,** 1331–1354 (2006).
27. Pangburn, M. K. Host recognition and target differentiation by factor H, a regulator of the alternative pathway of complement. *Immunopharmacology* **49,** 149–157 (2000).
28. Fearon, D. T. Regulation by membrane sialic acid of β 1H-dependent decay-dissociation of amplification C3 convertase of the alternative complement pathway. *Proc. Natl Acad. Sci. USA* **75,** 1971–1975 (1978).
29. Pangburn, M. K. & Muller-Eberhard, H. J. Complement C3 convertase: cell surface restriction of β1H control and generation of restriction on neuraminidase-treated cells. *Proc. Natl Acad. Sci. USA* **75,** 2416–2420 (1978).
30. McEver, R. P. Selectins: lectins that initiate cell adhesion under flow. *Curr. Opin. Cell Biol.* **14,** 581–586 (2002).
31. Rosen, S. D. Ligands for L-selectin: homing, inflammation, and beyond. *Annu. Rev. Immunol.* **22,** 129–156 (2004).
32. Rosen, S. D., Singer, M. S., Yednock, T. A. & Stoolman, L. M. Involvement of sialic acid on endothelial cells in organ-specific lymphocyte recirculation. *Science* **228,** 1005–1007 (1985).
33. Tyrrell, D. *et al.* Structural requirements for the carbohydrate ligand of E-selectin. *Proc. Natl Acad. Sci. USA* **88,** 10372–10376 (1991).
34. Johnston, G. I., Cook, R. G. & McEver, R. P. Cloning of GMP-140, a granule membrane protein of platelets and endothelium: sequence similarity to proteins involved in cell adhesion and inflammation. *Cell* **56,** 1033–1044 (1989).
35. Handa, K., Nudelman, E. D., Stroud, M. R., Shiozawa, T. & Hakomori, S. Selectin GMP-140 (CD62; PADGEM) binds to sialosyl-Lea and sialosyl-Lex, and sulfated glycans modulate this binding. *Biochem. Biophys. Res. Commun.* **181,** 1223–1230 (1991).
36. Moore, K. L., Varki, A. & McEver, R. P. GMP-140 binds to a glycoprotein receptor on human neutrophils: evidence for a lectin-like interaction. *J. Cell Biol.* **112,** 491–499 (1991).
37. Bevilacqua, M. *et al.* Selectins: a family of adhesion receptors. *Cell* **67,** 233 (1991).
38. Varki, A. Selectin ligands. *Proc. Natl Acad. Sci. USA* **91,** 7390–7397 (1994).
39. Kawashima, H. *et al. N*-acetylglucosamine-6-O-sulfotransferases 1 and 2 cooperatively control lymphocyte homing through L-selectin ligand biosynthesis in high endothelial venules. *Nature Immunol.* **6,** 1096–1104 (2005).
40. Moore, K. L. *et al.* Identification of a specific glycoprotein ligand for P-selectin (CD62) on myeloid cells. *J. Cell Biol.* **118,** 445–456 (1992).
41. Sako, D. *et al.* Expression cloning of a functional glycoprotein ligand for P-selectin. *Cell* **75,** 1179–1186 (1993).
42. Wilkins, P. P., Moore, K. L., McEver, R. P. & Cummings, R. D. Tyrosine sulfation of P-selectin glycoprotein ligand-1 is required for high affinity binding to P-selectin. *J. Biol. Chem.* **270,** 22677–22680 (1995).
43. Sako, D. *et al.* A sulfated peptide segment at the amino terminus of PSGL-1 is critical for P-selectin binding. *Cell* **83,** 323–331 (1995).
44. Pouyani, T. & Seed, B. PSGL-1 recognition of P-selectin is controlled by a tyrosine sulfation consensus at the PSGL-1 amino terminus. *Cell* **83,** 333–343 (1995).
45. Leppanen, A., Yago, T., Otto, V. I., McEver, R. P. & Cummings, R. D. Model glycosulfopeptides from P-selectin glycoprotein ligand-1 require tyrosine sulfation and a core 2-branched O-glycan to bind to L-selectin. *J. Biol. Chem.* **278,** 26391–26400 (2003).
46. Lowe, J. B. & Marth, J. D. A genetic approach to mammalian glycan function. *Annu. Rev. Biochem.* **72,** 643–691 (2003).
47. Robinson, S. D. *et al.* Multiple, targeted deficiencies in selectins reveal a predominant role for P-selectin in leukocyte recruitment. *Proc. Natl Acad. Sci. USA* **96,** 11452–11457 (1999).
48. Marshall, B. T. *et al.* Direct observation of catch bonds involving cell-adhesion molecules. *Nature* **423,** 190–193 (2003).
49. Norgard-Sumnicht, K. E., Varki, N. M. & Varki, A. Calcium-dependent heparin-like ligands for L-selectin in nonlymphoid endothelial cells. *Science* **261,** 480–483 (1993).
50. Nelson, R. M. *et al.* Heparin oligosaccharides bind L- and P-selectin and inhibit acute inflammation. *Blood* **82,** 3253–3258 (1993).

51. Skinner, M. P. *et al.* Characterization of human platelet GMP-140 as a heparin-binding protein. *Biochem. Biophys. Res. Commun.* **164,** 1373–1379 (1989).

52. Koenig, A., Norgard-Sumnicht, K. E., Linhardt, R. & Varki, A. Differential interactions of heparin and heparan sulfate glycosaminoglycans with the selectins: implications for the use of unfractionated and low molecular weight heparins as therapeutic agents. *J. Clin. Invest.* **101,** 877–889 (1998).

53. Xie, X. *et al.* Inhibition of selectin-mediated cell adhesion and prevention of acute inflammation by nonanticoagulant sulfated saccharides: studies with carboxyl-reduced and sulfated heparin and with trestatin A sulfate. *J. Biol. Chem.* **275,** 34818–34825 (2000).

54. Stevenson, J. L., Choi, S. H. & Varki, A. Differential metastasis inhibition by clinically relevant levels of heparins — correlation with selectin inhibition, not antithrombotic activity. *Clin. Cancer Res.* **11,** 7003–7011 (2005).

55. Ludwig, R. J. *et al.* The ability of different forms of heparins to suppress P-selectin function *in vitro* correlates to their inhibitory capacity on bloodborne metastasis *in vivo. Thromb. Haemost.* **95,** 535–540 (2006).

56. Varki, N. M. & Varki, A. Heparin inhibition of selectin-mediated interactions during the hematogenous phase of carcinoma metastasis: rationale for clinical studies in humans. *Semin. Thromb. Hemost.* **28,** 53–66 (2002).

57. Fukushima, K. *et al.* Characterization of sialosylated Lewis^x as a new tumor-associated antigen. *Cancer Res.* **44,** 5279–5285 (1984).

58. Fukuda, M. Possible roles of tumor-associated carbohydrate antigens. *Cancer Res.* **56,** 2237–2244 (1996).

59. Borsig, L., Wong, R., Hynes, R. O., Varki, N. M. & Varki, A. Synergistic effects of L- and P-selectin in facilitating tumor metastasis can involve non-mucin ligands and implicate leukocytes as enhancers of metastasis. *Proc. Natl Acad. Sci. USA* **99,** 2193–2198 (2002).

60. Laubli, H., Stevenson, J. L., Varki, A., Varki, N. M. & Borsig, L. L-selectin facilitation of metastasis involves temporal induction of Fut7-dependent ligands at sites of tumor cell arrest. *Cancer Res.* **66,** 1536–1542 (2006).

61. Zacharski, L. R. & Ornstein, D. L. Heparin and cancer. *Thromb. Haemost.* **80,** 10–23 (1998).

62. Albelda, S. M., Smith, C. W. & Ward, P. A. Adhesion molecules and inflammatory injury. *FASEB J.* **8,** 504–512 (1994).

63. Crocker, P. R. *et al.* Purification and properties of sialoadhesin, a sialic acid-binding receptor of murine tissue macrophages. *EMBO J.* **10,** 1661–1669 (1991).

64. Powell, L. D., Sgroi, D., Sjoberg, E. R., Stamenkovic, I. & Varki, A. Natural ligands of the B cell adhesion molecule CD22β carry N-linked oligosaccharides with alpha-2,6-linked sialic acids that are required for recognition. *J. Biol. Chem.* **268,** 7019–7027 (1993).

65. Crocker, P. R. *et al.* Sialoadhesin, a macrophage sialic acid binding receptor for haemopoietic cells with 17 immunoglobulin-like domains. *EMBO J.* **13,** 4490–4503 (1994).

66. Kelm, S. *et al.* Sialoadhesin, myelin-associated glycoprotein and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily. *Curr. Biol.* **4,** 965–972 (1994).

67. Angata, T. Molecular diversity and evolution of the Siglec family of cell-surface lectins. *Mol. Divers.* **10,** 555–566 (2006).

68. Crocker, P., Paulson, J. & Varki, A. Siglecs and their roles in the immune system. *Nature Rev. Immunol.* **7,** 255–266 (2007).

69. Pan, B. *et al.* Myelin-associated glycoprotein and complementary axonal ligands, gangliosides, mediate axon stability in the CNS and PNS: neuropathology and behavioral deficits in single- and double-null mice. *Exp. Neurol.* **195,** 208–217 (2005).

70. Vyas, A. A., Blixt, O., Paulson, J. C. & Schnaar, R. L. Potent glycan inhibitors of myelin-associated glycoprotein enhance axon outgrowth *in vitro. J. Biol. Chem.* **280,** 16305–16310 (2005).

71. Jones, C., Virji, M. & Crocker, P. R. Recognition of sialylated meningococcal lipopolysaccharide by Siglecs expressed on myeloid cells leads to enhanced bacterial uptake. *Mol. Microbiol.* **49,** 1213–1225 (2003).

72. Carlin, A. F., Lewis, A. L., Varki, A. & Nizet, V. Group B streptococcal capsular sialic acids interact with Siglecs (immunoglobulin-like lectins) on human leukocytes. *J. Bacteriol.* **89,** 1231–1237 (2007).

73. Blasius, A. L., Cella, M., Maldonado, J., Takai, T. & Colonna, M. Siglec-H is an IPC-specific receptor that modulates type I IFN secretion through DAP12. *Blood* **107,** 2474–2476 (2006).

74. Angata, T., Hayakawa, T., Yamanaka, M., Varki, A. & Nakamura, M. Discovery of Siglec-14, a novel sialic acid receptor undergoing concerted evolution with Siglec-5 in primates. *FASEB J.* **20,** 1964–1973 (2006).

75. Yang, L. J. *et al.* Sialidase enhances spinal axon outgrowth *in vivo. Proc. Natl Acad. Sci. USA* **103,** 11057–11062 (2006).

76. Poe, J. C. *et al.* CD22 regulates B lymphocyte function *in vivo* through both ligand-dependent and ligand-independent mechanisms. *Nature Immunol.* **5,** 1078–1087 (2004).

77. Grewal, P. K. *et al.* ST6Gal-I restrains CD22-dependent antigen receptor endocytosis and Shp-1 recruitment in normal and pathogenic immune signaling. *Mol. Cell. Biol.* **26,** 4970–4981 (2006).

78. Collins, B. E., Smith, B. A., Bengtson, P. & Paulson, J. C. Ablation of CD22 in ligand-deficient mice restores B cell receptor signaling. *Nature Immunol.* **7,** 199–206 (2006).

79. Oetke, C., Vinson, M. C., Jones, C. & Crocker, P. R. Sialoadhesin-deficient mice exhibit subtle changes in B- and T-cell populations and reduced immunoglobulin m levels. *Mol. Cell. Biol.* **26,** 1549–1557 (2006).

80. Razi, N. & Varki, A. Masking and unmasking of the sialic acid-binding lectin activity of CD22 (Siglec-2) on B lymphocytes. *Proc. Natl Acad. Sci. USA* **95,** 7469–7474 (1998).

81. Collins, B. E. *et al.* Masking of CD22 by *cis* ligands does not prevent redistribution of CD22 to sites of cell contact. *Proc. Natl Acad. Sci. USA* **101,** 6104–6109 (2004).

82. Kleene, R., Yang, H., Kutsche, M. & Schachner, M. The neural recognition molecule L1 is a sialic acid-binding lectin for CD24, which induces promotion and inhibition of neurite outgrowth. *J. Biol. Chem.* **276,** 21656–21663 (2001).

83. Kleene, R. & Schachner, M. Glycans and neural cell interactions. *Nature Rev. Neurosci.* **5,** 195–208 (2004).

84. Banerjee, M. & Chowdhury, M. Purification and characterization of a sperm-binding glycoprotein from human endometrium. *Hum. Reprod.* **9,** 1497–1504 (1994).

85. Chatterji, U., Sen, A. K., Schauer, R. & Chowdhury, M. Paracrine effects of a uterine agglutinin are mediated via the sialic acids present in the rat uterine endometrium. *Mol. Cell. Biochem.* **215,** 47–55 (2000).

86. Varki, A. & Altheide, T. K. Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Res.* **15,** 1746–1758 (2005).

87. Velasquez, J. G. *et al.* Role of sialic acid in bovine sperm–zona pellucida binding. *Mol. Reprod. Dev.* **74,** 617–628 (2006).

88. Michele, D. E. *et al.* Post-translational disruption of dystroglycan–ligand interactions in congenital muscular dystrophies. *Nature* **418,** 417–422 (2002).

89. Chiba, A. *et al.* Structures of sialylated O-linked oligosaccharides of bovine peripheral nerve α-dystroglycan The role of a novel O-mannosyl-type oligosaccharide in the binding of α-dystroglycan with laminin. *J. Biol. Chem.* **272,** 2156–2162 (1997).

90. Scholler, N., Hayden-Ledbetter, M., Hellström, K. E., Hellström, I. & Ledbetter, J. A. CD83 is a sialic acid-binding Ig-like lectin (Siglec) adhesion receptor that binds monocytes and a subset of activated CD8$^+$ T cells. *J. Immunol.* **166,** 3865–3872 (2001).

91. Blixt, O. *et al.* Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl Acad. Sci. USA* **101,** 17033–17038 (2004).

92. Brinkman-Van der Linden, E. C. M. *et al.* Loss of N-glycolylneuraminic acid in human evolution: implications for sialic acid recognition by siglecs. *J. Biol. Chem.* **275,** 8633–8640 (2000).

93. Nguyen, D. H., Hurtado-Ziola, N., Gagneux, P. & Varki, A. Loss of Siglec expression on T lymphocytes during human evolution. *Proc. Natl Acad. Sci. USA* **103,** 7765–7770 (2006).

94. Patel, N. *et al.* OB-BP1/Siglec-6. A leptin- and sialic acid-binding protein of the immunoglobulin superfamily. *J. Biol. Chem.* **274,** 22729–22738 (1999).

95. Sonnenburg, J. L., Altheide, T. K. & Varki, A. A uniquely human consequence of domain-specific functional adaptation in a sialic acid-binding receptor. *Glycobiology* **14,** 339–346 (2004).

96. Hayakawa, T. *et al.* A human-specific gene in microglia. *Science* **309,** 1693 (2005).

97. Angata, T., Varki, N. M. & Varki, A. A second uniquely human mutation affecting sialic acid biology. *J. Biol. Chem.* **276,** 40282–40287 (2001).

98. Angata, T., Margulies, E. H., Green, E. D. & Varki, A. Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. *Proc. Natl Acad. Sci. USA* **101,** 13251–13256 (2004).

99. Gagneux, P. *et al.* Human-specific regulation of α2–6 linked sialic acids. *J. Biol. Chem.* **278,** 48245–48250 (2003).

100. Nimmerjahn, F. & Ravetch, J. V. The antiinflammatory activity of IgG: the intravenous IgG paradox. *J. Exp. Med.* **204,** 11–15 (2007).

Heparan sulphate proteoglycans fine-tune mammalian physiology

Joseph R. Bishop[1], Manuela Schuksz[1] & Jeffrey D. Esko[1]

Heparan sulphate proteoglycans reside on the plasma membrane of all animal cells studied so far and are a major component of extracellular matrices. Studies of model organisms and human diseases have demonstrated their importance in development and normal physiology. A recurrent theme is the electrostatic interaction of the heparan sulphate chains with protein ligands, which affects metabolism, transport, information transfer, support and regulation in all organ systems. The importance of these interactions is exemplified by phenotypic studies of mice and humans bearing mutations in the core proteins or the biosynthetic enzymes responsible for assembling the heparan sulphate chains.

Physiology describes the integration of physical and chemical activities in cells and organs to maintain systemic homeostasis. Textbooks often arrange the subject according to major organ systems: digestive, endocrine, nervous, muscular, skeletal, urinary, reproductive, respiratory, integumentary, circulatory and immune, with the overarching themes of metabolism, transport, information transfer, support and regulation. Heparan sulphate proteoglycans (HSPGs) have essential roles in all of these systems (Table 1). Many excellent reviews have appeared on the involvement of proteoglycans in embryological development[1–3], and therefore this subject will not be considered further here. Instead, this review focuses on recent studies demonstrating their role in adult physiology, with particular emphasis on genetic mouse models and pathophysiological states in humans.

Structure and properties of heparan sulphate proteoglycans

HSPGs are composed of a core protein and one or more heparan sulphate (HS) glycosaminoglycan (GAG) chains — linear polysaccharides composed of alternating N-acetylated or N-sulphated glucosamine units (*N*-acetylglucosamine, GlcNAc, or *N*-sulphoglucosamine, GlcNS) and uronic acids (glucuronic acid, GlcA, or iduronic acid, IdoA). There are three subfamilies of HSPGs: the membrane-spanning proteoglycans (namely syndecans, betaglycan and CD44v3), the glycophosphatidyl-inositol (GPI)-linked proteoglycans (namely glypicans), and the secreted extracellular matrix (ECM) proteoglycans (namely agrin, collagen XVIII and perlecan). Haematopoietic cells also contain a secretory vesicle proteoglycan known as serglycin. Some HSPGs contain chondroitin sulphate/dermatan sulphate, a GAG chain that differs from HS in its component sugars (*N*-acetylgalactosamine (GalNAc) and GlcA/IdoA) and patterns of modification, and other types of glycan (N-linked and O-linked mucin-type chains).

The HS chains are assembled on core proteins by enzymes in the Golgi, using nucleotide sugars imported from the cytoplasm. During their assembly, they undergo a series of processing reactions involving GlcNAc N-deacetylation and N-sulphation, epimerization of GlcA to IdoA, and O-sulphation that generate relatively short segments of modified sugars interspersed by variable tracts of unmodified sugars[4] (Fig. 1). There is great structural heterogeneity in terms of chain length and size, the spacing of the modified tracts, and the extent of sulphation and epimerization within the modified segments. Processing also can occur by plasma-membrane-bound endosulphatases (SULF1 and

SULF2), which remove specific sulphate groups from the chains[5,6]. The HSPGs on the cell surface can be shed by proteolytic cleavage of the core protein (Fig. 2e) and by endoglycosidic cleavage of the HS chains by extracellular heparanase[7] (HPA; Fig. 2f). In this way, bound ligands can be liberated and allowed to diffuse away from a cell. Some data suggest that HS can localize to the nucleus (Fig. 2n) but the significance of these observations remains unclear[8].

Because of their high negative charge, the HS chains bind to a plethora of proteins, including the members of the fibroblast growth factor (FGF) family and their receptor tyrosine kinases, transforming growth factors (TGFs), bone morphogenetic proteins (BMPs), Wnt proteins, chemokines and interleukins, as well as enzymes and enzyme inhibitors, lipases and apolipoproteins, and ECM and plasma proteins. The most well-studied example is activation of the serpin protease inhibitor antithrombin by a specific pentasaccharide within heparin (a processed, highly sulphated form of HS), which interacts with positively charged amino-acid residues aligned along two α-helices in the protein. Binding results in a conformational change, which increases the rate of inactivation of proteases involved in coagulation (for example, factors IIa and Xa)[9].

The interaction of heparin and HS with FGFs and their receptors has also been characterized in great detail[10]. In this case, the GAG chain acts as a template that bridges FGF and the FGF receptor (FGFR; Fig. 2a) like the interaction of heparin with antithrombin and factor IIa. Formation of the complex effectively lowers the concentration of FGF required to initiate signalling through its receptor and extends the duration of the response[11]. Unlike heparin–antithrombin binding, the formation and function of FGF–FGFR complexes may depend more on the spatial distribution of negatively charged groups in HS rather than a specific sequence of sulphated sugars[12]. Other growth factor/receptor interactions may follow a similar binding and activation process.

Many functions of HSPGs depend on interactions of the core protein. For example, HSPGs at the plasma membrane can transfer spatial information about a cell's environment and either activate adhesion mechanisms or enhance cell motility. The cytoplasmic domains of syndecans interact with cytoskeletal proteins, as well as kinases (Src and calcium/calmodulin-dependent serine protein kinase, CASK) and phosphatidyl-inositol-4,5-bisphosphate ($PtdIns(4,5)P_2$)[13] (Fig. 2j). Often, signalling through HSPGs in this context occurs in collaboration with other cell-surface receptors (such as integrins) to facilitate cell attachment,

[1]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093, USA.

Table 1 | Genetic studies of proteoglycans in mammalian physiology

Organ system	Gene/mutation	Phenotype
Digestive system	$Sdc1^{-/-}$	Protein-losing enteropathy (PLE)[32]
	$Ndst1^{-/-}$	Hypertriglyceridaemia due to decreased binding and endocytosis of chylomicron and very-low-density lipoprotein remnants[20]
Endocrine system	$Sdc1^{tg}$, $Sdc3^{-/-}$, Hpa^{tg}	Loss of control over food intake, resistance to diet induced obesity, binding of melanocortin antagonists[16–18]
	$Gpc3^{-/-}$	Simpson–Golabi–Behmel syndrome, pre- and postnatal overgrowth, multiple growth-factor signalling systems[80,81]
Nervous system	$Sdc3^{-/-}$	Enhanced long-term potentiation, impaired hippocampal function, altered binding of heparin-binding growth-associated molecule (HB-GAM)[82]
	$Ndst1^{-/-}$	Cerebral hypoplasia, neural-tube closure defects[83], eye and lens defects[84,85]
	$Hs2st^{-/-}$, $Hsglce^{-/-}$	Eye defects[86,87]
Muscular system	$Agrn^{-/-}$	Defective neuromuscular junction and synapse formation, and acetylcholine receptor clustering[31,88]
	$Hspg2^{-/-}$	Acetylcholinesterase localization at synaptic junctions[89]
	$Sdc3^{-/-}$, $Sdc4^{-/-}$	Satellite cell defects[62]
Skeletal system	$Hspg2^{-/-}$	Cartilage and growth-plate defects[27,38]
	$Gpc3^{-/-}$	Reduced calcified trabecular bone, delayed osteoclast appearance[90]; BMP4 interactions, postaxial polydactyly and rib malformations[91]
	$Ext1^{Gt/Gt}$ (hypomorphic allele)	Growth-plate abnormalities, IHH signalling and spatial control over its distribution[36]
	$Ext2^{+/-}$	Rib exostoses, chondrocyte differentiation[39]
	$Ndst1^{-/-}$	Craniofacial defects[83,84]; lack of or delay in ossification[92]
	$Hsglce^{-/-}$	Shorter body length; excessive mineralization; lack of proximal phalanges and tarsal bones, and postaxial polydactyly, malformed ribcage and sternum[87]
	Hpa^{tg}	Enhanced osteogenic differentiation of bone marrow stromal cells, increases in trabecular bone mass, cortical thickness and rate of bone formation[93]
Integumentary system	$Hspg2^{-/-}$ (antisense)	Thin, poorly organized epidermis, incomplete stratification, impaired wound healing[55,56]
	$Sdc1^{-/-}$	Keratinocyte activation in wound healing[58]; reduction of secondary and tertiary branching in the mammary glands[41]
	$Sdc4^{-/-}$	Delayed cutaneous wound healing[57]
	Hpa^{tg}	Overgrowth of hair[63]; accelerated wound angiogenesis[94]; mammary hyperbranching[18]
Circulatory system	$Col18a1^{-/-}$	Increased microvascular growth[95]; increased angiogenesis associated with atherosclerotic plaques[96]
	$Sdc1^{-/-}$	Increased inflammation-mediated corneal angiogenesis[97]
	$Sdc4^{-/-}$	Impairment of fetal vessels in placenta[98]; delayed angiogenesis in wound granulation tissue[57]
Immune system	$Sdc1^{-/-}$	Enhanced leukocyte–endothelial interaction in the retina[97]; enhanced *Pseudomonas aeruginosa* infection rate and virulence[74]
	$Prg1^{-/-}$	Secretory granule defects in mast cells[50]; defective secretory granule maturation and granzyme B storage in cytotoxic T cells[53]; no effect on macrophages[54]
	$Ndst1^{-/-}$ (endothelia and leukocytes)	Decreased chemokine transcytosis and presentation, and neutrophil infiltration[49]
	$Ndst2^{-/-}$	Mast-cell deficiency, defective protease storage in granules[51,52]
Respiratory system	$Sdc1^{-/-}$	Airway hyperresponsiveness, hypersecretion, eosinophilia, local IL-4 response[73]; lung inflammation[76]
	$Ndst1^{-/-}$	Lung hypoplasia and surfactant insufficiency[84,99]
	$Hsglce^{-/-}$	Poorly inflated lungs, thickened, cell-rich alveolar walls[87]
Urinary system	$Col18a1^{-/-}$	Broadened basement membrane, reduced filtration capacity[30]
	$Hspg2^{-/-}$ (exon 3 deletion)	Exon 3 deletion shows proteinuria after protein loading[28,29]
	Hpa^{tg}	Increased levels of urinary protein and creatinine[18]; decreased amyloid deposits in the kidneys[72]
	$Sdc4^{-/-}$	Increased susceptibility to κ-carrageenan-induced renal damage[100]
	$Hs2st^{-/-}$	Renal agenesis[86]
	$Hsglce^{-/-}$	Renal agenesis[87]
Reproductive system	$Hs3st1^{-/-}$	Ovarian function, placenta development[66,67]

Null genotypes refer to systemic defects unless otherwise indicated. *Agrn*, agrin; *Col18a1*, collagen XVIII; *Ext*, exostosin (co-polymerase); *Gpc*, glypican; *Gt*, gene trap allele; *Hs2st*, uronyl 2-*O*-sulphotransferase; *Hs3st*, glucosamine 3-*O*-sulphotransferase; *Hs6st*, glucosamine 6-*O*-sulphotransferase; *Hsglce*; HS glucuronyl C5 epimerase; *Hpa*, heparanase; *Hspg2*, perlecan; *Ndst*, GlcNAc *N*-deacetylase/*N*-sulphotransferase; *Prg1*, serglycin; *Sdc*, syndecan; tg, transgenic expression.

spreading and motility. For example, the ECM protein fibronectin contains domains that simultaneously bind to the HS chains of syndecans and one or more integrins to induce cell spreading and focal adhesion formation[14]. In fact, almost every ECM molecule contains binding sites for HS (Fig. 2i), suggesting that the balance between adhesion and motility depends on the integration of signals mediated through proteoglycan-binding and integrin-based adhesion mechanisms. Recent studies show that the extracellular domains of proteoglycans can bind directly to protein ligands independently of HS chains[15]. How cells coordinate these various activities remains an open question.

Heparan sulphate proteoglycans in mammalian physiology

In mammalian physiology, protein–HSPG interactions promote activities specific to each organ system. To study these systems *in vivo*, numerous knockout and transgenic mice have been generated, often with surprising phenotypes (summarized in Table 1). Although the repertoire of strains does not include mutants in all of the relevant genes, the available strains show that HSPGs orchestrate metabolism, transport, information transfer, support and regulation at the systemic level, as well as the cellular level. The following examples illustrate these principles.

Heparan sulphate proteoglycans modulate nutritional metabolism

HSPGs control food intake by modulating feeding behaviour through a feedback loop in the hypothalamus. This was first discovered by chance in transgenic mice ectopically expressing syndecan-1 in the hypothalamic nuclei[16]. The transgenic expression of syndecan-1 mimicked an increase in syndecan-3 seen in nuclei during food deprivation. Consistent with this observation, mice deficient in syndecan-3 do not eat as much and show reduced adipose content and partial resistance to obesity, whereas transgenic syndecan-1 mice develop maturity-onset obesity as a result of excessive food intake[17]. The HS chains, rather than the core protein, seem to modulate this process, because transgenic mice that overexpress human heparanase also exhibit reduced food consumption and have decreased body weight[18]. The phenotype resembles that of mice with impaired action of the melanocortin system. Possible mechanisms include sequestration of antagonists, such as agouti-related protein, or direct interaction of syndecan-3 with the melanotropin-stimulating hormone (MSH) receptor.

In the circulatory system, HSPGs modulate lipid metabolism by serving as receptors for lipases, tethering them to endothelial cell surfaces, and by acting as clearance receptors in the liver[19]. Dietary triglycerides and triglycerides synthesized in the liver enter the circulation in chylomicrons and very-low-density lipoproteins (VLDL), respectively. As chylomicrons and VLDL circulate, they encounter lipases (endothelial, hepatic and lipoprotein lipases) that hydrolyse triglycerides from the core, generating fatty acids for energy metabolism in peripheral tissues and lipid storage in adipose tissue. These lipases are not usually found in the plasma but, instead, are tethered to cells and modulated by HSPGs. After hydrolysis of the triglyceride core, the remnant particles enter the liver sinusoids. Here, HSPGs facilitate their sequestration and receptor-mediated clearance by binding to apolipoproteins and bound lipases (Fig. 2g), subsequently allowing degradation of the particles in lysosomes (Fig. 2h). Recent studies of mice bearing a hepatocyte-specific knockout of N-deacetylase/N-sulphotransferase (Ndst1) showed that plasma triglycerides accumulate under both fasted and fed conditions. Crossbreeding studies with low-density lipoprotein (LDL) receptor mutants showed that HSPGs work in parallel with LDL receptors[20]. Analyses of lipoprotein uptake by cultured cells indicate that syndecans and possibly perlecan can mediate uptake, but the relevant proteoglycan receptors in vivo are not known[21,22]. There is evidence that binding of the ligand can stimulate uptake[21], suggesting that HSPGs

should be considered classical endocytic receptors like transferrin or LDL receptors, albeit with broader specificity.

Heparan sulphate proteoglycans organize basement membrane barriers

Epithelial cells sit on a basement membrane, which consists of the basal lamina (containing collagens IV, VI and XVIII, laminins, agrin and perlecan) connected to the underlying reticular matrix composed of fibronectin, fibrillar collagens and proteoglycans[23]. These matrices provide support and resistance to mechanical stress, and these properties are partly dependent on interactions of the HS chains with other basement membrane components (Fig. 2k). HSPGs also have direct roles in regulating the transport of solutes across these barriers. In the urinary system, a specialized thick (100–200 nm) basal lamina known as the glomerular basement membrane (GBM) separates the endothelial cells lining the capillaries from interdigitated podocytes and acts as a filtration barrier. The acidic GAG chains prevent negatively charged substances, such as proteins in the blood, from passing through the GBM and entering the filtrate. Consistent with this, loss of HS from the GBM results in proteinuria in cases of diabetic nephropathy[24]. Furthermore, perfusion of rat kidneys with bacterial heparinase[25] or transgenic expression of heparanase in the kidney[18] increases levels of protein and creatinine in urine.

In the adult kidney, agrin, perlecan and collagen XVIII are present in the GBM[26]. Because perlecan-null mutants are embryonic lethal[27], mutant mice were created with a deletion of exon 3, which encodes the HS attachment sites in the amino-terminal portion of the protein[28]. These mice are viable and do not exhibit changes in GBM ultrastructure or proteinuria under normal conditions. However, they do show filtration defects when stressed by protein loading, indicating the importance of the HS chains in the deleted domain[29]. Mutants lacking collagen XVIII exhibit thickening of the GBM and a mild increase in serum creatinine levels, which is suggestive of decreased filtration capacity[30]. The function of agrin in the kidney is unknown, because systemic knockouts are embryonic lethal[31]. Double and triple mutants may be needed to determine the relative contribution of each HSPG.

Protein-losing enteropathies (PLE) refer to disorders in which loss of plasma proteins occurs in the gastrointestinal tract. This can arise from coeliac disease, Crohn's disease and congenital disorders of glycosylation, as well as after infection or Fontan surgery (to correct congenital univentricular hearts). Recent studies show that patients with PLE have
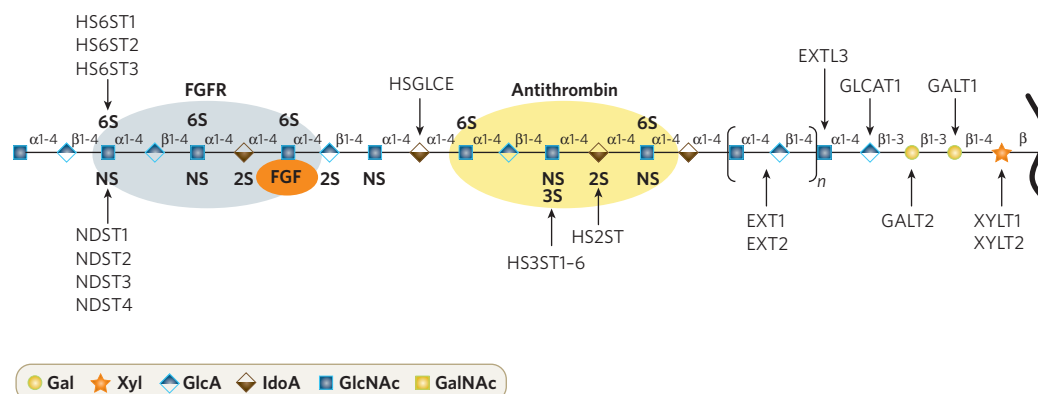
Figure 1 | Assembly of heparan sulphate and formation of binding sites for ligands. In mammals, as many as 26 enzymes participate in the formation of HS chains. HSPG core proteins are xylosylated co-translationally at specific serine residues (by XYLT1 and XYLT2). A tetrasaccharide primer (GlcA-Gal-Gal-Xyl, where Gal is galactose and Xyl is xylose) is assembled in the Golgi (by GALT1 and GALT2, which are galactosyltransferases, and by glucuronyltransferase 1 (GLCAT1)), and the first GlcNAc unit is added (by EXTL3). EXT1/EXT2 co-polymerize GlcAβ4 and GlcNAcα4 units into linear chains of 40–100 residues, which then undergo extensive modification by GlcNAc N-deacetylation and N-sulphation (denoted NS, by NDST1–4), C5 epimerization of D-GlcA to L-IdoA (by HS glucuronyl

C5 epimerase, HSGLCE) and O-sulphation at C2 of uronic acids (denoted 2S, by HS2ST), and at C6 (denoted 6S, by HS6ST1–3) and more rarely at C3 (denoted 3S, by HS3ST1–6) of glucosamine residues. The length of the sulphated and non-sulphated segments varies. Outside the cell, two endosulphatases (SULF1 and SULF2) catalyse the removal of specific 6-O-sulphate groups, and secreted heparanase (HPA) can fragment the chains (not shown). The pattern of negatively charged sulphates and uronic acids creates binding sites for various protein ligands, including growth factors (such as FGF), receptors (such as FGFR) and protease inhibitors (such as antithrombin), as well as other proteins (not shown).

diminished HSPGs, with prominent loss of syndecan-1 from the baso-lateral surface of intestinal epithelial cells[32]. The mechanism underlying the disappearance of syndecan (and possibly other HSPGs) is unknown, but the effect seems to correlate with inflammatory insult coupled with an underlying condition[33]. The proteoglycans may function in two ways: as an integral component of the basement membrane and as a buffer for inflammatory cytokines, such as tumor necrosis factor-α (TNF-α) and interferon-γ. Interestingly, treatment with heparin can alleviate this problem in some patients, possibly by binding to the cytokines and preventing inflammation[34].

Heparan sulphate proteoglycans in cell signalling and morphogenesis
Numerous studies in mice and other model organisms have documented the role of HSPGs in regulating morphogen gradients and growth-factor signalling reactions during development (Fig. 2l). Some of these processes continue into adulthood — for example, the formation of endochondral bones starts during embryogenesis and proceeds through puberty as body size increases. Axial bone growth occurs through growth plates in which chondrocytes undergo a coordinated programme of proliferation, hypertrophy and ossification under the control of Indian hedgehog (IHH), FGF18, BMPs, Wnts and probably other factors[35], many of which bind to HS. Homozygous mice carrying a hypomorphic allele of the HS-polymerizing enzyme EXT1 exhibit skeletal abnormalities[36]. Examination of the growth plates in these animals showed that the gradient of IHH extends further from its source, causing a broadening of the proliferative zone. Effects on spatial distribution of growth factors could be due to protection against proteolysis or uptake and clearance through endocytosis[37]. The specific proteoglycans affecting IHH diffusion are unknown. Perlecan may have a role, as null mutants exhibit severe disorganization of the columnar chondrocytes in growth plates and defective endochondral ossification[27,38]. Interestingly, deletion of exon 3 from perlecan has no reported effect on skeletal development, suggesting that the GAGs attached at this site are dispensable.

In humans, heterozygous null mutations in *EXT1* and *EXT2* cause hereditary multiple exostoses, an autosomal dominant disease characterized by the formation of cartilage-capped bony outgrowths (osteo-chondromas or exostoses) on growth plates throughout the body. Heterozygous mice also develop exostoses, but for unknown reasons the growths are limited to the ribs[39]. The mechanism underlying the
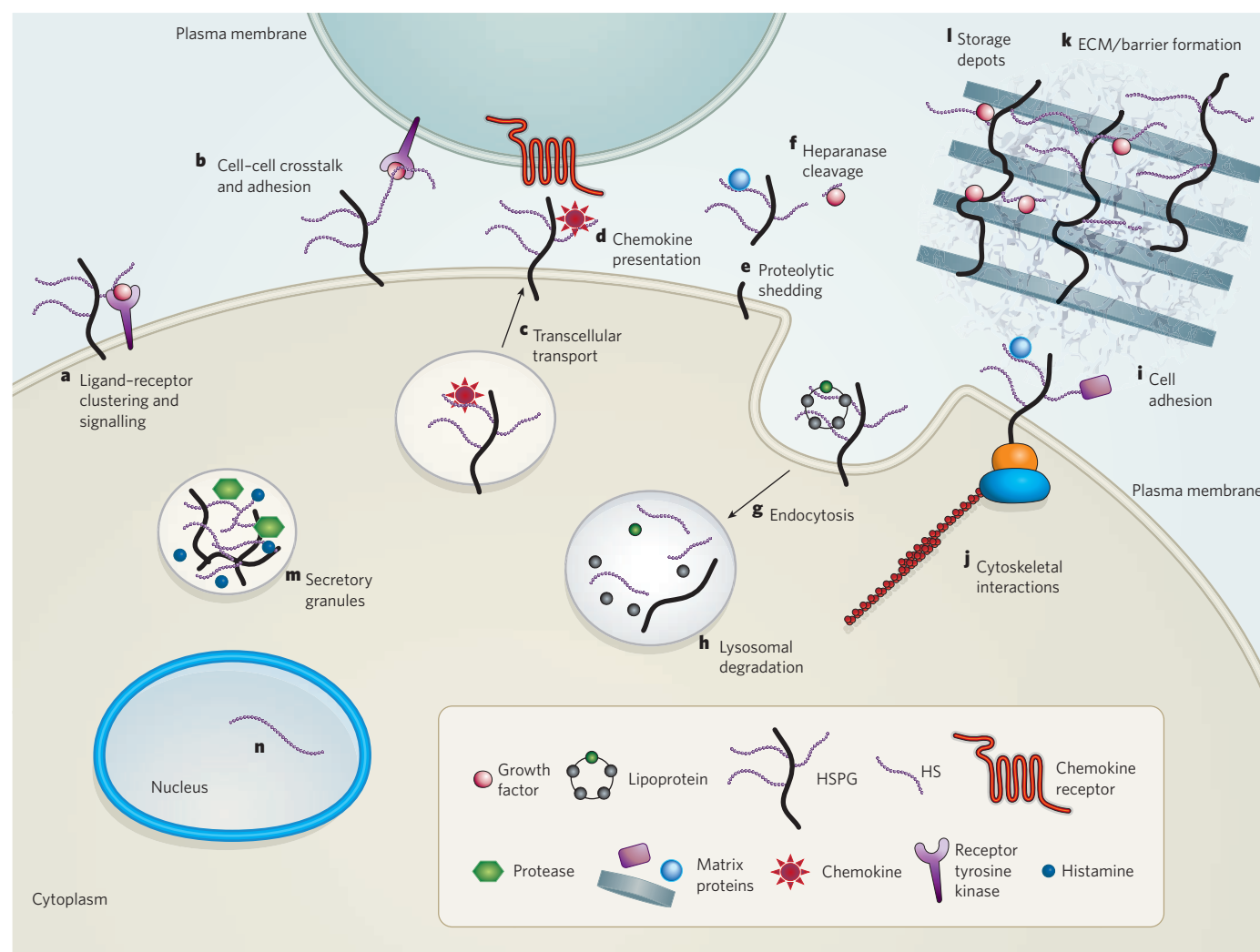
Figure 2 | Heparan sulphate proteoglycans have many roles in cell physiology. HSPGs function as co-receptors for growth factors and their receptor tyrosine kinases, which are present either on the same cell (**a**) or on adjacent cells (**b**). They transport chemokines across cells (**c**) and present them at the cell surface (**d**). Proteolytic processing leads to the shedding of syndecans and glypicans from the cell surface (**e**), and heparanase cleaves the HS chains (**f**), liberating bound ligands (such as growth factors). Cell-surface HSPGs are actively taken up by endocytosis (**g**) and can recycle back to the surface or be degraded in lysosomes (**h**). HSPGs also facilitate cell adhesion to the extracellular matrix (**i**) and form bridges to the cytoskeleton (**j**). Secreted HSPGs are involved in the formation of organized extracellular matrices that form physiological barriers (**k**) and sequester growth factors and morphogens for later release (**l**). Serglycin carrying highly sulphated heparin chains is packaged into secretory granules of haematopoetic cells (**m**). Finally, some experiments suggest that HS chains exist in the nucleus (**n**), although their function in this location is unknown.

formation of exostoses is unknown, but presumably relates to loss of binding of growth factors to truncated HS chains in growth plate chondrocytes.

The mammary glands develop postnatally and undergo regression and amplification under the control of endocrine hormones, oestrogen, progesterone, growth hormone and prolactin. Like most other endocrine factors, these hormones do not bind to HS but rather act on the mammary stroma to induce the local expression of soluble growth factors, many of which do bind to HS[40]. Genetic evidence supporting a role for HS in mammary development has been derived from studies of mice deficient in syndecan-1, which exhibit normal primary mammary duct formation but have a mild reduction in secondary and tertiary branching[41]. By contrast, hyperbranching occurs in the mammary epithelia of transgenic mice expressing human heparanase[18]. These studies do not distinguish between the effects of the mutations or the transgenes on mammary epithelial cells directly or by stromal interactions. Recent studies in mice show that selective inactivation of *Ndst1* in mammary epithelia using the Cre–*loxP* system causes a specific block in lobuloalveolar development during pregnancy and lactation (B. Crawford and J.E., unpublished observations) and that branching morphogenesis is affected in glands lacking both NDST1 and NDST2 or EXT1 (O. Garner and J.E., unpublished observations). These findings suggest that different HS growth factor interactions operate at different stages of gland development and that formation of the mammary branches and lobuloalveoli depends on HS generated specifically by the mammary epithelia.

Cellular crosstalk by heparan sulphate proteoglycans

HSPGs can also act across cells or '*in trans*' (Fig. 2b). At neuromuscular junctions, a highly specialized basal lamina assembles in the synaptic cleft. The motor neurons secrete agrin into the cleft, where it interacts with a muscle-specific kinase (MuSK, a transmembrane tyrosine receptor complex), and causes recruitment of intracellular casein kinase 2 and rapsyn, and subsequent clustering of acetylcholine receptors on the sarcolemma. In mice that lack agrin, neuromuscular junctions do not form properly, and acetylcholine receptors do not cluster[31,42]. Interestingly, undersulphation of the chains in mice with a systemic knockout of *Ndst1* has no effect on clustering[43], but treatment of cultured muscle cells with chlorate, a general sulphation inhibitor, or chondroitinase blocks it[44,45]. These findings suggest that one or more proteoglycans involved in receptor clustering may contain chondroitin sulphate chains rather than HS.

Activation of proteoglycans *in trans* was first shown during early development in zebrafish, when syndecan-2 expressed by ectodermal cells acts on signal-transducing receptors on the mesoderm, thus determining left–right axis formation[46]. It also occurs during blood vessel growth (angiogenesis) in mice. Arteries and arterioles are surrounded by mural cells, either vascular smooth-muscle cells in larger vessels or pericytes in small vessels. Mural cell proteoglycans can transactivate signalling pathways on the underlying endothelial cells[47]. This was elegantly demonstrated by preparing chimaeric cultures of mouse embryonic stem cells with defective *Ndst* expression (HS chains lacking sulphate) or VEGFR2, the primary receptor tyrosine kinase that mediates VEGF-dependent cell growth. Activation of the differentiation programme that drives angiogenesis resulted in chimaeric vessels that contained smooth-muscle cells expressing HSPGs, and HSPG-negative endothelial cells expressing VEGFR2. Transactivation of VEGFR2 by HSPGs on the smooth-muscle cells led to enhanced signal transduction by facilitating the formation of receptor–ligand complexes and trapping of the active VEGFR2 signalling complex on endothelial cells. How HSPGs coordinate growth-factor signalling reactions *in trans* with *cis* cell-autonomous reactions remains a mystery.

Heparan sulphate proteoglycans in injury and repair

Tissue injury results in inflammatory responses, recruitment of leukocytes into the damaged areas, and repair processes that depend on cell division and angiogenesis. In addition, stem cells may be activated to repopulate tissues with appropriate cell types.

Immediately after injury, macrophages and mast cells at the site of tissue damage release TNF-α and interleukin-1, which activate the endothelial cells in the nearby vasculature. Activation stimulates exocytosis of Weibel–Palade bodies, resulting in the appearance of platelet (P)-selectin at the cell surface, which initiates rolling of marginated leukocytes in the microvasculature. Chemokines presented on the endothelial surface (Fig. 2d) activate integrin receptors on leukocytes, resulting in firm adhesion and subsequent leukocyte extravasation across the endothelium at the site of injury. HSPGs have several roles in this process[48]. Recent studies of mice containing an endothelium-specific knockout of *Ndst1* showed that decreasing sulphation by ~60% diminished neutrophil infiltration[49]. However, this effect was not related to P-selectin but was due to weaker binding of leukocyte (L)-selectin on neutrophils to undersulphated proteoglycans on endothelial cells, which enhanced neutrophil rolling velocity. The mutation also reduced chemokine transcytosis across endothelial cells (Fig. 2c) and presentation on the cell surface, decreasing neutrophil adhesion and migration. The HSPGs involved in this process are unknown. Interestingly, altering *Ndst1* expression in leukocytes had no effect on neutrophil recruitment or T-cell-mediated responses, indicating that endothelial HS dominates in the system[49].

HSPGs also have a role in the formation of storage granules in mast cells. Mast cells degranulate as a result of direct injury, the binding of immunoglobulin (Ig)E–antigen complexes to Fc receptors, or receptor activation by certain complement factors, leading to the release of histamine and proteases into the interstitium. These components are normally stored in the secretory granules along with serglycin proteoglycan, which contains both heparin-like chains and chondroitin sulphate (Fig. 2m). Deletion of serglycin severely affects the storage of mast-cell-specific proteases[50]. Mast cells also fail to form properly in mice lacking *Ndst2*, and the few residual cells have reduced amounts of histamine and specific proteases in their secretory granules[51,52]. Interestingly, only minor effects on other haematopoietic cells have been noted in these mutants, even though serglycin seems to be the main proteoglycan in storage granules[53,54].

Many of the processes described above are coordinated during the repair process, in which damaged tissue is removed and tissue architecture is restored. Cutaneous wound-repair models have been studied in a number of HSPG knockouts. In engineered human skin, antisense disruption of perlecan expression in epidermal keratinocytes results in thin and poorly organized epidermis and incomplete stratification[55]. However, removal of the HS attachment sites in the N-terminal domain resulted in only a mild delay in cutaneous wound healing *in vivo*[56]. Inactivation of syndecan-1 or syndecan-4 also causes a mild delay in wound healing[57,58], which suggests that compensation might obscure the contribution of individual proteoglycans. Altering the HSPGs in the wound environment could affect growth-factor sequestration or activation, cell migration and proliferation, and angiogenesis[13]. Another hypothesis is that syndecan-1 and syndecan-4, shed from the cell surface by activated matrix metalloproteinases into the wound fluid, bind and protect elastase and cathepsin G from their physiological inhibitors, α1-antichymotrypsin and α1-protease inhibitor, thus regulating the proteolytic balance of the fluid[59]. A third possibility is that the proteoglycans aid in matrix contraction essential for tissue repair[60].

Regenerative processes that occur during repair depend on stem-cell differentiation. HSPGs are thought to affect proliferation, differentiation and maintenance of the stem-cell niche[61]. An interesting system that is under study involves skeletal muscle regeneration, which requires activation of quiescent satellite cells resident in the tissue. Stimulated satellite cells transiently express perlecan, glypican-1, and syndecan-3 and syndecan-4. Syndecan-3-null mice exhibit hyperplasia of myonuclei and satellite cells and defects in satellite-cell locomotion and differentiation. Syndecan-4-deficient mice also exhibit alterations in satellite-cell proliferation and reduced capacity to reconstitute damaged muscle[62]. Presumably, HSPGs participate in growth control in these systems by serving as co-receptors, but the relevant growth factors involved are not known. Proteoglycans could also have a role in cell migration. Studies

of regeneration in other organs (such as the liver, brain, skeletal system and bone marrow) suggest that the involvement of HSPGs is likely, but genetic studies are needed to verify this hypothesis[61].

Heparan sulphate proteoglycans are full of surprises

A somewhat surprising result of human heparanase overexpression in mice is overgrowth of hair[63]. During hair follicle cycling, heparanase expression occurs in synchrony with migration of follicular-stem-cell progeny into the lower part of the follicle, which is a prerequisite for hair shaft formation. Thus, overexpression of the enzyme could release growth factors that facilitate the migration of follicular-stem-cell progeny. Release of growth factors could also result in increased vascularization, which also facilitates faster hair growth. Studies of hair growth in mice lacking specific proteoglycans and biosynthetic enzymes have not yet been reported.

Haemostasis refers to the homeostatic control of the coagulation pathway to ensure blood flow after vascular injury. The process involves vascular constriction to limit the flow of blood to the injured area, platelet activation and aggregation to form a plug at the site of injury, activation of the coagulation cascade to form a fibrin clot and, ultimately, clot dissolution prior to tissue repair. Several of these steps are inhibited by exogenously supplied heparin, most notably coagulation by activation of antithrombin. Antithrombin activation depends on a specific pentasaccharide sequence containing a central 3-O-sulphated glucosamine residue catalysed by the action of the enzyme heparan sulphate 3-O-sulphotransferase (HS3ST1; Fig. 1). About one-third of a typical heparin preparation, which is purified and fractionated from porcine entrails, contains the active pentasaccharide. This sequence is also found in endothelial HS, in particular on the abluminal side of the endothelium[64], albeit in much lower abundance than in therapeutic heparin[65]. Surprisingly, mice lacking HS3ST1 do not exhibit a procoagulant phenotype, suggesting that endogenous HS bearing the high-affinity antithrombin-binding sequence might not have a role in general haemostasis[66]. There are other HS3ST isozymes that can form the high-affinity binding site, suggesting the possibility of compensation in the system.

Mice defective in HS3ST1 have fertility problems, probably related to defects in ovarian repair after follicle rupture at ovulation and subsequent formation of the corpus luteum. After follicular rupture and oocyte expulsion, proteases are activated, and a fibrin clot is formed, which serves as a provisional matrix for granulosa, theca and endothelial cells, and the formation of highly vascularized luteal tissue. Granulosa cells produce anticoagulant HSPGs that can bind and activate antithrombin, suggesting that the fertility defect in Hs3st1-deficient mice could result from dysregulation of serine proteases[67,68]. At least five HSPGs (perlecan, syndecan-1, -2 and -4, and glypican-1) are synthesized by granulosa cells, but which of these are involved in follicle development and ovulation has not been established[69].

The dark side of heparan sulphate proteoglycans

The examples above show how HSPGs have crucial roles in physiology. However, under certain conditions, HSPGs also contribute to pathophysiology. For example, in cancer, growth-factor-dependent signalling mediated by HSPGs facilitates primary tumour growth and angiogenesis. Tumour HSPGs differ in composition from HSPGs in corresponding normal tissue, which might affect the efficiency of growth-factor stimulation of tumour cells[70]. Similarly, the tumour microenvironment might affect the supporting vasculature, which differs in structure and integrity compared with vessels in normal tissue[23]. These findings suggest the possibility of selective targeting of tumour cells and tumour microvasculature by agents that bind HS or modify its synthesis.

HSPGs also promote amyloid deposition by facilitating formation of insoluble fibrils and stabilizing them against proteolytic degradation. The amyloid plaques seen in disorders such as Alzheimer's disease contain HSPGs, which is consistent with the finding that various amyloidogenic polypeptides bind HS[71]. Recent studies showed that transgenic overexpression of heparanase prevents amyloid deposition

in inflammation-associated amyloidosis in the liver and kidney, where extensive shortening of the HS chains occurred[72]. By contrast, other tissues less affected by the transgene remained susceptible to amyloid deposition. These findings provide direct *in vivo* evidence for HS participation in the development of amyloid disease and suggest the possibility of treating amyloid disorders by altering HS formation or inhibiting HS–amyloid interactions.

A number of studies have shown that syndecan-1 expressed by alveolar epithelia undergoes rapid shedding in response to lung injury, apparently through the matrix metalloproteinase matrilysin[73]. The pathogenic bacterium *Pseudomonas aeruginosa* exploits this response by using the shed ectodomains to increase infection rates and virulence[74]. Newborn mice deficient in syndecan-1 resist infection by *P. aeruginosa*, but become susceptible when the bacteria are mixed with purified syndecan-1 ectodomains or heparin[75]. Although bacteria exploit the ectodomains to increase virulence, the normal process of syndecan shedding seems to attenuate lung inflammation by confining chemokine expression and neutrophil influx to sites of injury and by inhibiting T-cell migration and accumulation[73,76].

Many other organisms (parasites, bacteria and viruses) are thought to use HSPGs as adhesion receptors for infection[77]. However, in most cases the association is based on studies of cultured cells using enzymes to degrade HS, chlorate to block sulphation or mutants to block assembly of the chains. In many cases, binding, invasion and replication have not been distinguished, nor have differences observed in culture been substantiated *in vivo*. Recent studies of *Toxoplasma gondii*, which was thought to require HSPGs for binding and infection, showed that binding and uptake in HS-deficient cells were not affected[78], but that replication of the organism inside vacuoles was reduced[79]. However, the difference in observed growth rate did not translate into altered infection rates *in vivo* in organs in which *Ndst1* was inactivated. This finding raises the possibility that HSPGs mediate signalling reactions between host and parasite. Although it could be argued that more extreme alterations in HS structure could confer resistance to infection, this example illustrates the difficulty in translating results obtained in cultured cells to functional interactions in tissues.

Conclusion

The studies described above and summarized in Table 1 demonstrate the importance of HSPGs in various organ systems. In fact, the data probably underrepresent their significance, because HSPGs interact with so many factors; one would expect few physiological systems to remain unaffected by changes in their composition. Most of these studies have been done in mice, which serve as an excellent model for mammalian physiology and human disease. Many strains bearing systemic mutations are already available, and conditional mutants have been reported for some of the essential enzymes. Thus, we can expect more detailed information about HSPG function in individual organ systems to emerge as investigators from various fields take advantage of these strains.

With a few exceptions, all multicellular organisms produce HSPGs, from ancient cnidarians (*Hydra*) to modern Mammalia. Although a comprehensive comparative study of HSPGs across phyla has not been done, the overall structure of HS seems largely conserved, whereas the core proteins have undergone expansion in number and diversity. Genetic studies of model organisms such as worms, flies and zebrafish have confirmed that many of the basic functions of HSPGs first described in cell-culture systems are conserved, and new mechanistic insights have been gained through the use of mosaic animals, conditional knockouts and gene-silencing techniques. Encouraged by these findings, we should exploit these powerful genetic systems to understand the role of HSPGs in adult physiology and as models for how their function goes awry in human disease. ∎

1. Bulow, H. E. & Hobert, O. The molecular diversity of glycosaminoglycans shapes animal development. *Annu. Rev. Cell Dev. Biol.* **22,** 375–407 (2006).
2. Hacker, U., Nybakken, K. & Perrimon, N. Heparan sulphate proteoglycans: the sweet side of development. *Nature Rev. Mol. Cell Biol.* **6,** 530–541 (2005).

3. Haltiwanger, R. S. & Lowe, J. B. Role of glycosylation in development. *Annu. Rev. Biochem.* **73,** 491–537 (2004).
4. Esko, J. D. & Selleck, S. B. Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annu. Rev. Biochem.* **71,** 435–471 (2002).
5. Dhoot, G. K. *et al.* Regulation of Wnt signaling and embryo patterning by an extracellular sulfatase. *Science* **293,** 1663–1666 (2001).
6. Morimoto-Tomita, M., Uchimura, K., Werb, Z., Hemmerich, S. & Rosen, S. D. Cloning and characterization of two extracellular heparin-degrading endosulfatases in mice and humans. *J. Biol. Chem.* **277,** 49175–49185 (2002).
7. Vlodavsky, I. *et al.* Mammalian heparanase: involvement in cancer metastasis, angiogenesis and normal development. *Semin. Cancer Biol.* **12,** 121–129 (2002).
8. Fedarko, N. S. & Conrad, H. E. A unique heparan sulfate in the nuclei of hepatocytes: structural changes with the growth state of the cells. *J.Cell Biol.* **102,** 587–599 (1986).
9. Huntington, J. A. Mechanisms of glycosaminoglycan activation of the serpins in hemostasis. *J. Thromb. Haemost.* **1,** 1535–1549 (2003).
10. Mohammadi, M., Olsen, S. K. & Ibrahimi, O. A. Structural basis for fibroblast growth factor receptor activation. *Cytokine Growth Factor Rev.* **16,** 107–137 (2005).
11. Forsten-Williams, K., Chua, C. C. & Nugent, M. A. The kinetics of FGF-2 binding to heparan sulfate proteoglycans and MAP kinase signaling. *J. Theor. Biol.* **233,** 483–499 (2005).
12. Kreuger, J., Spillmann, D., Li, J. P. & Lindahl, U. Interactions between heparan sulfate and proteins: the concept of specificity. *J. Cell Biol.* **174,** 323–327 (2006).
13. Alexopoulou, A. N., Multhaupt, H. A. & Couchman, J. R. Syndecans in wound healing, inflammation and vascular biology. *Int. J. Biochem. Cell Biol.* **39,** 505–528 (2006).
14. Couchman, J. R., Chen, L. G. & Woods, A. Syndecans and cell adhesion. *Int. Rev. Cytol.* **207,** 113–150 (2001).
15. Kirkpatrick, C. A. *et al.* The function of a *Drosophila* glypican does not depend entirely on heparan sulfate modification. *Dev. Biol.* **300,** 570–582 (2006).
16. Reizes, O. *et al.* Transgenic expression of syndecan-1 uncovers a physiological control of feeding behavior by syndecan-3. Cell **106,** 105–116 (2001).
17. Strader, A. D., Reizes, O., Woods, S. C., Benoit, S. C. & Seeley, R. J. Mice lacking the syndecan-3 gene are resistant to diet-induced obesity. *J. Clin. Invest.* **114,** 1354–1360 (2004).
18. Zcharia, E. *et al.* Transgenic expression of mammalian heparanase uncovers physiological functions of heparan sulfate in tissue morphogenesis, vascularization, and feeding behavior. *FASEB J.* **18,** 252–263 (2004).
19. Mahley, R. W. & Ji, Z. S. Remnant lipoprotein metabolism: key pathways involving cell-surface heparan sulfate proteoglycans and apolipoprotein E. *J. Lipid Res.* **40,** 1–16 (1999).
20. MacArthur, J. M. *et al.* Liver heparan sulfate proteoglycans mediate clearance of triglyceride-rich lipoproteins independently of LDL receptor family members. *J. Clin. Invest.* **117,** 153–164 (2007).
21. Fuki, I. V. *et al.* The syndecan family of proteoglycans. Novel receptors mediating internalization of atherogenic lipoproteins *in vitro*. *J. Clin. Invest.* **100,** 1611–1622 (1997).
22. Zeng, B. J., Mortimer, B. C., Martins, I. J., Seydel, U. & Redgrave, T. G. Chylomicron remnant uptake is regulated by the expression and function of heparan sulfate proteoglycan in hepatocytes. *J. Lipid Res.* **39,** 845–860 (1998).
23. Iozzo, R. V. Basement membrane proteoglycans: from cellar to ceiling. *Nature Rev. Mol. Cell Biol.* **6,** 646–656 (2005).
24. Raats, C. J. I., Van den Born, J. & Berden, J. H. M. Glomerular heparan sulfate alterations: mechanisms and relevance for proteinuria. *Kidney Int.* **57,** 385–400 (2000).
25. Kanwar, Y. S., Linker, A. & Farquhar, M. G. Increased permeability of the glomerular basement membrane to ferritin after removal of glycosaminoglycans (heparan sulfate) by enzyme digestion. *J. Cell Biol.* **86,** 688–693 (1980).
26. Groffen, A. J. *et al.* Agrin is a major heparan sulfate proteoglycan in the human glomerular basement membrane. *J. Histochem. Cytochem.* **46,** 19–27 (1998).
27. Arikawa-Hirasawa, E., Watanabe, H., Takami, H., Hassell, J. R. & Yamada, Y. Perlecan is essential for cartilage and cephalic development. *Nature Genet.* **23,** 354–358 (1999).
28. Rossi, M. *et al.* Heparan sulfate chains of perlecan are indispensable in the lens capsule but not in the kidney. *EMBO J.* **22,** 236–245 (2003).
29. Morita, H. *et al.* Heparan sulfate of perlecan is involved in glomerular filtration. *J. Am. Soc. Nephrol.* **16,** 1703–1710 (2005).
30. Utriainen, A. *et al.* Structurally altered basement membranes and hydrocephalus in a type XVIII collagen deficient mouse line. *Hum. Mol. Genet.* **13,** 2089–2099 (2004).
31. Gautam, M. *et al.* Defective neuromuscular synaptogenesis in agrin-deficient mutant mice. *Cell* **85,** 525–535 (1996).
32. Westphal, V. *et al.* Reduced heparan sulfate accumulation in enterocytes contributes to protein-losing enteropathy in a congenital disorder of glycosylation. *Am. J. Pathol.* **157,** 1917–1925 (2000).
33. Bode, L., Eklund, E. A., Murch, S. & Freeze, H. H. Heparan sulfate depletion amplifies TNF-α-induced protein leakage in an *in vitro* model of protein-losing enteropathy. *Am. J. Physiol. Gastrointest. Liver Physiol.* **288,** G1015–G1023 (2005).
34. Donnelly, J. P., Rosenthal, A., Castle, V. P. & Holmes, R. D. Reversal of protein-losing enteropathy with heparin therapy in three patients with univentricular hearts and Fontan palliation. *J. Pediatr.* **130,** 474–478 (1997).
35. Kronenberg, H. M. Developmental regulation of the growth plate. *Nature* **423,** 332–336 (2003).
36. Koziel, L., Kunath, M., Kelly, O. G. & Vortkamp, A. Ext1-dependent heparan sulfate regulates the range of Ihh signaling during endochondral ossification. *Dev. Cell* **6,** 801–813 (2004).
37. Lander, A. D., Nie, Q. & Wan, F. Y. Do morphogen gradients arise by diffusion? *Dev. Cell* **2,** 785–796 (2002).
38. Arikawa-Hirasawa, E. *et al.* Dyssegmental dysplasia, Silverman–Handmaker type, is caused by functional null mutations of the perlecan gene. *Nature Genet.* **27,** 431–434 (2001).
39. Stickens, D., Zak, B. M., Rougier, N., Esko, J. D. & Werb, Z. Mice deficient in Ext2 lack heparan sulfate and develop exostoses. *Development* **132,** 5055–5068 (2005).
40. Hovey, R. C., Trott, J. F. & Vonderhaar, B. K. Establishing a framework for the functional mammary gland: from endocrinology to morphology. *J. Mammary Gland Biol. Neoplasia* **7,** 17–38 (2002).
41. Liu, B. Y., McDermott, S. P., Khwaja, S. S. & Alexander, C. M. The transforming activity of Wnt effectors correlates with their ability to induce the accumulation of mammary progenitor cells. *Proc. Natl Acad. Sci. USA* **101,** 4158–4163 (2004).
42. Gautam, M., DeChiara, T. M., Glass, D. J., Yancopoulos, G. D. & Sanes, J. R. Distinct phenotypes of mutant mice lacking agrin, MuSK, or rapsyn. *Brain Res. Dev. Brain Res.* **114,** 171–178 (1999).
43. Jenniskens, G. J. *et al.* Phenotypic knockout of heparan sulfates in myotubes impairs excitation-induced calcium spiking. *FASEB J.* **17,** NIL606–NIL629 (2003).
44. Mook-Jung, I. & Gordon, H. Acetylcholine receptor clustering in C2 muscle cells requires chondroitin sulfate. *J.Neurobiol.* **28,** 482–492 (1995).
45. McDonnell, K. M. & Grow, W. A. Reduced glycosaminoglycan sulfation diminishes the agrin signal transduction pathway. *Dev. Neurosci.* **26,** 1–10 (2004).
46. Kramer, K. L. & Yost, H. J. Ectodermal syndecan-2 mediates left–right axis formation in migrating mesoderm as a cell-nonautonomous Vg1 cofactor. *Dev. Cell* **2,** 115–124(2002).
47. Jakobsson, L. *et al.* Heparan sulfate in *trans* potentiates VEGFR-mediated angiogenesis. *Dev. Cell* **10,** 625–634 (2006).
48. Parish, C. R. The role of heparan sulphate in inflammation. *Nature Rev. Immunol.* **6,** 633–643 (2006).
49. Wang, L., Fuster, M., Sriramarao, P. & Esko, J. D. Endothelial heparan sulfate deficiency impairs L-selectin- and chemokine-mediated neutrophil trafficking during inflammatory responses. *Nature Immunol.* **6,** 902–910 (2005).
50. Abrink, M., Grujic, M. & Pejler, G. Serglycin is essential for maturation of mast cell secretory granule. *J. Biol. Chem.* **279,** 40897–40905 (2004).
51. Humphries, D. E. *et al.* Heparin is essential for the storage of specific granule proteases in mast cells. *Nature* **400,** 769–772 (1999).
52. Forsberg, E. *et al.* Abnormal mast cells in mice deficient in a heparin-synthesizing enzyme. *Nature* **400,** 773–776 (1999).
53. Grujic, M. *et al.* Serglycin-deficient cytotoxic T lymphocytes display defective secretory granule maturation and granzyme B storage. *J. Biol. Chem.* **280,** 33411–33418 (2005).
54. Zernichow, L. *et al.* Serglycin is the major secreted proteoglycan in macrophages and has a role in the regulation of macrophage tumor necrosis factor-α secretion in response to lipopolysaccharide. *J. Biol. Chem.* **281,** 26792–26801 (2006).
55. Sher, I. *et al.* Targeting perlecan in human keratinocytes reveals novel roles for perlecan in epidermal formation. *J. Biol. Chem.* **281,** 5178–5187 (2006).
56. Zhou, Z. *et al.* Impaired angiogenesis, delayed wound healing and retarded tumor growth in perlecan heparan sulfate-deficient mice. *Cancer Res.* **64,** 4699–4702 (2004).
57. Echtermeyer, F. *et al.* Delayed wound repair and impaired angiogenesis in mice lacking syndecan-4. *J. Clin. Invest.* **107,** 9–14 (2001).
58. Stepp, M. A. *et al.* Defects in keratinocyte activation during wound healing in the syndecan-1-deficient mouse. *J. Cell Sci.* **115,** 4517–4531 (2002).
59. Kainulainen, V., Wang, H. M., Schick, C. & Bernfield, M. Syndecans, heparan sulfate proteoglycans, maintain the proteolytic balance of acute wound fluids. *J. Biol. Chem.* **273,** 11563–11569 (1998).
60. Midwood, K. S., Valenick, L. V., Hsia, H. C. & Schwarzbauer, J. E. Coregulation of fibronectin signaling and matrix contraction by tenascin-C and syndecan-4. *Mol. Biol. Cell* **15,** 5670–5677 (2004).
61. Cool, S. M. & Nurcombe, V. Heparan sulfate regulation of progenitor cell fate. *J. Cell Biochem.* **99,** 1040–1051 (2006).
62. Cornelison, D. D. *et al.* Essential and separable roles for syndecan-3 and syndecan-4 in skeletal muscle development and regeneration. *Genes Dev.* **18,** 2231–2236 (2004).
63. Zcharia, E. *et al.* Heparanase regulates murine hair growth. *Am. J. Pathol.* **166,** 999–1008 (2005).
64. De Agostini, A., Watkins, S. C., Slayter, H. S., Youssoufian, H. & Rosenberg, R. D. Localization of anticoagulantly active heparan sulfate proteoglycans in vascular endothelium: antithrombin binding on cultured endothelial cells and perfused rat aorta. *J. Cell Biol.* **111,** 1293–1304 (1990).
65. Marcum, J. A., Fritze, L., Galli, S. J., Karp, G. & Rosenberg, R. D. Microvascular heparin-like species with anticoagulant activity. *Am. J. Physiol.* **245,** H725–H733 (1983).
66. HajMohammadi, S. *et al.* Normal levels of anticoagulant heparan sulfate are not essential for normal hemostasis. *J. Clin. Invest.* **111,** 989–999 (2003).
67. Shworak, N. W., HajMohammadi, S., De Agostini, A. I. & Rosenberg, R. D. Mice deficient in heparan sulfate 3-*O*-sulfotransferase-1: normal hemostasis with unexpected perinatal phenotypes. *Glycoconj. J.* **19,** 355–361 (2002).
68. Hasan, S. *et al.* Coordinate expression of anticoagulant heparan sulfate proteoglycans and serine protease inhibitors in the rat ovary: a potent system of proteolysis control. *Biol. Reprod.* **66,** 144–158 (2002).
69. Hosseini, G., Liu, J. & De Agostini, A. I. Characterization and hormonal modulation of anticoagulant heparan sulfate proteoglycans synthesized by rat ovarian granulosa cells. *J. Biol. Chem.* **271,** 22090–22099 (1996).
70. Fuster, M. M. & Esko, J. D. The sweet and sour of cancer: glycans as novel therapeutic targets. *Nature Rev. Cancer* **5,** 526–542 (2005).
71. van Horssen, J., Wesseling, P., van den Heuvel, L. P., de Waal, R. M. & Verbeek, M. M. Heparan sulphate proteoglycans in Alzheimer's disease and amyloid-related disorders. *Lancet Neurol.* **2,** 482–492 (2003).
72. Li, J. P. *et al.* In vivo fragmentation of heparan sulfate by heparanase overexpression renders mice resistant to amyloid protein A amyloidosis. *Proc. Natl Acad. Sci. USA* **102,** 6473–6477 (2005).
73. Li, Q., Park, P. W., Wilson, C. L. & Parks, W. C. Matrilysin shedding of syndecan-1 regulates chemokine mobilization and transepithelial efflux of neutrophils in acute lung injury. *Cell* **111,** 635–646 (2002).
74. Park, P. W., Pier, G. B., Hinkes, M. T. & Bernfield, M. Exploitation of syndecan-1 shedding by *Pseudomonas aeruginosa* enhances virulence. *Nature* **411,** 98–102 (2001).
75. Haynes, A. *et al.* Syndecan 1 shedding contributes to *Pseudomonas aeruginosa* sepsis. *Infect. Immun.* **73,** 7914–7921 (2005).
76. Xu, J., Park, P. W., Kheradmand, F. & Corry, D. B. Endogenous attenuation of allergic lung inflammation by syndecan-1. *J. Immunol.* **174,** 5758–5765 (2005).
77. Dinglasan, R. R. & Jacobs-Lorena, M. Insight into a conserved lifestyle: protein–carbohydrate adhesion strategies of vector-borne pathogens. *Infect. Immun.* **73,** 7797–7807 (2005).
78. Bishop, J. R. & Esko, J. D. The elusive role of heparan sulfate in *Toxoplasma gondii* infection. *Mol. Biochem. Parasitol.* **139,** 267–269 (2005).

79. Bishop, J. R., Crawford, B. E. & Esko, J. D. Cell surface heparan sulfate promotes replication of *Toxoplasma gondii*. *Infect. Immun.* **73,** 5395–5401 (2005).

80. Cano-Gauci, D. F. *et al.* Glypican-3-deficient mice exhibit developmental overgrowth and some of the abnormalities typical of Simpson-Golabi-Behmel syndrome. *J. Cell Biol.* **146,** 255–264 (1999).

81. Chiao, E. *et al.* Overgrowth of a mouse model of the Simpson–Golabi–Behmel syndrome is independent of IGF signaling. *Dev. Biol.* **243,** 185–206 (2002).

82. Kaksonen, M. *et al.* Syndecan-3-deficient mice exhibit enhanced LTP and impaired hippocampus-dependent memory. *Mol. Cell. Neurosci.* **21,** 158–172 (2002).

83. Grobe, K. *et al.* Cerebral hypoplasia and craniofacial defects in mice lacking heparan sulfate *Ndst1* gene function. *Development* **132,** 3777–3786 (2005).

84. Ringvall, M. *et al.* Defective heparan sulfate biosynthesis and neonatal lethality in mice lacking *N*-deacetylase/*N*-sulfotransferase-1. *J. Biol. Chem.* **275,** 25926–25930 (2000).

85. Pan, Y., Woodbury, A., Esko, J. D., Grobe, K. & Zhang, X. Heparan sulfate biosynthetic gene *Ndst1* is required for FGF signaling in early lens development. *Development* **133,** 4933–4944 (2006).

86. Bullock, S. L., Fletcher, J. M., Beddington, R. S. & Wilson, V. A. Renal agenesis in mice homozygous for a gene trap mutation in the gene encoding heparan sulfate 2-sulfotransferase. *Genes Dev.* **12,** 1894–1906 (1998).

87. Li, J. P. *et al.* Targeted disruption of a murine glucuronyl C5-epimerase gene results in heparan sulfate lacking ʟ-iduronic acid and in neonatal lethality. *J. Biol. Chem.* **278,** 28363–28366 (2003).

88. Serpinskaya, A. S., Feng, G., Sanes, J. R. & Craig, A. M. Synapse formation by hippocampal neurons from agrin-deficient mice. *Dev. Biol.* **205,** 65–78 (1999).

89. Arikawa-Hirasawa, E., Rossi, S. G., Rotundo, R. L. & Yamada, Y. Absence of acetylcholinesterase at the neuromuscular junctions of perlecan-null mice. *Nature Neurosci.* **5,** 119–123 (2002).

90. Viviano, B. L. *et al.* Altered hematopoiesis in glypican-3-deficient mice results in decreased osteoclast differentiation and a delay in endochondral ossification. *Dev. Biol.* **282,** 152–162 (2005).

91. Paine-Saunders, S., Viviano, B. L., Zupicich, J., Skarnes, W. C. & Saunders, S. *glypican-3* controls cellular responses to *Bmp4* in limb patterning and skeletal development. *Dev. Biol.* **225,** 179–187 (2000).

92. Pallerla, S. R., Pan, Y., Zhang, X., Esko, J. D. & Grobe, K. Heparan sulfate *Ndst1* gene function variably regulates multiple signaling pathways during mouse development. *Dev. Dyn.* **236,** 556–563 (2007).

93. Kram, V. *et al.* Heparanase is expressed in osteoblastic cells and stimulates bone formation and bone mass. *J. Cell Physiol.* **207,** 784–792 (2006).

94. Zcharia, E. *et al.* Heparanase accelerates wound angiogenesis and wound healing in mouse and rat models. *FASEB J.* **19,** 211–221 (2005).

95. Li, Q. & Olsen, B. R. Increased angiogenic response in aortic explants of collagen XVIII/endostatin-null mice. *Am. J. Pathol.* **165,** 415–424 (2004).

96. Moulton, K. S. *et al.* Loss of collagen XVIII enhances neovascularization and vascular permeability in atherosclerosis. *Circulation* **110,** 1330–1336 (2004).

97. Gotte, M. *et al.* Role of syndecan-1 in leukocyte–endothelial interactions in the ocular vasculature. *Invest. Ophthalmol. Vis. Sci.* **43,** 1135–1141 (2002).

98. Ishiguro, K. *et al.* Syndecan-4 deficiency impairs the fetal vessels in the placental labyrinth. *Dev. Dyn.* **219,** 539–544 (2000).

99. Fan, G. *et al.* Targeted disruption of NDST-1 gene leads to pulmonary hypoplasia and neonatal respiratory distress in mice. *FEBS Lett.* **467,** 7–11 (2000).

100. Ishiguro, K. *et al.* Syndecan-4 deficiency increases susceptibility to κ-carrageenan-induced renal damage. *Lab. Invest.* **81,** 509–516 (2001).

Exploiting the defensive sugars of HIV-1 for drug and vaccine design

Christopher N. Scanlan[1], John Offer[2], Nicole Zitzmann[1] & Raymond A. Dwek[1]

The sustained effort towards developing an antibody vaccine against HIV/AIDS has provided much of our understanding of viral immunology. It is generally accepted that one of the main barriers to antibody neutralization of HIV is the array of protective structural carbohydrates that covers the antigens on the virus's surface. Intriguingly, however, recent findings suggest that these carbohydrates, which have evolved to protect HIV and promote its transmission, are also attractive therapeutic targets.

A remarkable feature of the human immunodeficiency virus (HIV) is the dense carbohydrate (glycan) array that surrounds the exposed envelope antigens. However, the HIV genome encodes no gene products capable of synthesizing carbohydrates: its surface antigens are glycosylated entirely by host cellular enzymes. This extensive glycosylation is known to affect almost every aspect of virus biology. The folding of viral glycoproteins, the transmission of the virus and the nature of the immune response to infection are all profoundly affected by this glycosylation.

The coating of HIV with immunologically 'self' glycans (that is, those synthesized by the host cell) has a predictable effect on viral immunogenicity: antibodies against most of the available antigenic surface of HIV do not normally occur. Here we review the apparently contradictory roles of HIV glycans as both powerful adaptations for virus survival and targets for therapeutic intervention. Our understanding of the extraordinary extent to which HIV relies on the human glycosylation pathway has exposed new vulnerabilities that are now the target of both drug and vaccine design.

Structure and selection of viral carbohydrates

Synthesis and structure of gp120/gp41 carbohydrates

The types of glycan found on HIV are determined by the interactions between the three-dimensional structure of the envelope proteins and the biosynthetic environment of the cell that the virus has infected[1]. However, the locations of N-linked attachment sites[2] (Asn-X-Ser/Thr-X, where X is not proline) and O-linked attachment sites[3] are directly encoded by the viral genome. Importantly, the positioning of N-linked glycans is well conserved between both isolates and clades, particularly when compared with the high degree of variability within the viral envelope[4,5].

The envelope gene is translated as gp160 and later cleaved by furin proteases into gp120 and gp41. The nascent gp160 is transported by Sec61 to the endoplasmic reticulum, where N-linked glycan precursors ($Glc_3Man_9GlcNAc_2$) are transferred co-translationally to the amide groups of asparagine residues. The first and second glucose residues of $Glc_3Man_9GlcNAc_2$ glycans are removed by glucosidases I and II. Monoglucosylated gp160 is a substrate for the chaperones calnexin and calreticulin. While bound to calnexin/calreticulin, gp160 (re)folds through interactions with other endoplasmic reticulum chaperones and disulphide isomerases such as ERp-57. The hydrolysis of the final glucose–mannose (Glcα1-2Man) bond by glucosidase II frees gp160 from calnexin/calreticulin and allows it to exit from the endoplasmic reticulum. Misfolded glycoproteins are reglucosylated by the 'folding sensor' UDP-glucose glucosyltransferase and can rebind to calnexin/ calreticulin for further refolding cycles. After release from calnexin/calreticulin, the oligomannose-bearing glycoprotein exits from the endoplasmic reticulum and moves into the Golgi, where further mannosidase trimming to $Man_5GlcNAc_2$ initiates the synthesis of complex-type glycans (Fig. 1). However, buried glycans are protected from the processing that normally occurs in the endoplasmic reticulum/Golgi, meaning that they remain as 'immature' precursor glycans[6,7]. In the case of gp160, the steric occlusion of closely spaced N-linked carbohydrates means that less than 50% of the envelope glycans exiting from the Golgi are properly processed[8,9]. A full discussion of N-linked glycosylation is beyond the scope of this review; however, the conserved functional roles of N-linked glycans in the endoplasmic reticulum[10] and the diversification of these structures in the Golgi[11] have been well reviewed elsewhere.

Most of the viral envelope surface is covered by carbohydrates (Fig. 2a). Structural knowledge of HIV gp120 has revealed that within this glycan shield oligomannose glycans are predominantly found in a cluster away from the receptor-binding sites and trimer interface[8] (Fig. 2b, c). This distribution fits well with the 'antigenic map' of gp120 (refs 8, 12–14), which divides the glycoprotein into three regions: the neutralizing face, the non-neutralizing face and the heavily glycosylated silent face (Fig. 3a). Interestingly, sites with oligomannose glycans are more strongly conserved than those with complex-type carbohydrates[4] and are located predominantly on the silent face. A potential explanation for the conservation of the oligomannose glycoforms is the interaction of gp120 with host mannose-specific lectins[15] (as discussed below).

An unusual feature of gp120 is the degree to which sugar–sugar interactions are formed between neighbouring glycans[16,17]. For most glycoproteins, N-linked carbohydrates exhibit considerable conformational freedom beyond that of the protein[18]. However, the glycans of HIV are constrained within tight clusters[16,17]. Unusually extensive electron density is visible for many of the carbohydrates present in the crystal structure of $gp120_{SIV}$ from the closely related simian immunodeficiency virus, indicating a rigid carbohydrate field stabilized by a sugar–sugar hydrogen-bond network[16,17]. Although not unique to HIV (for example, ordered oligomannose glycans have also been observed on the envelope glycoprotein of the Epstein–Barr virus[19]), this atypical clustering has direct functional, immunological and therapeutic consequences.

Biological roles of HIV glycans

Although HIV is generally specific for $CD4^+$ cells, the type of secondary co-receptor that is bound by gp120 defines the exact tropism of the virus. HIV specific for the chemokine co-receptor CCR5 (R5 viruses) is

found during early infection in most individuals, and primarily infects circulating, activated and memory CCR5[+] T lymphocytes and macrophages. CXCR4-specific HIV (X4 viruses) infect a wider range of CD4[+] cells, including naive T cells[20]. There is a direct link between glycosylation and viral tropism: a change in tropism from CCR5 to CXCR4 is closely connected with changes in N-linked glycosylation within the extended variable loops (V1/V2 and V3) of gp120 (refs 20, 21). However, beyond the positioning of glycans, it has also been established that many of the specific carbohydrate motifs found on HIV gp120 have immunological properties. For example, α2-6-linked sialic acids, which are characteristic of glycoproteins from CD4[+] T cells, are known ligands for CD22, an immunosuppressive lectin found on B cells[22]. Bisecting N-acetylglucosamine residues, found on gp120 from T-cell lines[23], can limit natural-killer-cell function. However, the extent to which the types of carbohydrate on HIV have important roles *in vivo* requires further study. The use of gp120 from infected cells rather than cultured cell lines with unrepresentative complex glycosylation will be crucial to the success of such studies.

Given that much of the antigenic surface of gp120 is mannosylated (Fig. 2), we might expect gp120 to be readily neutralized by the mannose-specific lectins of the innate immune system — notably, mannose-binding lectin (MBL). However, although MBL interacts with gp120 *in vitro*, and has been shown to neutralize laboratory-adapted strains of HIV, it only weakly neutralizes primary isolates[24,25]. Indeed, rather than serving as a target for the innate immune system, it seems that HIV uses host lectins for survival. For example, the oligomannose carbohydrates of gp120 bind to C-type lectins such as the mannose receptor, langerin and DC-SIGN (dendritic-cell-specific intercellular adhesion molecule-3-grabbing non-integrin)[26,27]. These lectins are expressed on dendritic cell subsets, including those located in (sub)mucosal tissue. It has been proposed that DC-SIGN[+] cells (and perhaps other C-type lectin-expressing cells such as macrophages[28]) sequester HIV-1 particles by means of a non-infectious pathway, and subsequently present the virus to T cells across an immunological synapse. This is known as *trans* infection. However, the biological significance of *trans* infection is controversial, with recent evidence indicating that direct *cis* infection of dendritic-cell subsets accounts for long-term transfer of the virus to T cells[29]. This is consistent with evidence indicating that the productive infection of dendritic cells by HIV-1 is dependent on, or at least significantly enhanced by, the presence of DC-SIGN[30]. Other lectins are also present on these cells and DC-SIGN-independent transmission of HIV from dendritic cells to T cells has been documented[28,31]. Nonetheless, DC-SIGN/lectin-mediated early infection, whether *cis* or *trans*, represents an attractive target for antimicrobial intervention (as discussed below).

Immune tolerance

The antigenic surfaces of viruses, prokaryotes and eukaryotes are covered by 'shields' of polysaccharides or glycoconjugates such as glycolipids and glycoproteins. Much, if not all, of this carbohydrate diversity is a product of antigenic selection or co-evolution[32,33]. Thus the humoral immune system is highly effective at discriminating self from non-self carbohydrates, as has been revealed by microarray analysis of human serum specificities[34]. Classical examples of immunological discrimination of carbohydrates include the rejection of xeno-tissues expressing α-galactose epitopes or heterologous blood-group antigens (the A, B and H antigens differ only in terms of the identity of the non-reducing terminal monosaccharide).

However, in the case of viruses such as HIV, the relationship between host and pathogen is subverted by the fact that the carbohydrates on the pathogen are themselves synthesized by the host; it might seem inevitable that B-cell self tolerance will limit the scale of the anticarbohydrate response. An interesting exception may be the primary virions that establish the initial infection. These are glycosylated by the donor host and may carry non-self antigens. Indeed, HIV-1 from an AA or AO donor is susceptible to antibody-mediated, complement-dependent neutralization by BB, BO or OO host antibodies and vice versa[35]. It is thought that transmission of viruses is decreased between individuals

(or species) with heterologous blood groups[35,36]. A strategy augmenting this naturally occurring phenomenon could potentially contribute to herd immunity against HIV.

Glycan microheterogeneity

About 1% of the human genome is dedicated to the diversification of glycans in the Golgi[11]. Analysis of gp120 N-linked glycans revealed that each contains an average of 5 (major) glycoforms[8]. Assuming independent variance for 25 sites, this represents a maximum of 5^{25}
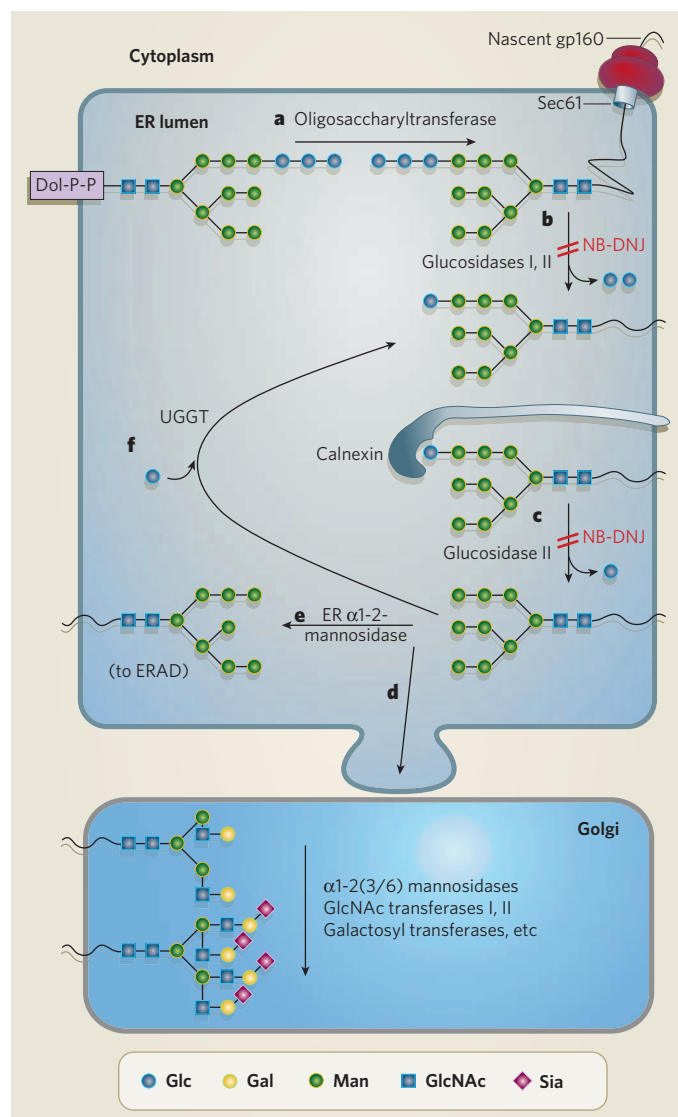
Figure 1 | HIV gp160 N-linked glycosylation. $Glc_3Man_9GlcNAc_2$ is transferred from dolichol pyrophosphate (Dol-P-P) to nascent polypeptides entering the endoplasmic reticulum (ER) through Sec61 (**a**), and this transfer is mediated by oligo-saccharyltransferase, which recognizes N-linked glycosylation sequences (Asn-X-Ser/Thr-X) in gp160. After removal of two glucose units (**b**), the monoglucosylated glycan ($Glc_1Man_9GlcNAc_2$) binds to calnexin or calreticulin and promotes glycoprotein folding. Hydrolysis of the final glucose–mannose bond by glucosidase II frees gp160 from calnexin (**c**), and it can exit from the ER (**d**) and enter the Golgi, where glycoproteins undergo further processing to become complex-type glycans. Glycoproteins with prolonged residence in the ER are eventually subject to trimming by ER α1-2-mannosidase (**e**), which removes a terminal mannose residue from the 'middle' branch (forming $Man_8(B)GlcNAc_2$), a signal for ER-associated degradation (ERAD). Misfolded glycoproteins are reglucosylated (**f**) by UDP-glucose glucosyl transferase (UGGT) and can then rebind calnexin for another cycle of refolding. Inhibitors of the pathway, such as NB-DNJ, can prevent formation of N-glycan intermediate structures that are crucial for the correct folding of the glycoprotein.

different gp120 glycoforms, for any given sequence. The potential chemical diversity of a single HIV clone exceeds the genetic diversity of the entire HIV epidemic. A direct consequence of microheterogeneity is that any neutralizing anticarbohydrate agent may only be effective against a subset of a given viral clone or quasi-species.

Humoral selection of HIV carbohydrates

Numerous studies have shown that glycosylation influences the binding of antibodies against gp120 (refs 5, 37). The acquisition (or loss) of a glycosylation site can have a dramatic effect on the immunogenicity of the surrounding protein surface. For example, glycans on the variable (V1/V2 (ref. 38), V3 (ref. 39) and V4 (ref. 40)) loops of gp120 have been shown to modulate the binding of monoclonal antibodies to these regions of gp120. These 'glycan-dependent' epitopes seem to result from glycan occlusion of proximal protein epitopes, or indirect distal conformational perturbation, rather than direct antibody–carbohydrate binding.

Analysis of HIV envelope sequences indicates that highly variable regions are often adjacent to N-linked glycosylation sites. Moreover, although relatively well conserved, the positions of N-linked glycans do shift during the course of infection. This has led to the concept of the 'evolving glycan shield' of gp120, in which changes in N-linked glycosylation enhance the rate of immune evasion[5]. In this model, the 'dynamic shielding' of the gp120 surface by large glycans amplifies the antigenic effect of relatively small sequence changes (to the Asn-X-Ser/Thr-X motif).

A potential, related role for N-linked glycans in immune escape might be for 'silent' mutations to accumulate within a region of the protein surface that is occluded by carbohydrate. Thus a cryptic pool of variation may evolve whose antigenic exposure is effectively suppressed, until the loss of a 'protective' glycan. This process would provide for the accumulation of diversity while avoiding the fitness costs associated with individual components of that variation. The basic requirements for this process are well established: the protection of protein epitopes by N-linked glycans[38–40] and the location of positive selection in and around shifting glycosylation sites[4,5]. Similar evolutionary 'capacitors' have been suggested for the transient expression of morphological diversity in fruitfly developmental genes[41] and in the prion-mediated translation of normally silent portions of the yeast genome[42].

Accumulation of glycans under antibody selection

The mean number of N-linked sugar sites has not changed significantly during the known course of the HIV-1 epidemic[4]. However, within the course of an infection, there is some evidence for a consistent accumulation of N-linked glycans after establishment of the virus in a new host[43,44]. Similarly, viruses isolated early on after transmission seem to have shorter variable loops and fewer glycosylation sites[45,46]. These observations support a model in which, early in infection and in the absence of antibody-mediated selection, glycosylation sites are dispensed with in favour of replicative efficiency, and only return when the virus is subject to mounting humoral neutralization pressure. However, this phenomenon is not universal, and is proposed to be more prevalent in some viral subtypes than others[45,47,48]. Furthermore, longitudinal studies have reported this effect in only some cases[47,49], indicating that infection-specific factors determine the rate of glycan accumulation.

The most detailed analysis so far sampled the diversity of glycosylation-site mutations within viral quasi-species over time. Divergent envelope sequences (that is, those of quasi-species that had spread most from the inoculum sequence) had acquired extra N-linked glycosylation sites, whereas non-divergent ones had not[50]. Interestingly, within separate, divergent quasi-species there was a remarkable degree of convergent evolution at glycosylation hotspots, suggesting an unexpected constraint on the potential diversity of HIV. The convergent evolution of glycosylation sites may have important implications for both vaccine design and antiviral therapeutic lectins: viruses resistant to such therapy would revert to a sensitive phenotype upon passage to an uninfected host.

Therapeutics to exploit the glycan shield of HIV

Inhibition of glycan biosynthesis prevents infectious virion formation

Several classes of drug are known to inhibit key stages of mammalian glycan biosynthesis. Notable examples are imino sugars, such as N-butyldeoxynojirimycin (NB-DNJ), which can inhibit the trimming of $Glc_3Man_9GlcNAc_2$ to $Glc_1Man_9GlcNAc_2$ in the endoplasmic reticulum and prevent entry of its carrier glycoprotein into the calnexin/calreticulin folding cycle. NB-DNJ treatment leads to structural alterations in gp120 (ref. 51), and viruses expressed in the presence of this drug fail to undergo productive post-receptor-binding rearrangements or fusion with the target-cell membrane[51–53].

Interestingly, despite its inhibition of a host process, NB-DNJ is moderately well tolerated in humans[54,55]. It is already used at low concentrations for the treatment of glycolipid storage disorders[56], although at the higher concentrations needed for antiviral activity NB-DNJ causes unwanted inhibition of disaccharidases in the intestine[54,55]. The molecular basis for the enhanced sensitivity of viral, but not host, glycoproteins to glucosidase inhibitors remains unknown and warrants further investigation. Potential explanations may involve the link between lectin-mediated retention of glucosylated gp120 in the endoplasmic reticulum
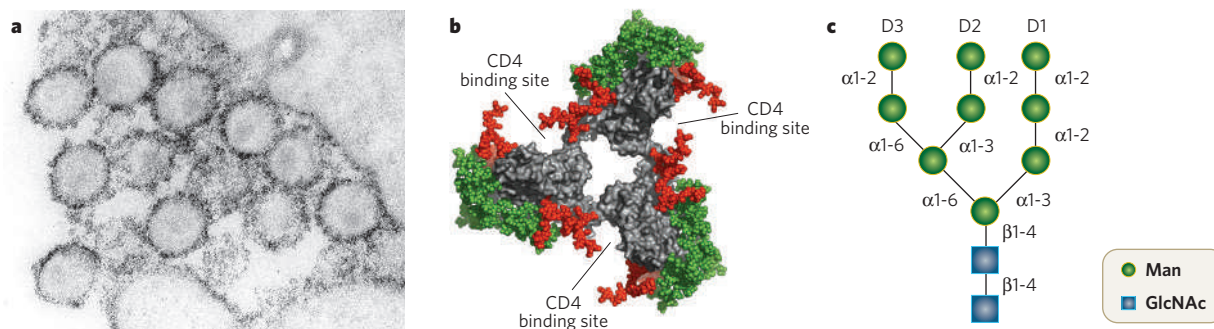
Figure 2 | The carbohydrate shield of gp120. a, Electron micrograph of HIV-1. Carbohydrates are stained with ruthenium red (dark), showing extensive occlusion of the antigenic surface by host-derived carbohydrates. The outer domain of gp120 is particularly rich in conserved glycosylation sites. (Image courtesy of CDC public image database.) **b**, The glycosylation surface of gp120 can therefore be roughly partitioned into two regions: the oligomannose glycans (green), found on the densely glycosylated outer domain, and the complex sugars (red), which are distributed on the more exposed receptor-binding sites and hypervariable loops[8]. The site-specific analysis of monomeric gp120 used to provide these data may not fully reflect the occlusion that occurs within a trimer, so the abundance of oligomannose glycans is likely to exceed that of the related monomer. (Molecular model of gp120 based on crystal structure of CD4-liganded core[14,97], with perspective shown being that from the viral membrane. Model courtesy of M. Wormald.) **c**, Representation of the structure of the oligomannose glycan $Man_9GlcNAc_2$, showing the structure, glycosidic linkages and identity of the D1, D2 and D3 termini of the A, B and C arms.

and gp120's complex disulphide bond network that is required for productive infection by T-cell thiol-reactive proteins[57,58]. Alternatively, the sensitivity of gp120 to glucosidase inhibition may simply be an example of the proposed correlation between N-glycosylation density and calnexin/calreticulin dependence[10].

Lectin-mediated HIV neutralization

The glycans of gp120 are attractive targets for exogenous lectins that prevent viral infection or transmission. The primary mechanisms of these lectins are competitive inhibition of the association of gp120 with DC-SIGN[59] or disruption of receptor-induced conformational changes and/or inhibition of membrane fusion[60]. A wide range of plant, animal and microbial lectins have been assayed for efficacy against HIV. Many of these lectins bind to the mannose residues of gp120, but other monosaccharides (such as N-acetylglucosamine, galactose and fucose) are also candidates. Lectin specificity for HIV, and not for host proteins, seems to rely on the unusual density of the sugars found on gp120 and their specificity for terminal residues, such as mannose, that are not normally observed in mammalian proteins[61].

In the absence of an effective vaccine for HIV/AIDS (acquired immunodeficiency syndrome), microbicidal lectins are promising complements to traditional barrier protection; at least one, cyanovirin, has shown efficacy in vivo[62]. Cyanovirin is a small bacterial lectin with a specificity for α1-2-linked mannose residues, similar to that of the antibody 2G12 (see below). Although systemic use of antiviral lectins has yet to be investigated, their ability to prevent cell–cell virus transfer[63] could make them a valuable therapeutic tool. The more general specificity of lectins, compared with that of traditional small-molecule inhibitors (such as those targeting the viral reverse transcriptase and protease[64]), makes them potentially powerful antivirals. Many independent mutations[65,66] are required for viral escape. However, the development of viable therapeutic lectins is not without challenges, and the mucosal or systemic immunogenicity of these foreign antigens and the risk of cross-reactivity with host carbohydrates both need to be addressed[61].

The future of anticarbohydrate therapeutics

HIV has evolved to exploit the human glycosylation machinery. Although this yields considerable advantages for the virus, the sheer extent of this glycosylation provides a window of discrimination between host and virus. Of course, it is possible that HIV could evolve to dispense with its glycosylation under drug or lectin selection. However, because glucosidases I and II are host enzymes, viral mutants that escape the drugs targeting these proteins might be less likely to arise. It has also been argued that escape from the selection pressure targeted at HIV glycosylation would remove a major defensive strength from the virus[67]. Immune escape from antiviral therapy, although possible, might be bought at a significant cost to the virus's inherent ability to evade immunity. Moreover, as described above, in the absence of selection, identical glycans might convergently evolve (that is, reappear) in a new host. Thus, at an epidemiological level, resistance may be limited by the selection pressures that drive the formation of the glycan shield of gp120.

Vaccines to exploit the glycan shield of HIV
Neutralization of HIV by a carbohydrate-binding antibody

Most antibodies against HIV-1 are either directed against non-neutralizing epitopes (for example, those found on monomeric gp120 but not on functional trimers; Fig. 3a) or exert a selection pressure that HIV rapidly evades through changes in its envelope sequence[68,69]. Similarly, most immunogens based on the envelope proteins of HIV elicit a narrow antibody response, specific only for non-neutralizing epitopes or those that are poorly conserved between strains[70,71]. Nonetheless, a few antibodies, isolated from infected individuals, do exhibit a potent neutralizing activity against a broad range of HIV isolates[68,72–78]. Remarkably, given the generally poor immunogenicity of heterogeneous, self glycans outlined above, one of these broadly neutralizing antibodies, IgG1 2G12, binds directly and exclusively to the carbohydrates of gp120 (refs 15, 75, 79–81; Fig. 3b).
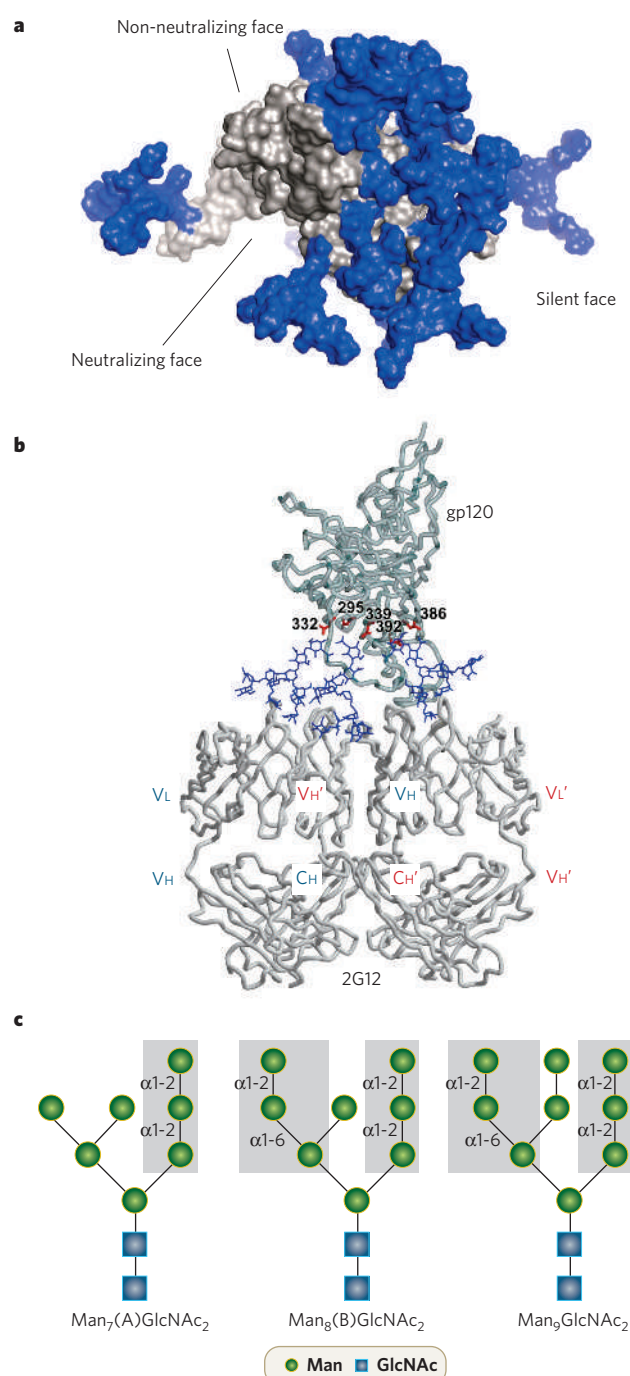
Figure 3 | Antigenicity and glycosylation of HIV gp120. a, Antigenic map of monomeric HIV gp120, based on the crystal structure of gp120 (ref. 13). The neutralizing face contains the receptor-binding sites, and the non-neutralizing face contains epitopes that are exposed on free, monomeric gp120 but are hidden by adjacent gp120/gp41 subunits on the functional, trimeric form of gp120/gp41 (ref. 98). The silent face is heavily glycosylated (with N-linked glycans (blue)) and, with the known exception of the 2G12 epitope, immunosilent. **b**, Interaction between gp120 and IgG 2G12. The domain-exchanged configuration of 2G12 results from a non-covalent interaction between two heavy chains (red and blue) that provides an extended paratope capable of high-affinity (multivalent) interaction with gp120 sugars[81]. The positioning of the reducing termini of the 2G12-bound glycans is consistent with mutagenesis data that indicate that the oligomannose glycans at Asn 332 and Asn 392 of gp120 are critical for 2G12 binding. (Model of 2G12–gp120 complex reproduced, with permission, from ref. 81.) **c**, The Manα1-2Manα1-2Man and Manα1-2Manα1-6Man motifs recognized by 2G12 (refs 80, 99; grey shading) are found on gp120's main oligomannose glycoforms, thus minimizing the effective microheterogeneity of these glycans.

The structure and specificity of the 2G12 epitope

The 2G12 epitope is formed in the dense cluster of oligomannose glycans found on the outer domain of gp120 (refs 15, 79). The antibody itself adopts a unique domain-exchanged configuration, with the variable (V_H) domains of the two heavy chains crossing over at the junction between the first constant (C_H1) domain and the V_H domain to form a non-standard I-shaped antibody[81]. The antibody-binding fragment (Fab) accommodates Manα1-2Man residues, provided by the non-reducing termini of $Man_{6-9}GlcNAc_2$ glycans, along the extended paratope formed by adjacent V_H domains. The 2G12 structure, combined with studies of its epitope and specificity (Fig. 3b, c), reveals how this antibody is capable of high-affinity interaction with a broad range of viral isolates (41%, with low activity against clade C, owing to the absence of a required key glycosylation site in gp120

(refs 15, 76)). The glycan cluster is conserved between isolates, and the carbohydrate-recognition motif is conserved between glycoforms. This conservation is consistent with the hypothesis that the 2G12 epitope is maintained by the convergent evolution of glycosylation sites driven by neutralizing antibody selection[43]. The clustering of self sugars in a non-self manner provides the basis for immunological discrimination[15,79].

Progress and obstacles for synthetic mannose vaccines against HIV

Chemical synthesis directed towards molecular mimics of the 2G12 epitope have yielded antigens that can interact with 2G12 (refs 80, 82–87). The advantage of the synthetic approach is the potential variation in design for antigen optimization and presentation. However, one potential disadvantage of the chemical approaches reported so far for
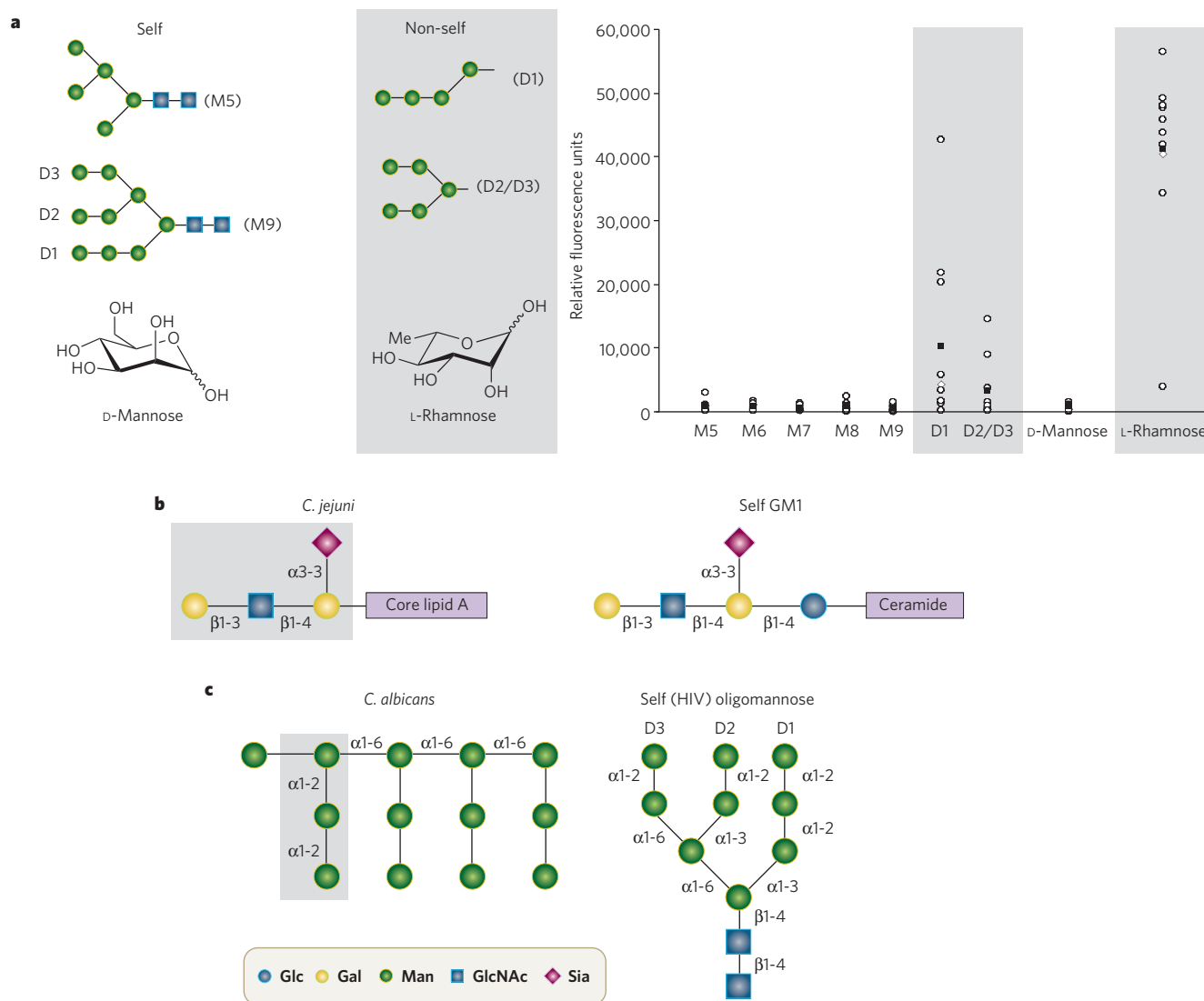
Figure 4 | Self/non-self discrimination and antigenic mimicry. a, The right panel shows the extent of binding of human sera (*n* = 10, open circles; mean values are shown as filled boxes) to a glycan microarray, as determined using fluorescently labelled antihuman antibodies[34]. The discrimination between self mannosides and non-self mannosides (grey shading) is closely regulated, presenting a challenge to vaccine design. The carbohydrates that bind 2G12 — Manα1-2Manα1-2Manα1-3Man (compound D1, left panel) and (Manα1-2Manα1-6)(Manα1-2Manα1-3)Man (compound D2/D3) — are naturally antigenic. These antigenic structures are nonetheless immunosilent in the context of a self glycan.

The immunological discrimination is evident at an atomic level: changing the optical configuration and substitution at a single carbon position changes the antigenicity of the monosaccharide D-mannose, compared with L-rhamnose. Thus L-rhamnose may be a powerful antigenic monosaccharide for inclusion in synthetic 'mannose' vaccines. **b,** Carbohydrate mimicry between *C. jejuni* and self ganglioside GM1. Autoimmune neuropathies, including Guillain–Barré syndrome, are a direct consequence of this molecular mimicry[100]. **c,** Carbohydrate mimicry between the mannan of *C. albicans* and the oligomannose glycans of gp120.

2G12 is the degree of internal flexibility evident within the compounds. Thus a unique biophysical feature of gp120 (the inter-glycan hydrogen-bond network figure) is not replicated[17]. The high affinity of 2G12 for gp120 derives at least in part from the lowered entropic penalty normally associated with the binding of unconstrained glycans[81]. The inherent flexibility of synthetic vaccines may limit the affinity maturation of any cognate antibody or B-cell receptor, not just 2G12. Thus, in this limited case, specific antigenicity for 2G12 is also a measure of immunogenicity.

The seemingly unique status of 2G12-like antibodies presents a serious challenge to the synthetic strategy[88]. If 2G12-like antibodies are hardly ever elicited by what is probably their optimal cognate epitope (the silent face of gp120), then how can mimics of the same structure be expected to be any more effective? Implicit in this challenge are two assumptions that warrant inspection. The first is that all protective anti-oligomannose antibodies against gp120 would need to resemble the highly mutated, domain-exchanged configuration of the 2G12 Fab. The second is that the best immunogen for a given epitope is that with the closest homology.

It is tempting to assume that the unusual nature of the 2G12 epitope[15,79] and the specific structural basis for its recognition[81] are related. However, this does not exclude the possibility of alternative modes of recognition of gp120 glycans. For example, it would be sterically possible for two Fab arms to bind to widely spaced carbohydrates within a single gp120 monomer, or for the trimeric form of gp120 to accommodate binding of Fabs across different monomeric subunits. Provided that the immunological constraints that seem to have driven the formation of 2G12 (discrimination from self glycosylation, high avidity and minimal microheterogeneity[79]) are observed, it may be possible for alternative structural determinants to be included in a gp120-mimetic immunogen. Moreover, the tendency for convergent evolution[43,50] of oligomannose glycosylation sites suggests that an oligomannose vaccine might have the unusual property of being assisted by the HIV-resistance mechanisms that normally protect the virus from antibody neutralization.

Convergent approaches towards anticancer and HIV vaccine design
The challenge of eliciting high-affinity antibodies against glycans that are fundamentally self units is not without precedent. Elevated or altered expression of self glycolipids (such as Globo-H, GM2 and GM3) and glycoproteins (including the T_N, sialyl-TN (ST_N), and sialyl-LewisA (SLeA), SLeX and SLeY epitopes) is associated with many cancers[89]. The concept of clustering of normally self antigens to improve immunogenicity, and the importance of carrier peptides such as keyhole limpet haemocyanin[90], has been investigated, along with strategies such as immunization with peptide mimetics of glycan structures[91] (an approach that has also been explored for HIV[92]). Although these approaches have yet to translate into clinical vaccines (see page 1000), it has been established that high-titre, class-switched antibodies against self carbohydrate epitopes can be elicited[93,94]. This provides an important rational justification for research into carbohydrate vaccines against HIV-1. Constructs, carriers and adjuvants that prove successful for cancer antigens should be considered for use in HIV vaccine development.

Drawing clues from nature to develop a carbohydrate vaccine for HIV-1
An interesting feature of the humoral anticarbohydrate repertoire is the inherent specificity against oligosaccharide motifs that constitute the 2G12 epitope. For example, the serum antibody recognition of the D1 and D2/D3 termini is similar to the specificity of 2G12 itself[34,80] (Fig. 4a). However, unlike 2G12, serum anti-oligomannose antibodies do not bind strongly to these structures when part of the larger self ($Man_{5-9}GlcNAc_2$) structures. So self motifs are antigenic when out of context.

The ability to overcome self tolerance to carbohydrate structures — the goal of both anticancer and HIV vaccines — is already exhibited by some pathogenic antigens. Infection with *Campylobacter jejuni* elicits IgG antibodies against the bacterial lipo-oligosaccharides that also bind identical structures found on the gangliosides of peripheral nervous tissue[95] (resulting in Guillain–Barré syndrome; Fig. 4b). The role of immunogenic carriers (for example, core lipid A) in breaking immune tolerance to a self antigen might be explored in the design of oligomannose vaccines. Thus, a component of pathogenesis in one disease (in this case Guillain–Barré) might be translated into a tool for vaccine design against another (HIV).

The search for immunogens that can elicit 2G12-like antibodies should include those pathogens whose mannose epitopes already seem to elicit a strong antimannose response. HIV is not the only pathogen to include the Manα1-2Man motif in its surface. Several such pathogens are endemic to the human population; some, such as *Candida albicans,* are particularly associated with HIV-1 infection[96] (Fig. 4c). The B-cell clone that was the ancestor of 2G12 may have initially recognized mannose-type structures on a pathogen other than HIV. A conceptually related approach is to chemically modify the glycan structures of the 2G12 epitope to include antigenic carbohydrates. Again, microarray analysis of serum antibodies indicates that rhamnose (6-deoxymannose), although structurally related to mannose, is strongly antigenic[34]. Introduction of antigenic motifs into normally immunosilent epitopes might help to break the immune silence of HIV carbohydrates.

Perspective and future directions
Immune selection against HIV-1 has driven the evolution of a dense array of N-linked carbohydrate sites[43,50]. Although the targeting of a host cell property presents challenges in the form of either toxicity (in the case of lectins and glycosylation inhibitors) or autoimmunity (in the case of vaccines), the extensive and unusual nature of HIV glycosylation may provide a window of discrimination between the function and structure of human glycans and those of HIV. From the perspective of vaccine design, it is known that the human immune system is capable of providing at least one solution to the recognition of self glycans on HIV. In the broadest sense, research towards a carbohydrate vaccine for HIV must address the question of how or why some antigens are able to elicit an immune response to carbohydrate structures with self components. Observations from autoimmunity suggest that the answer to this enigma may already exist in nature: the challenge is to find a solution specific to HIV. ■

1. Clark, G. F. *et al.* Viewing AIDS from a glycobiological perspective: potential linkages to the human fetoembryonic defence system hypothesis. *Mol. Hum. Reprod.* **3,** 5–13 (1997).
2. Ben-Dor, S., Esterman, N., Rubin, E. & Sharon, N. Biases and complex patterns in the residues flanking protein N-glycosylation sites. *Glycobiology* **14,** 95–101 (2004).
3. Gupta, R., Birch, H., Rapacki, K., Brunak, S. & Hansen, J. E. O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins. *Nucleic Acids Res.* **27,** 370–372 (1999).
4. Zhang, M. *et al.* Tracking global patterns of N-linked glycosylation site variation in highly variable viral glycoproteins: HIV, SIV, and HCV envelopes and influenza hemagglutinin. *Glycobiology* **14,** 1229–1246 (2004).
5. Wei, X. *et al.* Antibody neutralization and escape by HIV-1. *Nature* **422,** 307–312 (2003).
6. Crispin, M. D. *et al.* Monoglucosylated glycans in the secreted human complement component C3: implications for protein biosynthesis and structure. *FEBS Lett.* **566,** 270–274 (2004).
7. Janssen, B. J. *et al.* Structures of complement component C3 provide insights into the function and evolution of immunity. *Nature* **437,** 505–511 (2005).
8. Zhu, X., Borchers, C., Bienstock, R. J. & Tomer, K. B. Mass spectrometric characterization of the glycosylation pattern of HIV-gp120 expressed in CHO cells. *Biochemistry* **39,** 11194–11204 (2000).
9. Cutalo, J. M., Deterding, L. J. & Tomer, K. B. Characterization of glycopeptides from HIV-I(SF2) gp120 by liquid chromatography mass spectrometry. *J. Am. Soc. Mass Spectrom.* **15,** 1545–1555 (2004).
10. Kwong, P. D., Wyatt, R., Sattentau, Q. J., Sodroski, J. & Hendrickson, W. A. Oligomeric modeling and electrostatic analysis of the gp120 envelope glycoprotein of human immunodeficiency virus. *J. Virol.* **74,** 1961–1972 (2000).
11. Wyatt, R. *et al.* The antigenic structure of the HIV gp120 envelope glycoprotein. *Nature* **393,** 705–711 (1998).
12. Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393,** 648–659 (1998).
13. Sanders, R. W. *et al.* The mannose-dependent epitope for neutralizing antibody 2G12 on human immunodeficiency virus type 1 glycoprotein gp120. *J. Virol.* **76,** 7293–7305 (2002).
14. Chen, B. *et al.* Structure of an unliganded simian immunodeficiency virus gp120 core. *Nature* **433,** 834–841 (2005).
15. Chen, B. *et al.* Determining the structure of an unliganded and fully glycosylated SIV gp120 envelope glycoprotein. *Structure* **13,** 197–211 (2005).
16. Woods, R. J., Pathiaseril, A., Wormald, M. R., Edge, C. J. & Dwek, R. A. The high degree of internal flexibility observed for an oligomannose oligosaccharide does not alter the overall topology of the molecule. *Eur. J. Biochem.* **258,** 372–386 (1998).

17. Szakonyi, G. *et al.* Structure of the Epstein–Barr virus major envelope glycoprotein. *Nature Struct. Mol. Biol.* **13,** 996–1001 (2006).

18. Moore, J. P., Kitchen, S. G., Pugach, P. & Zack, J. A. The CCR5 and CXCR4 coreceptors — central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. *AIDS Res. Hum. Retroviruses* **20,** 111–126 (2004).

19. Pollakis, G. *et al.* N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. *J. Biol. Chem.* **276,** 13433–13441 (2001).

20. Haas, K. M. *et al.* CD22 ligand binding regulates normal and malignant B lymphocyte survival *in vivo*. *J. Immunol.* **177,** 3063–3073 (2006).

21. Mizuochi, T. *et al.* Diversity of oligosaccharide structures on the envelope glycoprotein gp 120 of human immunodeficiency virus 1 from the lymphoblastoid cell line H9. Presence of complex-type oligosaccharides with bisecting *N*-acetylglucosamine residues. *J. Biol. Chem.* **265,** 8519–8524 (1990).

22. Spear, G. T., Zariffard, M. R., Xin, J. & Saifuddin, M. Inhibition of DC-SIGN-mediated *trans* infection of T cells by mannose-binding lectin. *Immunology* **110,** 80–85 (2003).

23. Ying, H. *et al.* Interaction of mannose-binding lectin with HIV type 1 is sufficient for virus opsonization but not neutralization. *AIDS Res. Hum. Retroviruses* **20,** 327–335 (2004).

24. Feinberg, H., Mitchell, D. A., Drickamer, K. & Weis, W. I. Structural basis for selective recognition of oligosaccharides by DC-SIGN and DC-SIGNR. *Science* **294,** 2163–2166 (2001).

25. Wu, L. & KewalRamani, V. N. Dendritic-cell interactions with HIV: infection and viral dissemination. *Nature Rev. Immunol.* **6,** 859–868 (2006).

26. 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).

27. Nobile, C. *et al.* Covert human immunodeficiency virus replication in dendritic cells and in DC-SIGN-expressing cells promotes long-term transmission to lymphocytes. *J. Virol.* **79,** 5386–5399 (2005).

28. Turville, S. G. *et al.* HIV gp120 receptors on human dendritic cells. *Blood* **98,** 2482–2488 (2001).

29. Boggiano, C., Manel, N. & Littman, D. R. Dendritic cell-mediated *trans*-enhancement of HIV-1 infectivity is independent of DC-SIGN. *J. Virol.* **81,** 2519–2523 (2007).

30. Varki, A. Nothing in glycobiology makes sense, except in the light of evolution. *Cell* **126,** 841–845 (2006).

31. Gagneux, P. & Varki, A. Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* **9,** 747–755 (1999).

32. Blixt, O. *et al.* Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl Acad. Sci. USA* **101,** 17033–17038 (2004).

33. Neil, S. J., McKnight, A., Gustafsson, K. & Weiss, R. A. HIV-1 incorporates ABO histo-blood group antigens that sensitize virions to complement-mediated inactivation. *Blood* **105,** 4693–4699 (2005).

34. Seymour, R. M., Allan, M. J., Pomiankowski, A. & Gustafsson, K. Evolution of the human ABO polymorphism by two complementary selective pressures. *Proc. R. Soc. Lond. B* **271,** 1065–1072 (2004).

35. Lowe, J. B. & Marth, J. D. A genetic approach to mammalian glycan function. *Annu. Rev. Biochem.* **72,** 643–691 (2003).

36. Reitter, J. N., Means, R. E. & Desrosiers, R. C. A role for carbohydrates in immune evasion in AIDS. *Nature Med.* **4,** 679–684 (1998).

37. Cole, K. S., Steckbeck, J. D., Rowles, J. L., Desrosiers, R. C. & Montelaro, R. C. Removal of N-linked glycosylation sites in the V1 region of simian immunodeficiency virus gp120 results in redirection of B-cell responses to V3. *J. Virol.* **78,** 1525–1539 (2004).

38. McCaffrey, R. A., Saunders, C., Hensel, M. & Stamatatos, L. N-linked glycosylation of the V3 loop and the immunologically silent face of gp120 protects human immunodeficiency virus type 1 SF162 from neutralization by anti-gp120 and anti-gp41 antibodies. *J. Virol.* **78,** 3279–3295 (2004).

39. Bolmstedt, A. *et al.* Influence of N-linked glycans in V4–V5 region of human immunodeficiency virus type 1 glycoprotein gp160 on induction of a virus-neutralizing humoral response. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **12,** 213–220 (1996).

40. Queitsch, C., Sangster, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417,** 618–624 (2002).

41. True, H. L., Berlin, I. & Lindquist, S. L. Epigenetic regulation of translation reveals hidden genetic variation to produce complex traits. *Nature* **431,** 184–187 (2004).

42. Dacheux, L. *et al.* Evolutionary dynamics of the glycan shield of the human immunodeficiency virus envelope during natural infection and implications for exposure of the 2G12 epitope. *J. Virol.* **78,** 12625–12637 (2004).

43. Sagar, M., Wu, X., Lee, S. & Overbaugh, J. Human immunodeficiency virus type 1 V1-V2 envelope loop sequences expand and add glycosylation sites over the course of infection, and these modifications affect antibody neutralization sensitivity. *J. Virol.* **80,** 9586–9598 (2006).

44. Derdeyn, C. A. *et al.* Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science* **303,** 2019–2022 (2004).

45. Li, M. *et al.* Genetic and neutralization properties of subtype C human immunodeficiency virus type 1 molecular env clones from acute and early heterosexually acquired infections in Southern Africa. *J. Virol.* **80,** 11776–11790 (2006).

46. Frost, S. D. *et al.* Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. *Proc. Natl Acad. Sci. USA* **102,** 18514–18519 (2005).

47. Li, B. *et al.* Evidence for potent autologous neutralizing antibody titers and compact envelopes in early infection with subtype C human immunodeficiency virus type 1. *J. Virol.* **80,** 5211–5218 (2006).

48. Zhang, H. *et al.* Characterization of HIV-1 subtype C envelope glycoproteins from perinatally infected children with different courses of disease. *Retrovirology* [online] **3,** 73 (2006) (doi:10.1186/1742-4690-3-73).

49. Blay, W. M. *et al.* Consistent patterns of change during the divergence of human immunodeficiency virus type 1 envelope from that of the inoculated virus in simian/human immunodeficiency virus-infected macaques. *J. Virol.* **80,** 999–1014 (2006).

50. Fischer, P. B., Karlsson, G. B., Butters, T. D., Dwek, R. A. & Platt, F. M. *N*-butyldeoxynojirimycin-mediated inhibition of human immunodeficiency virus entry correlates with changes in antibody recognition of the V1/V2 region of gp120. *J. Virol.* **70,** 7143–7152 (1996).

51. Fischer, P. B. *et al.* The alpha-glucosidase inhibitor *N*-butyldeoxynojirimycin inhibits human immunodeficiency virus entry at the level of post-CD4 binding. *J. Virol.* **69,** 5791–5797 (1995).

52. Zitzmann, N. *et al.* Imino sugars inhibit the formation and secretion of bovine viral diarrhea virus, a pestivirus model of hepatitis C virus: implications for the development of broad spectrum anti-hepatitis virus agents. *Proc. Natl Acad. Sci. USA* **96,** 11878–11882 (1999).

53. Fischl, M. A. *et al.* The safety and efficacy of combination *N*-butyl-deoxynojirimycin (SC-48334) and zidovudine in patients with HIV-1 infection and 200–500 CD4 cells/mm³. *J. Acquir. Immune Defic. Syndr.* **7,** 139–147 (1994).

54. Tierney, M. *et al.* The tolerability and pharmacokinetics of *N*-butyl-deoxynojirimycin in patients with advanced HIV disease (ACTG 100). The AIDS Clinical Trials Group (ACTG) of the National Institute of Allergy and Infectious Diseases. *J. Acquir. Immune Defic. Syndr. Hum. Retrovirol.* **10,** 549–553 (1995).

55. Butters, T. D., Dwek, R. A. & Platt, F. M. Imino sugar inhibitors for treating the lysosomal glycosphingolipidoses. *Glycobiology* **15,** 43R–52R (2005).

56. Ou, W. & Silver, J. Role of protein disulfide isomerase and other thiol-reactive proteins in HIV-1 envelope protein-mediated fusion. *Virology* **350,** 406–417 (2006).

57. Barbouche, R., Miquelis, R., Jones, I. M. & Fenouillet, E. Protein-disulfide isomerase-mediated reduction of two disulfide bonds of HIV envelope glycoprotein 120 occurs post-CXCR4 binding and is required for fusion. *J. Biol. Chem.* **278,** 3131–3136 (2003).

58. Helenius, A. & Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **73,** 1019–1049 (2004).

59. Balzarini, J., Van Herrewege, Y., Vermeire, K., Vanham, G. & Schols, D. Carbohydrate-binding agents efficiently prevent dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin (DC-SIGN)-directed HIV-1 transmission to T lymphocytes. *Mol. Pharmacol.* **71,** 3–11 (2007).

60. Esser, M. T. *et al.* Cyanovirin-N binds to gp120 to interfere with CD4-dependent human immunodeficiency virus type 1 virion binding, fusion, and infectivity but does not affect the CD4 binding site on gp120 or soluble CD4-induced conformational changes in gp120. *J. Virol.* **73,** 4360–4371 (1999).

61. Balzarini, J. Inhibition of HIV entry by carbohydrate-binding proteins. *Antiviral Res.* **71,** 237–247 (2006).

62. Tsai, C. C. *et al.* Cyanovirin-N inhibits AIDS virus infections in vaginal transmission models. *AIDS Res. Hum. Retroviruses* **20,** 11–18 (2004).

63. Chiba, H. *et al.* Actinohivin, a novel anti-HIV protein from an actinomycete that inhibits syncytium formation: isolation, characterization, and biological activities. *Biochem. Biophys. Res. Commun.* **282,** 595–601 (2001).

64. Kutilek, V. D., Sheeter, D. A., Elder, J. H. & Torbett, B. E. Is resistance futile? *Curr. Drug Targets Infect. Disord.* **3,** 295–309 (2003).

65. Balzarini, J. *et al.* Mutational pathways, resistance profile, and side effects of cyanovirin relative to human immunodeficiency virus type 1 strains with N-glycan deletions in their gp120 envelopes. *J. Virol.* **80,** 8411–8421 (2006).

66. Witvrouw, M. *et al.* Resistance of human immunodeficiency virus type 1 to the high-mannose binding agents cyanovirin N and concanavalin A. *J. Virol.* **79,** 7777–7784 (2005).

67. Balzarini, J. Targeting the glycans of gp120: a novel approach aimed at the Achilles heel of HIV. *Lancet Infect. Dis.* **5,** 726–731 (2005).

68. Pantophlet, R. & Burton, D. R. gp120: target for neutralizing HIV-1 antibodies. *Annu. Rev. Immunol.* **24,** 739–769 (2006).

69. Burton, D. R., Stanfield, R. L. & Wilson, I. A. Antibody vs. HIV in a clash of evolutionary titans. *Proc. Natl Acad. Sci. USA* **102,** 14943–14948 (2005).

70. Moore, J. P. *et al.* Primary isolates of human immunodeficiency virus type 1 are relatively resistant to neutralization by monoclonal antibodies to gp120, and their neutralization is not predicted by studies with monomeric gp120. *J. Virol.* **69,** 101–109 (1995).

71. Gilbert, P. B. *et al.* Correlation between immunologic responses to a recombinant glycoprotein 120 vaccine and incidence of HIV-1 infection in a phase 3 HIV-1 preventive vaccine trial. *J. Infect. Dis.* **191,** 666–677 (2005).

72. Zhou, T. *et al.* Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445,** 732–737 (2007).

73. Zolla-Pazner, S. Identifying epitopes of HIV-1 that induce protective antibodies. *Nature Rev. Immunol.* **4,** 199–210 (2004).

74. Stiegler, G. *et al.* A potent cross-clade neutralizing human monoclonal antibody against a novel epitope on gp41 of human immunodeficiency virus type 1. *AIDS Res. Hum. Retroviruses* **17,** 1757–1765 (2001).

75. Trkola, A. *et al.* Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J. Virol.* **70,** 1100–1108 (1996).

76. Binley, J. M. *et al.* Comprehensive cross-clade neutralization analysis of a panel of anti-human immunodeficiency virus type 1 monoclonal antibodies. *J. Virol.* **78,** 13232–13252 (2004).

77. Saphire, E. O. *et al.* Crystal structure of a neutralizing human IGG against HIV-1: a template for vaccine design. *Science* **293,** 1155–1159 (2001).

78. Ofek, G. *et al.* Structure and mechanistic analysis of the anti-human immunodeficiency virus type 1 antibody 2F5 in complex with its gp41 epitope. *J. Virol.* **78,** 10724–10737 (2004).

79. Scanlan, C. N. *et al.* The broadly neutralizing anti-human immunodeficiency virus type 1 antibody 2G12 recognizes a cluster of α1→2 mannose residues on the outer face of gp120. *J. Virol.* **76,** 7306–7321 (2002).

80. Lee, H. K. *et al.* Reactivity-based one-pot synthesis of oligomannoses: defining antigens recognized by 2G12, a broadly neutralizing anti-HIV-1 antibody. *Angew. Chem. Int. Ed. Engl.* **43,** 1000–1003 (2004).

81. Calarese, D. A. *et al.* Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science* **300,** 2065–2071 (2003).

82. Dudkin, V. Y. *et al.* Toward fully synthetic carbohydrate-based HIV antigen design: on the critical role of bivalency. *J. Am. Chem. Soc.* **126,** 9560–9562 (2004).

83. Geng, X., Dudkin, V. Y., Mandal, M. & Danishefsky, S. J. In pursuit of carbohydrate-based HIV vaccines, part 2: the total synthesis of high-mannose-type gp120 fragments — evaluation of strategies directed to maximal convergence. *Angew. Chem. Int. Ed. Engl.* **43,** 2562–2565 (2004).

84. Adams, E. W. *et al.* Oligosaccharide and glycoprotein microarrays as tools in HIV glycobiology; glycan-dependent gp120/protein interactions. *Chem. Biol.* **11,** 875–881 (2004).

85. Wang, L. X., Ni, J., Singh, S. & Li, H. Binding of high-mannose-type oligosaccharides and synthetic oligomannose clusters to human antibody 2G12: implications for HIV-1 vaccine design. *Chem. Biol.* **11,** 127–134 (2004).

86. Ni, J., Song, H., Wang, Y., Stamatos, N. M. & Wang, L. X. Toward a carbohydrate-based HIV-1 vaccine: synthesis and immunological studies of oligomannose-containing glycoconjugates. *Bioconjug. Chem.* **17,** 493–500 (2006).

87. Wang, L. X. Toward oligosaccharide- and glycopeptide-based HIV vaccines. *Curr. Opin. Drug Discov. Devel.* **9,** 194–206 (2006).

88. Roux, K. H. *et al.* Electron microscopic and immunochemical analysis of the broadly neutralizing HIV-1-specific, anti-carbohydrate antibody, 2G12. *Mol. Immunol.* **41,** 1001–1011 (2004).

89. Freire, T., Bay, S., Vichier-Guerre, S., Lo-Man, R. & Leclerc, C. Carbohydrate antigens: synthesis aspects and immunological applications in cancer. *Mini Rev. Med. Chem.* **6,** 1357–1373 (2006).

90. Kagan, E. *et al.* Comparison of antigen constructs and carrier molecules for augmenting the immunogenicity of the monosaccharide epithelial cancer antigen Tn. *Cancer Immunol. Immunother.* **54,** 424–430 (2005).

91. Riemer, A. B. *et al.* Induction of IgG antibodies against the GD2 carbohydrate tumor antigen by vaccination with peptide mimotopes. *Eur. J. Immunol.* **36,** 1267–1274 (2006).

92. Pashov, A. *et al.* Antigenic properties of peptide mimotopes of HIV-1-associated carbohydrate antigens. *J. Biol. Chem.* **280,** 28959–28965 (2005).

93. Slovin, S. F. *et al.* Fully synthetic carbohydrate-based vaccines in biochemically relapsed prostate cancer: clinical trial results with α-*N*-acetylgalactosamine-*O*-serine/threonine conjugate vaccine. *J. Clin. Oncol.* **21,** 4292–4298 (2003).

94. Vliegenthart, J. F. Carbohydrate based vaccines. *FEBS Lett.* **580,** 2945–2950 (2006).

95. Komagamine, T. & Yuki, N. Ganglioside mimicry as a cause of Guillain–Barré syndrome. *CNS Neurol. Disord. Drug Targets* **5,** 391–400 (2006).

96. Gemmill, T. R. & Trimble, R. B. Overview of N- and O-linked oligosaccharide structures found in various yeast species. *Biochim. Biophys. Acta* **1426,** 227–237 (1999).

97. Kwong, P. D. *et al.* Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure* **8,** 1329–1339 (2000).

98. Pantophlet, R. *et al.* Fine mapping of the interaction of neutralizing and nonneutralizing monoclonal antibodies with the CD4 binding site of human immunodeficiency virus type 1 gp120. *J. Virol.* **77,** 642–658 (2003).

99. Calarese, D. A. *et al.* Dissection of the carbohydrate specificity of the broadly neutralizing anti-HIV-1 antibody 2G12. *Proc. Natl Acad. Sci. USA* **102,** 13372–13377 (2005).

100. Willison, H. J. The immunobiology of Guillain–Barré syndromes. *J. Peripher. Nerv. Syst.* **10,** 94–112 (2005).

Synthesis and medical applications of oligosaccharides

Peter H. Seeberger[1] & Daniel B. Werz[1,2]

Our understanding of the different glycoconjugates present on cells, proteins and entire organisms is lagging far behind advances in genomics and proteomics. Carbohydrate sequencing and the synthesis of defined oligosaccharides are two key technologies that have contributed to progress in glycomics research. Synthetic tools and high-throughput experiments such as carbohydrate arrays are beginning to affect biological research. These techniques are now being applied to the development of carbohydrate-based diagnostics, vaccines and therapeutics.

Glycomics is the study of the structural and functional aspects of the various glycoconjugates present on proteins, cells and, in some cases, entire organisms. Compared with its counterparts, genomics[1] and proteomics[2] — which deal with nucleic acids and proteins, respectively — the field of glycomics is much less developed. The types of bio-oligomers and biopolymers covered by the term glycoconjugate are diverse. Different natural products such as glycoproteins, glycolipids, glycosaminoglycans and glycosylphosphatidylinositol anchors are summarily known as sugars[3]. Even when only the oligosaccharide chains are considered, studies of processes involving carbohydrates are complicated because, unlike peptides and oligonucleotides (which are usually linear), many oligosaccharides have branched structures. Because they are not under direct genetic control, glycoconjugates are typically heterogeneous. Amplification methods — such as the polymerase chain reaction (PCR) for nucleic acids or bacterial expression systems for protein production — do not exist for glycoconjugates. Consequently, until recently, isolation of carbohydrates was the only way to procure these natural products[3]. Overall progress in glycobiology has suffered from a lack of tools such as those that are readily available for studying nucleic acids and proteins, including automated sequencing[4,5], automated synthesis[6,7], high-throughput microarray screening, and detailed structure elucidation, including X-ray analyses. Carbohydrate synthesis[8] methods are time-consuming and practised mostly by specialized laboratories[9–11]. The products of such syntheses have aided the development of modern sequencing methods[12,13].

Improved synthetic protocols for solution-phase oligosaccharide synthesis[9–11], as well as the use of automated carbohydrate assembly, have provided more straightforward access to usable quantities of pure oligosaccharides. Synthetic carbohydrates have allowed the development of chemical approaches to glycomics that provide a molecular picture of biological processes involving carbohydrates. Synthetic sugars are beginning to be used in the development of diagnostic tests, vaccines and carbohydrate therapeutics. Some of these exciting advances in this rapidly growing field are highlighted in this review.

The usual first step when investigating a biological signal-transduction event is to establish which molecule is responsible for the activity. If the biomolecule is a nucleic acid or a protein, an answer can be obtained quickly because there are reliable, automated sequencing techniques. If, however, a carbohydrate is involved, sequencing is less straightforward. Carbohydrate analysis has improved tremendously in the past two decades[12–18], but there is still no single method to determine the composition of all glycoconjugates. Given the structural diversity of glycoconjugates, different analytical approaches may persist for the analysis of different classes of sugar.

Synthesis of carbohydrates

Once a particular oligosaccharide (or a set of oligosaccharides) has been identified as being responsible for a biological effect, it often has to be synthesized in order to establish or confirm its structure assignment. In addition, defined oligosaccharides and their analogues are key tools for biochemical, biophysical and biological studies. The synthesis of carbohydrates has been pursued for more than a century, and many oligosaccharides can now be synthesized, albeit with considerable effort[6,9–11]. Specialized laboratories synthesize oligosaccharides using processes that may take months to years owing to the structural complexity of carbohydrates. This situation is reminiscent of the solution-phase total syntheses of peptides and DNA that were practised before the advent of automated solid-phase synthesis. A host of improvements has accelerated solution-phase oligosaccharide syntheses[19].

The one-pot synthesis strategy aims to automate the planning of oligosaccharide synthesis[20]. On the basis of the relative reactivity of hundreds of monosaccharide 'building blocks', a computer program known as Optimer selects the appropriate building blocks, as well as the order in which they should be added to the reaction vessel during oligosaccharide assembly (Fig. 1a). The Optimer method works well for oligosaccharides of up to six units, but requires an extensive set of building blocks. An automated instrument known as the 'Golgi apparatus'[21] has been used for enzymatic syntheses[22], building on the superb regiospecificity and stereospecificity of glycosyltransferases that avoid the need for protective groups to assemble oligosaccharides[23] (see page 1008). Engineered organisms such as yeast have been used for the production of N-glycosylated proteins[24,25]. A library of glycoengineered cell lines is expected to yield a plethora of specific glycovariants that have previously been unobtainable in mammalian cells.

The development of a fully automated oligosaccharide synthesis process by chemical means has been viewed with considerable scepticism in light of the complexity of carbohydrates, the large number of monomers needed and possible connections between monosaccharide units. Recent bioinformatics studies explored the diversity of mammalian oligosaccharide connectivities using a comprehensive database of

[1]Laboratory for Organic Chemistry, Swiss Federal Institute of Technology (ETH) Zürich, Wolfgang-Pauli-Strasse 10, CH-8093 Zürich, Switzerland. [2]Present address: Institute for Organic and Biomolecular Chemistry, University of Göttingen, Tammannstrasse 2, D-37077 Göttingen, Germany.

isolated N-linked and O-linked glycans, and glycosphingolipids (D.B.W., R. Ranzinger, S. Herget, A. Adibekian, C.-W. von der Lieth and P.H.S., unpublished observations). Data mining has revealed that nature uses only a small proportion of the theoretically possible connections, so the complexity of glycospace (that is, the body of different structures that can, in principle, be constructed) is significantly reduced. According to the results of this database analysis (D.B.W., R. Ranzinger, S. Herget, A. Adibekian, C.-W. von der Lieth and P.H.S., unpublished observations), just 36 building blocks are needed to assemble 75% of known mammalian oligosaccharides by chemical methods. The challenge of synthesizing these 36 building blocks is surmountable when considering that about 100 different amino-acid monomers are commercially available for peptide synthesis.

Each monosaccharide building block is synthesized at a multi-gram scale and can be used for the assembly of various targets. The feasibility of assembling most carbohydrates using a limited, defined set of building blocks is still questioned, but example structures of increasing complexity are being reported. Temporary protective groups mark sites of further glycosylation, and permanent protective groups mask hydroxyl groups to be unveiled at the end of the synthesis. Besides controlling regioselectivity by orthogonal protective groups to account for branching of the carbohydrate chain, the stereochemistry at the anomeric carbon must be controlled. Placement of participating protective groups at the C2 hydroxyl or amine groups ensures the formation of *trans*-glycosidic linkages, whereas non-participating groups are needed for the preferential installation of *cis*-glycosides.

Automated solid-phase oligosaccharide synthesis (Fig. 1b) has been developed from insights gained from oligopeptide and oligonucleotide assembly[26]. The first building block is added to a polystyrene resin equipped with an easily cleavable linker containing a free hydroxyl group[27]. An activating agent induces couplings involving glycosyl phosphate and glycosyl trichloroacetimidate building blocks[26]. Unlike oligonucleotide and peptide couplings, glycosidic bond formation occurs mostly at low temperatures and requires a reaction chamber that can be cooled. Excess building blocks (that is, a 5–10-fold molar excess, sometimes applied twice) are added to the chamber for each coupling. Mass action to drive coupling reactions to completion and to achieve high yields is also common to peptide and oligonucleotide syntheses. Washing and filtration remove any side products or remaining reagents before selective removal of a temporary protective group readies the next hydroxyl group for subsequent coupling. Coupling efficiencies can be assessed by spectrometric read-out after protecting-group removal when temporary protecting groups that absorb ultraviolet radiation, such as 9-fluorenylmethyloxycarbonyl (Fmoc), are used[28]. Originally, this coupling–deprotection cycle was automated using a converted peptide synthesizer[26]. An automated oligosaccharide synthesizer prototype with parallel synthesis capability is currently being tested.

After completion of the oligosaccharide sequence, the fully protected product is cleaved from solid support. After global deprotection, the oligosaccharide is purified and its structure verified. A series of increasingly complex oligosaccharides has been assembled, each within 1 day or less, using the automated oligosaccharide synthesizer. This compares favourably with the weeks to months taken using solution-phase methods[28]. Initial syntheses contained only the synthetically less challenging *trans*-glycosidic linkages, but *cis*-glycosides such as α-galactoses have now also been selectively incorporated[29].

At present, automated oligosaccharide synthesis resembles the early days of automated peptide and oligonucleotide assembly: many carbohydrate structures, both simple and complex, can be synthesized by automation. The problems and drawbacks — such as the excess of buildings blocks used, the difficulties in incorporating certain monosaccharides such as sialic acid, and the double bond in the linking moiety that restricts the deprotection conditions — have been recognized. Some of these limitations have now been addressed, but certain challenges remain. Although commercially available monomeric building blocks are quite expensive, it seems likely that the cost of these reagents will decrease with increasing demand. One of the reasons why automated
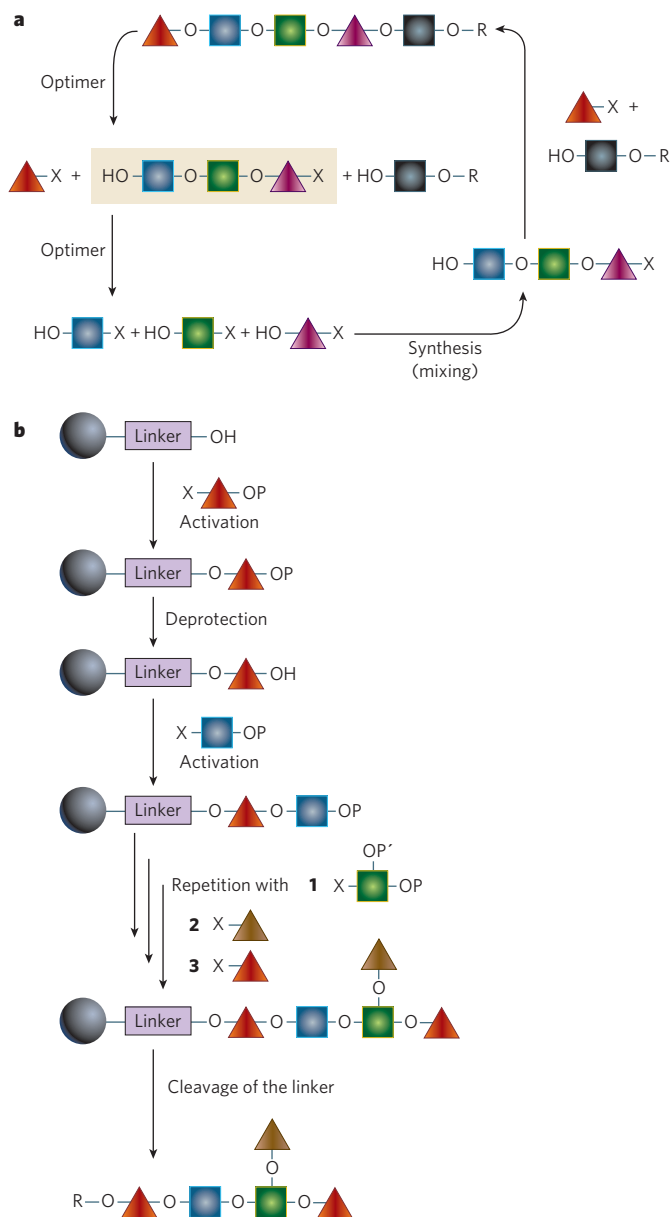
Figure 1 | Schematic representation of strategies for automated oligosaccharide assembly. a, Automated solution-phase synthesis using the Optimer-based one-pot approach for a pentasaccharide. A computer program selects appropriate monosaccharide building blocks according to their relative reactivity values in order to achieve best yields for oligosaccharide assembly. **b**, Automated solid-phase assembly using five monosaccharide building blocks on a polymer resin (black circle), attached by means of a linker. The cycle — consisting of activation and deprotection steps — is performed five times for the assembly of a pentasaccharide. Finally, the linker is cleaved to procure the desired oligosaccharide. Coloured triangles and squares represent different sugar monomer building blocks. P, P′, temporary protecting groups; R, hydrocarbon residue to be functionalized; X, leaving group.

solid-phase synthesis of oligosaccharides is not currently performed by a larger number of groups is the fact that the synthesis instrument is not yet commercially available. However, the strength of the chemical approach to incorporating unnatural linkages nicely complements the evolving enzymatic technologies. A combination of improved isolation and purification strategies of naturally occurring carbohydrates, enzymatic approaches, the use of engineered organisms and chemical

Table 1 | Synthetic tools for studying the interactions of carbohydrates

Carbohydrates attached to	Source of oligosaccharides	Use	Refs
Gold surface for surface plasmon resonance	Isolated and synthetic	Accurate measurement of carbohydrate–protein interactions	87
Gold beads	Isolated and synthetic	Visualization of carbohydrate–protein and carbohydrate–cell interactions	88
Affinity columns	Synthetic	Purification of carbohydrate-binding proteins	89
Magnetic beads	Synthetic	Purification of carbohydrate-binding proteins	90, 91
Activity-based probes	Synthetic	Elucidation of enzyme mechanisms	92
Fluorescent labelling	Isolated and synthetic	Visualization of carbohydrate location *in vitro* and *in vivo*	89
Quantum dots	Synthetic	*In vitro* and *in vivo* imaging	91
Fluorescent polymers	Synthetic	Detection of bacteria	93

synthetic approaches such as automated solid-phase synthesis is expected to provide scientists with rapid access to defined oligosaccharide libraries for glycomics investigations in the near future.

Carbohydrate arrays

Several methods have been established to study the interactions of carbohydrates with various other molecules (Table 1). The DNA microarray has been a key tool in genomics research[30], and protein arrays are used to identify protein–protein interactions and potential inhibitors[31]. Similarly, carbohydrate microarrays have been used in glycomics research to examine the interactions of carbohydrates with other molecules. The chip-based format of microarrays offers important advantages over common screening techniques such as enzyme-linked immunosorbent assays (ELISAs), because several thousand binding events can be screened on a single glass slide and only minuscule amounts of analyte and ligand are required. Assay miniaturization is particularly suitable for glycomics, because access to pure oligosaccharides is the limiting factor. The first carbohydrate microarrays relied on isolated saccharides that were non-covalently attached to membranes[32,33]. A flurry of methodological studies evaluated different aspects of microarray design and adopted oligonucleotide and protein array techniques for carbohydrate chips (Fig. 2). Synthetic monosaccharides and oligosaccharides were covalently attached via different linkers to glass[34], plastic[35] and gold surfaces[36], or placed on beads in fibre-optic wells[37]. Initial proof-of-principle studies focused on known interactions between lectins (carbohydrate-binding proteins) and sugars. Current screening efforts rely on carbohydrate arrays in which chemically or enzymatically synthesized and isolated oligosaccharides with a linker on the reducing terminus are covalently attached to glass slides[38]. Standard DNA printing and scanning equipment is used to produce and analyse the carbohydrate microarrays[39,40].

Initially, carbohydrate–protein interactions important to the process of HIV infection were analysed, and epitope mapping of HIV-related antibodies was performed[40,41]. The knowledge gained from the microarray experiments was essential to efforts directed at the creation of carbohydrate-based HIV vaccine candidates[42,43] (see page 1038).

During the past 3 years, a host of carbohydrate–protein interactions has been studied using oligosaccharide arrays[44]. National and regional consortia such as the Consortium for Functional Glycomics (NIH, USA) and the ETH Zürich Glycomics Initiative are making this technology widely accessible to life science researchers.

Carbohydrate microarrays can be used to address the interactions of sugars with other types of molecule, as well as entire cells. Carbohydrate–RNA interactions were screened by incubating labelled RNA with aminoglycoside microarrrays[45]. Mechanisms responsible for antibiotic resistance were studied using these arrays together with resistance-causing enzymes[46].

The binding of cells to microarray surfaces allows the detection and typing of bacteria in blood[47]. The interaction of eukaryotic cells with carbohydrate arrays has also been demonstrated[48]. The oligosaccharide binding preferences of different types of avian and human influenza strains can be determined using carbohydrate microarrays[49].

The structure–activity relationship of glycosaminoglycan polysaccharides, including heparin and chondroitin sulphate, is poorly understood. Carbohydrate microarrays containing synthetic[50] or isolated[51] heparin oligosaccharides have served to identify specific sequences recognized by different fibroblast growth factors[52]. Immobilization of a series of chondroitin tetrasaccharides has aided the identification of a novel tumour necrosis factor-α antagonist[53] and the exact chondroitin sequences bound by proteins[54].

In addition to the application of carbohydrate arrays to the study of biomolecular interactions, oligosaccharide microarrays are beginning to be used as diagnostics and to conduct epidemiological studies. Hundreds of human sera have been screened for antibodies against the malaria toxin glycosylphosphatidylinositol (GPI) anchor[55], and a correlation between the presence of specific antibodies and resistance to severe malaria has been established (F. Kamena, M. Tamborrini, X. Liu, G. Pluschke and P.H.S., unpublished observations). The search for serological markers of autoimmune diseases has yielded results for Crohn's disease[56]. Epitope mapping of antibodies against tumour-associated antigens is also feasible[57].

The presence of certain oligosaccharide structures in a glycoprotein sample can be determined using arrays of immobilized lectins[58]. These carbohydrate-binding proteins recognize terminal saccharide units and have been used to analyse the dynamic bacterial glycome[59].

Oligosaccharide therapeutics

Given the prevalent role of carbohydrates in a wide range of biological processes it may seem surprising that there are few carbohydrate-based therapeutics and diagnostics on the market. In addition to monosaccharide-inspired drugs such as the influenza virus treatment Tamiflu[60,61] (oseltamivir phosphate; Roche) two blockbuster drugs, acarbose (Precose, Glucobay; Bayer) and heparin, stand out. Both oligosaccharides were derived by isolation and reached the clinic before a detailed structure–activity relationship had been established. In addition, aminoglycosides — naturally occurring pseudo-oligosaccharides — have
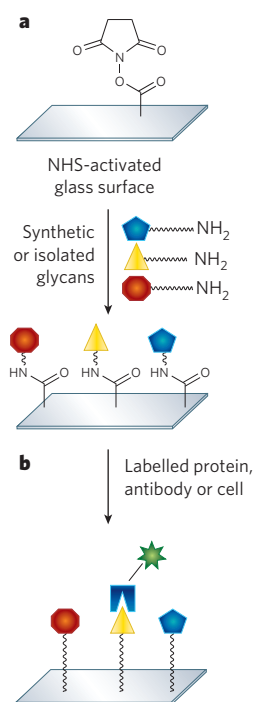
Figure 2 | Carbohydrate microarrays. a, Carbohydrate microarrays can be constructed from synthetic or isolated oligosaccharides that contain a primary amine. The glycans are covalently attached to the glass surface of a microscope slide that has been functionalized with a reactive group (for example, *N*-hydroxysuccinimide, NHS). **b,** Incubation of the carbohydrate microarray with a protein, antibody or cell that has been labelled (for example, with a fluorescent group) can be used to determine which oligosaccharide binds to that protein, antibody or cell. In this particular example, the protein represented by a blue wedge specifically binds to the oligosaccharide represented by the yellow triangle, but not to the other oligosaccharides that appear on the glass slide.

been used clinically to treat infectious diseases induced by a variety of Gram-negative bacteria[62]. The antibiotic activity of aminoglycosides is due to their inhibition of protein synthesis, which results from their binding to bacterial ribosomes[62].

Heparin

The oldest carbohydrate-based drug is isolated from animal organs and has been used clinically as an antithrombotic agent since the 1940s. Heparin activates the serine protease inhibitor antithrombin III, which blocks thrombin and factor Xa in the coagulation cascade[63]. This drug is a highly heterogeneous mixture of polysaccharides and is associated with severe side effects, including heparin-induced thrombocytopenia, bleeding and allergic reactions. Chemically or enzymatically fragmented heparins (low-molecular-weight heparins, LMWHs) are also heterogeneous, but are more bioavailable, with a longer half-life, a more predictable anticoagulant activity and fewer side effects *in vivo*.

After the specific pentasaccharide responsible for the anticoagulant property was identified in the early 1980s (ref. 64; Fig. 3a), a herculean effort lasting more than 10 years was begun to establish a structure–function relationship using synthetic oligosaccharides[64]. As a result of this drug-development effort, a synthetic pentasaccharide known as Arixtra (fondaparinux sodium; GlaxoSmithKline) has been available since 2002 (ref. 65). However, Arixtra does have some clinical shortcomings, such as an exceedingly long half-life *in vivo* and little to no dose-dependent activity in certain indications[66]. Thus, LMWHs still have the highest market share of all antithrombotics, and the need for additional synthetic heparin molecules with specific activities persists.

Recent advances in heparin sequencing[16], heparin synthesis[67–70] and heparin microarray technology[50,51] have provided the tools to identify specific sequences or sequence families that interact with proteins such as chemokines. The chemical synthesis of a broad range of heparin analogues should allow researchers to study the molecular mechanism of angiogenesis and to modulate wound healing and other medically relevant processes (see page 1030).

Acarbose

Carbohydrates such as starch and sucrose are principal components of food, and have to be enzymatically broken down in the intestinal tract. Acarbose[71], a pseudo-oligosaccharide of microbial origin, is produced by fermentation. This α-glucosidase and α-amylase inhibitor interferes with and regulates intestinal carbohydrate digestion, controls the rate of absorption of monosaccharides and influences the intermediary carbohydrate metabolism. It is used to treat type 2 diabetes.

Carbohydrate-based vaccines

The cell surfaces of bacteria, parasites and viruses exhibit oligosaccharides that are often distinct from those of their hosts. Specific types of glycoconjugate are often more highly expressed on the surface of tumours than on normal cells[72]. Such cell-surface carbohydrate markers are the basis for carbohydrate-based detection systems and vaccines. An immune response against the carbohydrate antigens that results in the death of target cells is required for a carbohydrate-based vaccine. Such vaccines have been widely used against a host of diseases for several decades[73]. The carbohydrate antigens for antibacterial vaccines were isolated from biological sources. Recently, intense efforts focused on the use of defined carbohydrate antigens that are synthesized rather than isolated. A carbohydrate-based approach has also been pursued for anticancer vaccine candidates[74–76] (see page 1000). However, one of the early carbohydrate-based anticancer vaccine candidates recently failed in a Phase III clinical trial.

Antibacterial vaccines

Polysaccharide capsules, glycoproteins or glycolipids cover the cell surfaces of many bacteria. Capsular polysaccharides are either homopolymers or made up of between two and six repeating sugar units. Capsular polysaccharides elicit type-specific protective immune responses in adults but not in infants, who do not respond with antibodies that confer
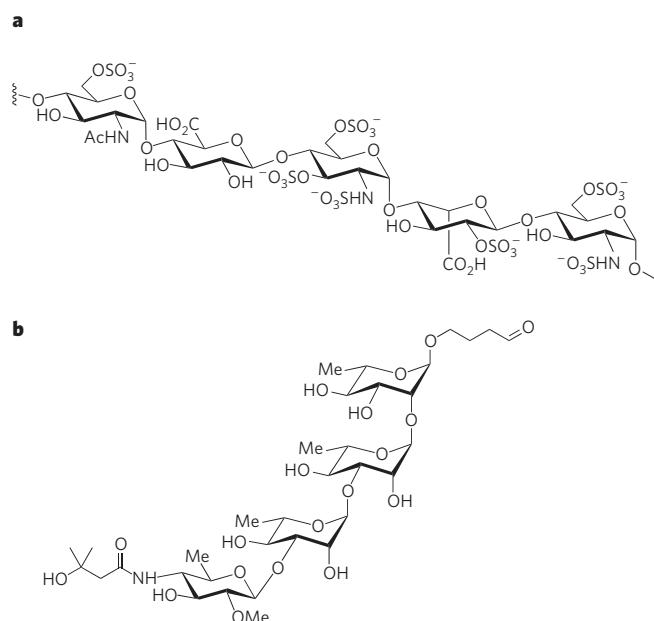
Figure 3 | Carbohydrates used in medicine and for vaccine development. a, A pentasaccharide sequence of heparin. This sequence is responsible for binding to antithrombin III. **b**, Synthetic spore surface tetrasaccharide of *B. anthracis*. This molecule was used for the generation of anticarbohydrate antibodies to detect anthrax spores and is currently being used in vaccine development.

protection. Conjugate vaccines consisting of a carbohydrate antigen and an immunogenic protein can overcome this immunogenicity problem and produce high titres of protective antibodies.

Improved analytical tools have helped to identify the exact chemical structure of carbohydrate antigens and have aided the development of new vaccines. Several vaccines based on purified capsular polysaccharides or on neoglycoconjugates are now commercially available, such as vaccines against *Neisseria meningitidis*, *Streptococcus pneumoniae*, *Haemophilus influenza* type b (Hib) and *Salmonella typhi*[77]. Meningitis caused by Hib has essentially been eradicated in areas where national vaccination programmes using protein conjugate vaccines have been implemented.

Vaccine development could benefit greatly from the new glycomics technologies. The identification of specific oligosaccharide antigens has been aided substantially by sequencing and carbohydrate arrays. The procurement of defined oligosaccharides using improved solution- and solid-phase methods has become fast enough to be used reiteratively in drug-development efforts.

A synthetic oligosaccharide-based conjugate vaccine is now used in Cuba, where the large-scale synthesis, pharmaceutical development and clinical evaluation of a conjugate vaccine composed of a synthetic capsular polysaccharide antigen of Hib was achieved. Long-term protective antibody titres compared favourably with products prepared with the Hib polysaccharide extracted from bacteria[78].

A tetrasaccharide has been discovered on the surface of spores of the biowarfare agent *Bacillus anthracis*[79]. Once the durable form of the pathogen has been inhaled it will kill most victims if treatment is not commenced immediately. Synthesis of a species-specific tetrasaccharide antigen[80,81] (Fig. 3b) allowed the production of antibodies that specifically recognize *B. anthracis* in the presence of the closely related opportunistic human pathogen *Bacillus cereus*[82]. Challenge experiments to create a conjugate vaccine against anthrax are ongoing.

Antiparasite vaccines

Like bacteria, many parasites have unique glycoconjugates on their surfaces. The specific carbohydrates may serve as a starting point for the creation of conjugate vaccines, but efforts towards this goal have been

hampered by the fact that the parasites are very difficult to culture and because glycoconjugates cannot be obtained in pure form or sufficient quantity by isolation.

Malaria

Plasmodium falciparum is the most pathogenic of the single-celled parasites of the genus *Plasmodium* that are responsible for malaria. Malaria infects 5–10% of humans worldwide and kills more than 2 million people each year. Infected mosquitoes transmit the parasite, which leads to the common symptoms of chills and fever. Drug resistance is a growing problem at a time when there is still no effective vaccine.

P. falciparum expresses a large amount of GPI on its cell surface[83]. This glycolipid triggers an inflammatory cascade that is responsible for much of malaria's morbidity and mortality. When a protein conjugate of a synthetic hexasaccharide GPI malaria toxin was administered to mice before infection, this resulted in a highly reduced mortality rate of only 10–20%, compared with 100% without vaccination[55]. Cross-reactivity of the antibodies with human GPI structures was not observed owing to the differences between human and *P. falciparum* GPI. Immunization of mice did not alter the infection rate or overall parasitaemia, indicating that the antibody against the GPI neutralized toxicity without killing the parasites[55].

Preclinical studies involving protein conjugates of synthetic GPI antigens are currently underway. To support such vaccine development efforts, methods for the large-scale synthesis of oligosaccharide antigens have been developed by taking advantage of the latest advances in carbohydrate synthesis technology. Very small amounts of synthetic antigen (10^{-9}–10^{-7} g per person) are required, and the production of several kilograms of antigen per year will suffice.

Leishmaniasis

Leishmaniasis, which is caused by another protozoan parasite, is transmitted by sandflies and affects more than 12 million people worldwide. *Leishmania* resides in macrophages, making them difficult to treat. In a search for a potent vaccine, the lipophosphoglycans (LPGs)[84] that are ubiquitous on the cell surfaces of the parasites and are composed of a GPI anchor, a repeating phosphorylated disaccharide and different cap oligosaccharides became a target. The cap tetrasaccharide has been the focus of efforts towards a conjugate vaccine based on a synthetic antigen. The branched tetrasaccharide was assembled by automated solid-phase synthesis[85] and conjugated to a virosomal particle to enhance immunogenicity. These highly immunogenic conjugates yielded antibodies that selectively recognize parasite-infected livers[86]. Challenge studies in an animal model are currently underway.

Recent advances and future development

For many years the lack of tools for studying glycobiology prevented biologists and medical researchers from addressing research problems that involve carbohydrates. During the past decade, sequencing and synthesis technologies that are commonly used to study nucleic acids and proteins have become available for glycomics as well. Now, carbohydrate sequencing of glycoconjugates is often possible even though sample preparation is complicated by carbohydrate microheterogeneity and the absence of amplification procedures. Automated solid-phase synthesis, improved methods for solution-phase oligosaccharide assembly, enzymatic methods and the use of engineered cells have complemented each other, allowing oligosaccharide synthesis to take a big step forward by granting access to different classes of glycoconjugate. In turn, these methods have helped procure oligosaccharides and their non-natural analogues for the creation of high-throughput screening methods such as carbohydrate arrays.

The identification of specific oligosaccharides, by sequencing followed by comparison with synthetic oligosaccharides, has yielded insight into the interactions of carbohydrates and proteins. Oligosaccharide involvement at key positions of signalling pathways is beginning to emerge and a molecular understanding of carbohydrate binding to proteins is evolving. Detailed structural studies — including studies of protein–carbohydrate

interactions — using X-ray crystallography will become commonplace in the near future. Further improvements in the methods by which oligosaccharides are sequenced and synthesized will be needed to make their routine use possible for non-specialists.

A better understanding of the biological roles of carbohydrates and improved sequencing and synthesis techniques are beginning to influence the design of diagnostic and therapeutic approaches. Carbohydrate arrays help to define new disease markers by screening the sera of patients. Bacterial and viral detection and typing can be achieved using carbohydrate microarrays. Synthetic access to oligosaccharides of infectious agents that are hard to culture and isolate (for example, *B. anthracis* and *P. falciparum*) facilitates antibody production for specific detection of these pathogens. These anticarbohydrate antibodies may become important for passive immunization. The first conjugate vaccine candidates containing synthetic oligosaccharide antigens are reaching preclinical and clinical trials against bacterial (for example, Hib), viral (for instance, HIV) and parasitic (for example, malaria and leishmaniasis) infections. The trend to produce defined vaccine antigens using chemical and enzymatic methods, as well as engineered cells, is likely to increase, and synthetic vaccines are expected to complement already existing vaccines containing purified polysaccharides.

As our understanding of carbohydrate involvement in signalling cascades — in particular of those that involve glycosaminoglycans — expands, carbohydrate-mediated processes will become the target of drug-development efforts using small organic molecules. Glycomics has just gone beyond the initial proof-of-principle studies for diagnostics and therapeutic candidates. Improved tools and a better molecular understanding should convince those biologists and medical researchers who previously avoided carbohydrates to address questions involving this class of molecule. The excitement of glycomics is just beginning, with many discoveries to be made and applications to be developed. ∎

1. Insight: Human genomics and medicine. *Nature* **429,** 439–481 (2004).
2. Insight: Proteomics. *Nature* **422,** 191–237 (2003).
3. Varki, A. *et al. Essentials of Glycobiology* (Cold Spring Harbor Laboratory Press, 1999).
4. Hunkapiller, T., Kaiser, R. J., Koop, B. F. & Hood, L. Large-scale and automated DNA sequence determination. *Science* **354,** 59–67 (1991).
5. Grant, G. A., Crankshaw, M. W. & Gorka, J. Edman sequencing as tool for characterization of synthetic peptides. *Methods Enzymol.* **289,** 395–419 (1997).
6. Caruthers, M. H. Gene synthesis machines: DNA chemistry and its uses. *Science* **230,** 281–285 (1985).
7. Merrifield, B. Concept and early development of solid-phase peptide synthesis. *Methods Enzymol.* **289,** 3–13 (1997).
8. Ernst, B., Hart, G. W. & Sinaÿ, P. *Carbohydrates in Chemistry and Biology Pt 1 Chemistry of Saccharides* (Wiley-VCH, Weinheim, 2000).
9. Mayer, T. G., Kratzer, B. & Schmidt, R. R. Synthesis of a GPI anchor of the yeast *Saccharomyces cerevisiae*. *Angew. Chem. Int. Ed. Engl.* **33,** 2177–2181 (1994).
10. Seifert, J., Lergenmuller, M. & Ito, Y. Synthesis of an α-(2,3)-sialylated complex-type undecasaccharide. *Angew. Chem. Int. Ed.* **39,** 531–534 (2000).
11. Wang, Z.-G. *et al.* Toward fully synthetic homogeneous glycoproteins: a high mannose core containing glycopeptide containing carrying full H-type 2 human blood group specifity. *Angew. Chem. Int. Ed.* **40,** 1728–1732 (2001).
12. Ashline, D., Singh, S., Hanneman, A. & Reinhold, V. Congruent strategies for carbohydrate sequencing. 1. Mining structural details by MSn. *Anal. Chem.* **77,** 6250–6262 (2005).
13. Laroy, W., Contreras, R. & Callewaert, N. Glycome mapping on DNA sequencing equipment. *Nature Protocols* **1,** 397–405 (2006).
14. Dell, A. & Morris, H. R. Glycoprotein structure determination by mass spectrometry. *Science* **291,** 2351–2356 (2001).
15. Venkataraman, G., Shriver, Z., Raman, R. & Sasisekharan, R. Sequencing complex polysaccharides. *Science* **286,** 537–542 (1999).
16. Chi, L. *et al.* Preparation and structural determination of large oligosaccharides derived from acharan sulfate. *Carbohydr. Res.* **341,** 864–869 (2006).
17. Saad, O. M. & Leary, J. A. Heparin sequencing using enzymatic digestion and ESI-MSn with HOST: a heparin/HS oligosaccharide sequencing tool. *Anal. Chem.* **77,** 5902–5911 (2005).
18. Nishimura, S. *et al.* High-throughput protein glycomics: combined use of chemoselective glycoblotting and MALDI-TOF/TOF mass spectrometry. *Angew. Chem. Int. Ed.* **44,** 91–96 (2004).
19. Sears, P. & Wong, C.-H. Toward automated synthesis of oligosaccharides and glycoproteins. *Science* **291,** 2344–2350 (2001).
20. Zhang, Z. *et al.* Programmable one-pot oligosaccharide synthesis. *J. Am. Chem. Soc.* **121,** 734–753 (1999).
21. Nishimura, S. Automated glycosynthesizer 'Golgi' by mimicking biosynthetic process. *Tanpakushitsu Kakusan Koso* **48,** 1220–1225 (2003).
22. Koeller, K. M. & Wong, C.-H. Enzymes for chemical synthesis. *Nature* **409,** 232–234 (2001).
23. Wong, C.-H., Halcomb, R. L., Ichikawa, Y. & Kajimoto, T. Enzymes in organic synthesis: application to the problems of carbohydrate recognition. *Angew. Chem. Int. Ed. Engl.* **34,** 412–432 (1995).

24. Gerngross, T. U. Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. *Nature Biotechnol.* **22,** 1409–1414 (2004).

25. Li, H. *et al.* Optimization of humanized IgGs in glycoengineered *Pichia pastoris. Nature Biotechnol.* **24,** 210–215 (2006).

26. Plante, O. J., Palmacci, E. R. & Seeberger, P. H. Automated solid-phase synthesis of oligosaccharides. *Science* **291,** 1523–1527 (2001).

27. Andrade, R. B., Plante, O. J., Melean, L. G. & Seeberger, P. H. Solid-phase oligosaccharide synthesis: preparation of complex structures using a novel linker and different glycosylating agents. *Org. Lett.* **1,** 1811–1814 (1999).

28. Love, K. R. & Seeberger, P. H. Automated solid-phase synthesis of protected tumor-associated antigen and blood group determinant oligosaccharides. *Angew. Chem. Int. Ed.* **43,** 602–605 (2004).

29. Werz, D. B., Castagner, B. & Seeberger, P. H. Automated synthesis of the tumor-associated carbohydrate antigens Gb-3 and Globo-H: incorporation of α-galactosidic linkages. *J. Am. Chem. Soc.* **129,** 2770–2771 (2007).

30. Lockhart, D. J. & Winzeleer, E. A. Genomics, gene expression and DNA arrays. *Nature* **405,** 827–836 (2000).

31. Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293,** 2101–2105 (2001).

32. Wang, D. *et al.* Carbohydrate microarrays for the recognition of cross-reactive molecular markers of microbes and host cells. *Nature Biotechnol.* **20,** 275–281 (2002).

33. Fukui, S. *et al.* Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate–protein interactions. *Nature Biotechnol.* **20,** 1011–1017 (2002).

34. Park, S. & Shin, I. Fabrication of carbohydrate chips for studying protein–carbohydrate interactions. *Angew. Chem. Int. Ed.* **41,** 3180–3182 (2002).

35. Fazio, F. *et al.* Synthesis of sugar arrays in microtiter plate. *J. Am. Chem. Soc.* **124,** 14397–14402 (2002).

36. Houseman, B. T. & Mrksich, M. Carbohydrate arrays for the evaluation of protein binding and enzymatic modification. *Chem. Biol.* **9,** 443–454 (2002).

37. Adams, E. W. *et al.* Encoded fiber optic microsphere arrays for probing protein–carbohydrate interactions. *Angew. Chem. Int. Ed.* **42,** 5317–5320 (2003).

38. Culf, A. S., Cuperlovic-Culf, M. & Ouellette, R. J. Carbohydrate microarrays: survey of fabrication techniques. *OMICS* **10,** 289–310 (2006).

39. Ratner, D. M. *et al.* Probing protein–carbohydrate interactions with microarrays of synthetic oligosaccharides. *ChemBioChem* **5,** 379–383 (2004).

40. Blixt, O. *et al.* Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proc. Natl Acad. Sci. USA* **101,** 17033–17038 (2004).

41. Adams, E. W. *et al.* Oligosaccharide and glycoprotein microarrays as tools in HIV-glycobiology: glycan dependent gp120/protein interactions. *Chem. Biol.* **11,** 875–881 (2004).

42. Scanlan, C. N. *et al.* The carbohydrate epitope of neutralizing anti HIV-1 antibody 2G12. *Adv. Exp. Med. Biol.* **535,** 205–218 (2003).

43. Rudd, P. M., Wormald, M. R. & Dwek, R. A. Sugar-mediated ligand receptor interactions in the immune system. *Trends Biotechnol.* **22,** 524–530 (2004).

44. Paulson J. C., Blixt, O. & Collins, B. E. Sweet spots in functional glycomics. *Nature Chem. Biol.* **2,** 238–248 (2006).

45. Disney, M. D. & Seeberger, P. H. Aminoglycoside microarrays to explore carbohydrate–RNA interactions. *Chem. Eur. J.* **10,** 3308–3314 (2004).

46. Disney, M. D., Magnet, S., Blanchard, J. S. & Seeberger, P. H. Aminoglycoside microarrays to study antibiotic resistance. *Angew. Chem. Int. Ed.* **43,** 1591–1594 (2004).

47. Disney, M. D. & Seeberger, P. H. The use of carbohydrate microarrays to study carbohydrate–cell interactions and to detect pathogens. *Chem. Biol.* **11,** 1701–1707 (2004).

48. Nimrichter, L. *et al.* Intact cell adhesion to glycan microarrays. *Glycobiology* **14,** 197–203 (2004).

49. Stevens, J. *et al.* Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* **312,** 404–410 (2006).

50. de Paz, J. L., Noti, C. & Seeberger, P. H. Microarrays of synthetic heparin oligosaccharides. *J. Am. Chem. Soc.* **128,** 2766–2767 (2006).

51. de Paz, J. L., Spilmann, D. & Seeberger, P. H. Microarrays of heparin oligosaccharides obtained by nitrous acid depolymerization of heparin. *Chem. Commun.* 3116–3117 (2006).

52. Noti, C., de Paz, J. L. & Seeberger, P. H. Preparation and use of microarrays containing synthetic heparin oligosaccharides for the rapid analysis of heparin–protein interactions. *Chem. Eur. J.* **12,** 8664–8686 (2006).

53. Tully, S. E., Rawat, M. & Hsieh-Wilson, L. C. Discovery of a TNF-α antagonist using chondroitin sulfate microarrays. *J. Am. Chem. Soc.* **128,** 7740–7741 (2006).

54. Gama, C. I. *et al.* Sulfation patterns of glycosaminoglycans encode molecular recognition and activity. *Nature Chem. Biol.* **2,** 467–473 (2006).

55. Schofield, L. *et al.* Synthetic GPI as a candidate anti-toxic vaccine in a model of malaria. *Nature* **418,** 785–789 (2002).

56. Dotan, I. *et al.* Antibodies against laminaribioside and chitobioside are novel serologic markers in Crohn's disease. *Gastroenterology* **131,** 366–378 (2006).

57. Huang, C. Y. *et al.* Carbohydrate microarray for profiling the antibodies interacting with Globo H tumor antigen. *Proc. Natl Acad. Sci. USA* **103,** 15–20 (2006).

58. Pilobello, K. T., Krishnamoorthy, L., Slawek, D. & Mahal, L. K. Development of a lectin microarray for the rapid analysis of protein glycopatterns. *ChemBioChem* **6,** 985–989 (2005).

59. Hsu, K. L., Pilobello, K. T. & Mahal, L. K. Analyzing the dynamic bacterial glycome with a lectin microarray approach. *Nature Chem. Biol.* **2,** 153–157 (2006).

60. De Clercq, E. Antiviral agents against influenza A viruses. *Nature Rev. Drug Discov.* **5,** 1015–1025 (2006).

61. Farina, V. & Brown, J. D. Tamiflu: the supply problem. *Angew. Chem. Int. Ed.* **45,** 7330–7334 (2006).

62. Weizman, H. & Thor, Y. in *Carbohydrate-Based Drug Discovery* (ed. Wong, C.-H.) 661–684 (Wiley-VCH, Weinheim, 2003).

63. Capila, I. & Linhardt, R. J. Heparin–protein interactions. *Angew. Chem. Int. Ed.* **41,** 390–412 (2002).

64. Petitou, M., Casu, B. & Lindahl, U. 1976–1983, a critical period in the history of heparin: the discovery of the antithrombin binding site. *Biochimie* **85,** 83–89 (2003).

65. Petitou, M. & van Boeckel, C. A. A. A synthetic antithrombin binding pentasaccharide is now a drug! What comes next? *Angew. Chem. Int. Ed.* **43,** 3118–3133 (2004).

66. Giangrande, P. L. F. Fondaparinux (Arixtra): a new anticoagulant. *Int. J. Clin. Pract.* **56,** 615–617 (2002).

67. Noti, C. & Seeberger, P. H. Chemical approaches to define structure–activity relationships of heparin-like glycosaminoglycans. *Chem. Biol.* **12,** 731–756 (2005).

68. Codée, J. D. C. *et al.* A modular strategy toward the synthesis of heparin-like oligosaccharides using monomeric building blocks in a sequential glycosylation strategy. *J. Am. Chem. Soc.* **127,** 3767–3773 (2005).

69. de Paz, J.-L. *et al.* The activation of fibroblast growth factors by heparin: synthesis, structure, and biological activity of heparin-like oligosaccharides. *ChemBioChem* **2,** 673–685 (2001).

70. Lee, J.-C. *et al.* Synthesis of heparin oligosaccharides. *J. Am. Chem. Soc.* **126,** 476–477 (2004).

71. Truscheit, E. *et al.* Chemistry and biochemistry of microbial α-glucosidase inhibitors. *Angew. Chem. Int. Ed. Engl.* **20,** 744–761 (1981).

72. Hakomori, S. Tumor-associated carbohydrate antigens defining tumor malignancy: basis for development of anti-cancer vaccines. *Adv. Exp. Med. Biol.* **491,** 369–402 (2001).

73. Goldblatt, D. Recent developments in bacterial conjugate vaccines. *J. Med. Microbiol.* **47,** 563–567 (1998).

74. Danishefsky, S. J. & Allen, J. R. From the laboratory to the clinic: a retrospective on fully synthetic carbohydrate based anticancer vaccines. *Angew. Chem. Int. Ed.* **39,** 836–863 (2000).

75. Zhang, S. *et al.* Selection of tumor antigens as targets for immune attack using immunohistochemistry: I. Focus on gangliosides. *Int. J. Cancer* **3,** 42–49 (1997).

76. Slovin, S. F., Keding, S. J. & Ragupathi, G. Carbohydrate vaccines as immunotherapy for cancer. *Immunol. Cell Biol.* **83,** 418–428 (2005).

77. Ada, G. & Isaacs, D. Carbohydrate–protein conjugate vaccines. *Clin. Microbiol. Infect.* **9,** 79–85 (2003).

78. Verez-Bencomo, V. *et al.* A synthetic conjugate polysaccharide vaccine against *Haemophilus influenzae* type b. *Science* **305,** 522–525 (2004).

79. Daubenspeck, J. M. *et al.* Novel oligosaccharide side chains of the collagen-like region of BclA, the major glycoprotein of the *Bacillus anthracis* exosporium. *J. Biol. Chem.* **279,** 30945–30953 (2004).

80. Werz, D. B. & Seeberger, P. H. Total synthesis of antigen *Bacillus anthracis* tetrasaccharide — creation of an anthrax vaccine candidate. *Angew. Chem. Int. Ed.* **44,** 6315–6318 (2005).

81. Adamo, R., Saksena, R. & Kováč, P. Synthesis of the β-anomer of the spacer-equipped tetrasaccharide side chain of the major glycoprotein of the *Bacillus anthracis* exosporium. *Carbohydr. Res.* **340,** 2579–2582 (2005).

82. Tamborrini, M. *et al.* Anti-carbohydrate antibodies for the detection of anthrax spores. *Angew. Chem. Int. Ed.* **45,** 6581–6582 (2006).

83. Berhe, S., Schofield, L., Schwarz, R. T. & Gerold, P. Conservation of structure among glycosylphosphatidylinositol toxins from different geographic isolates of *Plasmodium falciparum. Mol. Biochem. Parasitol.* **103,** 273–278 (1999).

84. Turco, S. J. & Descoteaux, A. The lipophosphoglycan of *Leishmania* parasites. *Annu. Rev. Microbiol.* **46,** 65–94 (1992).

85. Hewitt, M. C. & Seeberger, P. H. Automated solid-phase synthesis of a branched *Leishmania* cap tetrasaccharide. *Org. Lett.* **3,** 3699–3702 (2001).

86. Liu, X. *et al.* Enhancement of the immunogenicity of synthetic carbohydrates by conjugation to virosomes — a leishmaniasis vaccine candidate. *ACS Chem. Biol.* **1,** 161–164 (2006).

87. Houseman, B. T., Gawalt, E. S. & Mrksich, M. Maleimide-functionalized self-assembled monolayers for the preparation of peptide and carbohydrate biochips. *Langmuir* **19,** 1522–1531 (2003).

88. de la Fuente, J. M. *et al.* Gold glyconanoparticles as water-soluble polyvalent models to study carbohydrate interactions. *Angew. Chem. Int. Ed.* **40,** 2258–2261 (2001).

89. Ratner, D. M., Adams, E. W., Disney, M. D. & Seeberger, P. H. Tools for glycomics: mapping interactions of carbohydrates in biological systems. *ChemBioChem* **5,** 1375–1383 (2004).

90. Rye, P. D. Sweet and sticky: carbohydrate-coated magnetic beads. *Nature Biotechnol.* **14,** 155–157 (1996).

91. Sun, X. L., Cui, W., Haller, C. & Chaikof, E. L. Site-specific multivalent carbohydrate labeling of quantum dots and magnetic beads. *ChemBioChem* **5,** 1593–1596 (2004).

92. Ballell, L. *et al.* A new chemical probe for proteomics of carbohydrate-binding proteins. *ChemBioChem* **6,** 291–295 (2005).

93. Disney, M. D., Zheng, J., Swager, T. M. & Seeberger, P. H. *J. Am. Chem. Soc.* **126,** 13343–13346 (2004).

ARTICLES

The von Neumann relation generalized to coarsening of three-dimensional microstructures

Robert D. MacPherson[1] & David J. Srolovitz[2]

Cellular structures or tessellations are ubiquitous in nature. Metals and ceramics commonly consist of space-filling arrays of single-crystal grains separated by a network of grain boundaries, and foams (froths) are networks of gas-filled bubbles separated by liquid walls. Cellular structures also occur in biological tissue, and in magnetic, ferroelectric and complex fluid contexts. In many situations, the cell/grain/bubble walls move under the influence of their surface tension (capillarity), with a velocity proportional to their mean curvature. As a result, the cells evolve and the structure coarsens. Over 50 years ago, von Neumann derived an exact formula for the growth rate of a cell in a two-dimensional cellular structure (using the relation between wall velocity and mean curvature, the fact that three domain walls meet at 120° and basic topology). This forms the basis of modern grain growth theory. Here we present an exact and much-sought extension of this result into three (and higher) dimensions. The present results may lead to the development of predictive models for capillarity-driven microstructure evolution in a wide range of industrial and commercial processing scenarios—such as the heat treatment of metals, or even controlling the 'head' on a pint of beer.

We consider physical systems in which space is divided into cells/grains/bubbles, which we refer to here as 'domains'. We assume that the growth of these domains is driven by capillarity. This applies to grains in metals or ceramics, and to bubbles in foams[1,2]. The rate of change of the volume of a domain is given by a physical constant $-M\gamma$ times the integral of the mean curvature $\mathcal{K}$ over the walls of the domain[3,4]. Here γ is the surface tension of the domain wall and M is a kinetic coefficient that describes its mobility. If the domains are grains in a metal or a ceramic, then a wall between two adjacent domains moves with velocity $-M\gamma\mathcal{K}$ (the Gibbs–Thompson effect)[5]. If the domains are bubbles in a foam, then $-\gamma\mathcal{K}$ is the pressure difference between adjacent bubbles, and the gas diffuses through the wall at a rate $-M\gamma\mathcal{K}$ per unit area, where M is now proportional to the diffusivity[6].

The integral of the mean curvature over all of the walls of a domain is a difficult quantity to use in theories for domain growth. In 1952, von Neumann[1] found a manageable expression for the growth of domains in two-dimensional space:

$$\frac{\mathrm{d}A}{\mathrm{d}t} = -M\gamma\left(2\pi - \sum_{i=1}^{n}\alpha_i\right) = -2\pi M\gamma\left(1 - \frac{1}{6}n\right) \qquad (1)$$

Here $\mathrm{d}A/\mathrm{d}t$ is the rate of change of the area of a domain, α_i is the exterior (turning) angle at a triple junction on that domain (where three domain walls meet—see Fig. 1a), and n is the number of such triple junctions around a particular domain. A more general derivation of the same result was provided by Mullins[3]. The derivation of the von Neumann–Mullins formula is this: the domain is surrounded by a closed curve, the total integral of the (mean) curvature $\mathcal{K}$ around which (and indeed around any closed curve) is 2π. However, α_i of this total curvature is localized in an abrupt turn at the ith triple junction, so it does not contribute to the integral of $\mathcal{K}$ along the walls. If the surface tension γ is the same on each wall, the equilibrium angle is

$\alpha_i = 2\pi/6$. Like von Neumann and Mullins[1,3], we explicitly assume that the domain walls move sufficiently slowly that this angle remains fixed at its equilibrium value. We treat $\alpha_i = 2\pi/6$ as a boundary condition for three domain walls meeting at a junction (in any dimension ≥ 2).

The von Neumann–Mullins relation is remarkable in that it is both exact and purely topological; the rate of growth of any domain only depends on n, its number of sides (or corners where three domains meet, that is, triple junctions)[1,3]. This relation forms the basis of much of what is known today about isotropic domain growth. Fifty years ago, the grain growth pioneer Smith wrote "It is greatly to be hoped that he [von Neumann], or some other mathematician, will be able to deduce similar relations in three dimensions."[7] Until now, no exact extension of von Neumann's two-dimensional result into three (or higher) dimensions has been found (despite a half century of intense effort in the materials, physics, engineering and mathematics communities[8–13]). The purpose of this Article is to provide an exact extension of this relation into all dimensions, and show that the von Neumann–Mullins relation, and its physically important extension to three dimensions, are simply special cases of the more general mathematical result.

Our three-dimensional von Neumann–Mullins relation is this: the rate of change of the volume V of a domain **D** is given by:

$$\frac{\mathrm{d}V}{\mathrm{d}t} = -2\pi M\gamma\left(\mathcal{L}(\mathbf{D}) - \frac{1}{6}\sum_{i=1}^{n}e_i(\mathbf{D})\right) \qquad (2)$$

The quantity $\mathcal{L}(\mathbf{D})$ is a natural measure of the linear size of domain **D** and $\sum_{i}^{n}e_i(\mathbf{D})$, where e_i is the length of triple line (edge) i and the summation is over all n triple lines of **D** (see Fig. 2a). This result is exact. A proof is provided in the Supplementary Information.

We digress to give the definition and properties of the quantity $\mathcal{L}(\mathbf{D})$. Consider a line, ℓ, through the origin of some coordinate

[1]School of Mathematics, Institute for Advanced Study, Princeton, New Jersey 08540, USA. [2]Department of Physics, Yeshiva University, New York, New York 10033, USA.
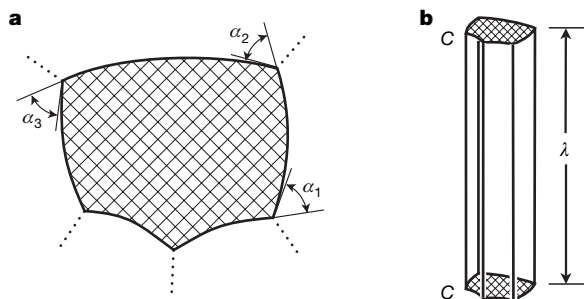
Figure 1 | A two-dimensional grain and its extension into three dimensions. a, A schematic illustration of a grain in a two-dimensional network structure. The dotted lines indicate domain boundaries of adjacent domains, and α_i is the turning angle at triple point i. **b**, The same grain extended into a three-dimensional prism of length λ and cross-sectional shape C.

system in three dimensions; p is a point on ℓ, and $\ell_p^\perp$ is the plane perpendicular to line ℓ at point p. We define the Euler width of a domain $\mathbf{D}$ in the direction ℓ as $E_\ell(\mathbf{D}) = \int_p \chi(\ell_p^\perp \cap \mathbf{D}) \mathrm{d}p$ (see Fig. 2b), where the integral is over all points p along line ℓ and $\chi(\ell_p^\perp \cap \mathbf{D})$ is the Euler characteristic of the intersections of the domain $\mathbf{D}$ with plane $\ell_p^\perp$. The Euler characteristic is the number of objects or pieces in the plane $\ell_p^\perp \cap \mathbf{D}$ minus the number of holes in those objects/pieces. (For a convex object with no holes, $\chi = 1$.) The mean width of $\mathbf{D}$, $\mathcal{L}(\mathbf{D})$, is twice the Euler width, averaged over all lines ℓ through the origin. If the object $\mathbf{D}$ is a line or a curve, the mean width is its length; for a convex object, the mean width is twice the average length of the projection onto a line ℓ, so for a sphere it is twice its diameter. The mean width is additive in the sense that if $\mathbf{D}_1$ and $\mathbf{D}_2$ are objects with mean width $\mathcal{L}(\mathbf{D}_1)$ and $\mathcal{L}(\mathbf{D}_2)$, the union of these two objects has mean width $\mathcal{L}(\mathbf{D}_1 \cup \mathbf{D}_2) = \mathcal{L}(\mathbf{D}_1) + \mathcal{L}(\mathbf{D}_2) - \mathcal{L}(\mathbf{D}_1 \cap \mathbf{D}_2)$. Note that the same additivity rule applies to the volume or surface area of a pair of objects. Hadwiger's theorem states that any measure of the linear dimension of a convex body that is additive and continuous is simply proportional to the mean width[14-18]. The mean width of an object can be computed analytically for many shapes, including all flat-faced polyhedra, and computed numerically for arbitrary shapes, as is shown in detail in Supplementary Information.

A few comments on the main result, equation (2), are in order. First, unlike the von Neumann–Mullins result in two dimensions, this result is not purely topological. The rate of change of the domain volume depends on the mean width of the domain and the total length of the triple lines. It does not, however, depend explicitly on grain shape. Note that the summation of the lengths of all triple lines (in equation (2)) can be described as the mean width of the set of triple lines, $\mathcal{L}(\mathrm{edge}(\mathbf{D}))$, where $\mathrm{edge}(\mathbf{D})$ is the set of triple lines.
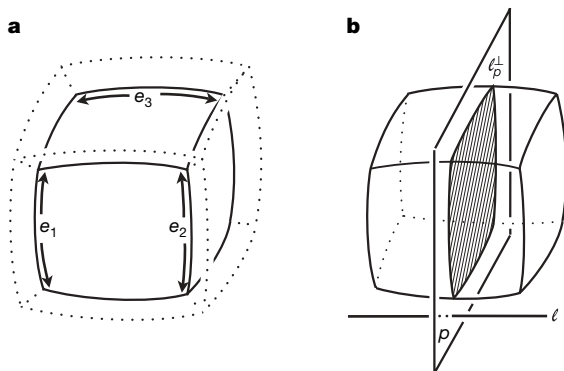
Figure 2 | Notation used in description of the three-dimensional von Neumann–Mullins relation. a, A grain with six faces and twelve edges from a three-dimensional network. e_i are the lengths of edge i. **b**, The intersection of a domain (**D**) and plane $\ell_p^\perp$ which has its normal parallel to ℓ at point p.

We can rewrite the right side of equation (2) as $-2\pi M\gamma \mathcal{L}(\mathbf{D})$ $(1 - f/6)$, where $f = \mathcal{L}(\mathrm{edge}(\mathbf{D}))/\mathcal{L}(\mathbf{D})$. If we write the number of faces on a domain (a face is a region of the surface of $\mathbf{D}$ bounded by triple lines) as m, then it can be seen that f scales as $m^{1/2}$. Using this relation, we can rewrite equation (2) approximately as $\mathrm{d}D^2/\mathrm{d}t = C_1 M\gamma(6 - C_2 m^{1/2})$, where we have assumed that all lengths are proportional to the same linear dimension D of domain $\mathbf{D}$, and C_1 and C_2 are constants. This result looks very similar to the classic two-dimensional von Neumann–Mullins result (a similar result was found by Hilgenfeldt et al.[10]), that is, the right side of the equation is topological. Unlike the exact extension of the von Neumann relation to three dimensions (equation (2)), this result is approximate and simply shows the correct scaling. As both of the terms on the right side of equation (2) are proportional to the linear dimension of the domain, it can be seen that $\mathrm{d}D^3/\mathrm{d}t$ is simply proportional to D or that $D \propto t^{1/2}$. Such parabolic growth is typical of capillarity-driven domain coarsening[19,20].

We can deduce the two-dimensional von Neumann–Mullins relation from equation (2) by considering a prism of cross-sectional shape C with n sides and length λ in three dimensions (see Fig. 1b). For large λ (holding C constant), the prism appears as a line such that $\mathcal{L}(\mathbf{D}) \approx \lambda$ and the sum of the edge lengths is simply $n\lambda$. Inserting these into equation (2) yields $\mathrm{d}V/\mathrm{d}t \approx \lambda \mathrm{d}A/\mathrm{d}t \approx -2\pi M\gamma (\lambda - \lambda n/6)$. Cancelling λ from each side and taking the limit that $\lambda \to \infty$ exactly yields the two-dimensional von Neumann–Mullins relation.

The von Neumann–Mullins relation can be restated by noting that $\mathrm{d}A/\mathrm{d}t = -(\pi M\gamma/3)(6 - n) = -2\pi M\gamma[\chi(\mathbf{D} - \chi(\mathrm{vert}(\mathbf{D}))/\mathbf{6}]$, where $\chi(\mathrm{vert}(\mathbf{D}))$ is the Euler characteristic of the vertices of the domain $\mathbf{D}$ (that is, the number of triple points). We note that this form of the von Neumann–Mullins relation is very similar to our three-dimensional result written in the form $\mathrm{d}V/\mathrm{d}t = -2\pi M\gamma$ $(\mathcal{L}(\mathbf{D}) - \mathcal{L}(\mathrm{edge}(\mathbf{D}))/6))$. The similarity between the two suggests that there may be a more general expression that works in all dimensions. This is, in fact, true and can be expressed as:

$$\frac{\mathrm{d}V_d}{\mathrm{d}t} = -2\pi M\gamma \left(H_{d-2}(\mathbf{D}_d) - \frac{1}{6} H_{d-2}(\mathbf{D}_{d-2}) \right) \qquad (3)$$

where d is the dimension of space, V_d is the volume of domain $\mathbf{D}_d$ in d dimensions, and $\mathbf{D}_{d-2}$ is the $(d-2)$-dimensional feature of the domain (for example, vertices in two dimensions, edges in three dimensions, …). In this expression, H_{d-2} is known as the Hadwiger $(d-2)$-measure in geometric probability[14,15]. The definition of H is akin to that of $\mathcal{L}$ above. Consider a $(d-2)$-dimensional plane, ℓ, through the origin in d dimensions; p is a point on ℓ, and $\ell_p^\perp$ defines the two-dimensional plane perpendicular to ℓ at p. $H_{d-2}(\mathbf{D}_d)$ is equal to the average of $E(\mathbf{D}_d) = \int_p \chi(\ell_p^\perp \cap \mathbf{D}_d)\mathrm{d}p$ over all planes ℓ through the origin. In two dimensions, $H_0(D_2) = \chi(D_2)$ and $H_0(D_0) = \chi(\mathrm{vert}(D))$. In three dimensions, $H_1(\mathbf{D}_3) = \mathcal{L}(\mathbf{D}_3)$ and $H_1(\mathbf{D}_1) = \mathcal{L}(\mathrm{edge}(\mathbf{D}_3))$. These agree with the two-dimensional von Neumann relation and our exact three-dimensional result, respectively.

As most modern models for the evolution of polycrystalline microstructures start with a postulated extension of the two-dimensional von Neumann–Mullins result into three dimensions, the new, exact theory for the evolution of each grain provides a firm foundation for the development of rigorous statistical models for microstructure evolution that respect the underlying geometric and topological constraints of a space-filling network[21,22].

1. von Neumann, J. in *Metal Interfaces* (ed. Herring, C.) 108–110 (American Society for Metals, Cleveland, 1952).
2. Mullins, W. W. in *Metal Surfaces: Structure, Energetics, and Kinetics* (eds Robertson, W. D. & Gjostein, N. A.) 17–66 (American Society for Metals, Metals Park, Ohio, 1963).
3. Mullins, W. W. Two-dimensional motion of idealized grain boundaries. *J. Appl. Phys.* **27**, 900–904 (1956).

4. Glazier, J. A. Grain growth in three dimensions depends on grain topology. *Phys. Rev. Lett.* **70**, 2170–2173 (1993).

5. Burke, J. E. & Turnbull, D. Recrystallization and grain growth. *Prog. Metal Phys.* **3**, 220–292 (1952).

6. Durian, D. J., Weitz, D. A. & Pine, D. J. Scaling behavior in shaving cream. *Phys. Rev. A* **44**, R7902–R7905 (1991).

7. Smith, C. S. in *Metal Interfaces* (ed. Herring, C.) 65–113 (American Society for Metals, Cleveland, 1952).

8. Rhines, F. N. & Craig, K. R. Mechanism of steady-state grain growth in aluminum. *Metall. Trans.* **5**, 413–425 (1974).

9. Rivier, N. On the structure of random tissues or froths, and their evolution. *Phil. Mag. B* **47**, L45–L49 (1983).

10. Hilgenfeldt, S., Kraynik, A. M., Koehler, S. A. & Stone, H. A. An accurate von Neumann's law for three-dimensional foams. *Phys. Rev. Lett.* **86**, 2685–2688 (2001).

11. Hilgenfeldt, S., Kraynik, A. M., Reinelt, D. A. & Sullivan, J. M. The structure of foam cells: Isotropic plateau polyhedra. *Europhys. Lett.* **67**, 484–490 (2004).

12. Glicksman, M. E. Capillarity-mediated grain growth in 3-D. *Mater. Sci. Forum* **467–470**, 1025–1031 (2004).

13. Glicksman, M. E. Analysis of 3-D network structures. *Phil. Mag.* **85**, 3–31 (2005).

14. Hadwiger, H. *Vorlesungen über Inhalt, Oberfläche und Isoperimetrie* (Springer, Berlin, 1957).

15. Klain, D. A. & Rota, G.-C. *Introduction to Geometric Probability* (Cambridge Univ. Press, Cambridge, UK, 1997).

16. Federer, H. Curvature measures. *Trans. Am. Math. Soc.* **93**, 418–491 (1959).

17. Fu, J. H. G. Curvature measures of subanalytic sets. *Am. J. Math.* **116**, 819–880 (1994).

18. Bröcker, L. & Kuppe, M. Integral geometry of tame sets. *Geom. Dedicata* **82**, 285–323 (2000).

19. Atkinson, H. V. Theories of normal growth in pure, single phase systems. *Acta Metall.* **36**, 469–491 (1988).

20. Mullins, W. W. Estimation of the geometrical rate constant in idealized three dimensional grain growth. *Acta Metall.* **37**, 2979–2984 (1989).

21. Hillert, M. On the theory of normal and abnormal grain growth. *Acta Metall.* **13**, 227–238 (1965).

22. Abbruzzese, G., Heckelmann, I. & Lucke, K. Statistical theory of two-dimensional grain growth. *Acta Metall. Mater.* **40**, 519–532 (1992).

ARTICLES

Cell tracing shows the contribution of the yolk sac to adult haematopoiesis

Igor M. Samokhvalov[1], Natalia I. Samokhvalova[1] & Shin-ichi Nishikawa[1]

The first haematopoietic stem cells (HSCs) appear in the aorta-gonad-mesonephros (AGM) region, major vitelline and umbilical vessels, and placenta; however, whether they arise locally or from immigrant yolk sac precursor cells remains unclear. This issue is best addressed by measuring cell-lineage relationships rather than cell potentials. To undertake long-term *in vivo* tracing of yolk sac cells, we designed a non-invasive pulse-labelling system based on Cre/*loxP* recombination. Here we show that in Runx1$^{+/-}$ (runt-related transcription factor 1) heterozygous mice, yolk sac cells expressing Runx1 at embryonic day 7.5 develop into fetal lymphoid progenitors and adult HSCs. During mid-gestation the labelled (embryonic day 7.5) yolk sac cells colonize the umbilical cord, the AGM region and subsequently the embryonic liver. This raises the possibility that some HSCs associated with major embryonic vasculature are derived from yolk sac precursors. We observed virtually no contribution of the labelled cells towards the yolk sac vasculature, indicating early segregation of endothelial and haematopoietic lineages.

Early studies seemed to indicate that in the mammalian embryo, HSCs originate in the yolk sac and subsequently colonize the fetal liver and adult bone marrow. However, it was later demonstrated that during mouse development, the first adult type (definitive) HSCs arise in the AGM region, major umbilical and vitelline vessels, and placenta about 3 days after the onset of yolk sac haematopoiesis[1,2]. Adoptive transfer experiments demonstrated that multipotent, low-level, long-term repopulating cells are generated exclusively within the embryo proper in the para-aortic splanchnopleura but not in the yolk sac, suggesting that definitive HSCs emerge intraembryonically[3,4]. However, the removal of an uncommitted embryo cell from its natural context and transferring it into the microenvironment of a bioassay system can influence the developmental decisions of that cell. The adoptive transfer experiments do not exclude the possibility that yolk sac cells, if allowed to complete normal development in the embryo, can mature to become definitive HSCs. Indeed, yolk sac cells injected transplacentally or into newborn recipients are capable of making a long-term, multipotent contribution to the haematopoietic system, suggesting that an additional *in vivo* maturation step is needed for them to acquire essential properties of HSCs[5,6].

Here we investigate whether early, extra-embryonic blood islands can give rise to HSCs in the adult animal. For this purpose we developed a tightly controlled, non-invasive technique of tagging early yolk sac blood-island cells. The system uses the tamoxifen-inducible *MER-Cre-MER* recombinase gene[7] placed under the control of the *Runx1* locus (also known as *AML1*, *Cbfa2* or *Pebp2a2*). *Runx1* is indispensable for the establishment of the definitive haematopoietic system[8,9], and its expression marks the site of *de novo* generation of haematopoietic progenitors[10]. Inactivation of one of the *Runx1* alleles may reveal the 'hidden' capacity of the yolk sac to generate the HSC lineage[11]. To test this possibility we generated 'knock-in' mice in which a *Runx1* allele was inactivated by targeted insertion of the *MER-Cre-MER* gene fusion (see Supplementary Fig. 1).

Expression of *Runx1* in early development

Using *Runx1-lacZ* (*Runx1$^{lacZ/wt}$*) knock-in mice[12], we investigated the pattern of *Runx1* expression during early embryonic stages (Fig. 1).

Irregular and low-level *Runx1* expression is first seen at embryonic day (E)6.5, mostly in the ectoplacental cone and extra-embryonic visceral endoderm. At around E7.5, intensive β-galactosidase (β-gal) staining appears in the proximal visceral yolk sac region. At this stage we also detected low-level β-gal staining in the ectoplacental cone and associated trophectoderm. However, Runx1-positive cells were completely absent within the embryo proper or allantois at E7.5–E8.25 presomitic stages, ranging from the early bud neural plate to the late headfold, even after over-staining for β-gal (number of embryos tested, $n > 50$; see Supplementary Fig. 2). With the onset of circulation, the bloodstream pushes Runx1$^+$ cells along yolk sac blood vessels into the embryonic vasculature (Fig. 1, inset III). At E8.5 (10–12 somite pairs) practically all staining is confined within blood circulatory cells (Fig. 1, inset IV) except for a few endothelial cells of the omphalomesenteric artery that start to express *Runx1* at a low level, as shown previously[10]. After stable circulation is established, at approximately E9.0 (15–18 somite pairs), the endothelium of the umbilical blood vessels becomes positive for Runx1 (data not shown). By E9.5 (21–23 somite pairs) *Runx1* expression is induced simultaneously in the fetal liver primordium and in the AGM region (Fig. 1).

At E8.5–E9.0, when β-gal staining of yolk sac circulatory cells noticeably fades[10], small, densely stained clusters of cells are discernable, initially in the proximal region of the yolk sac (Supplementary Fig. 3a). Upon expansion of the yolk sac, the clusters multiply in number and spread more distally. They are variable in size and are closely associated with the yolk sac vasculature (Supplementary Fig. 3d, e). At E9.0—immediately before the appearance of definitive progenitors in the embryo proper[13], and the onset of *Runx1* expression in the umbilical vessels, the AGM region and the fetal liver primordium—these yolk sac clusters dissipate into single circulatory cells (Supplementary Fig. 3f). At about E9.5 (22–25 somite pairs), in addition to blood cells, umbilical and vitelline vessels, the AGM region, and the liver primordium, *Runx1* expression is also induced in some non-haematopoietic tissues (Fig. 1).

Tight control of cell labelling *in vivo*

For specific labelling of yolk sac blood islands, we generated an analogous knock-in mouse line in which the *lacZ* reporter was substituted

with the *MER-Cre-MER* gene fusion. In contrast to the *Runx1-lacZ* allele[12], the *MER-Cre-MER* insert is not preceded by a splice acceptor and therefore its expression is driven solely by the proximal P2 promoter (Supplementary Fig. 1). This confines *MER-Cre-MER* expression to an area well inside the observed β-gal⁺ domain.

For genetic cell labelling we used embryos obtained by crossing the heterozygous *Runx1-MER-Cre-MER* knock-in mice ($Runx1^{Cre/wt}$) with homozygous Cre-reporter mouse strains such as the *Rosa26-loxP-STOP-loxP-lacZ* or *R26R* ($Rosa26^{R26R/R26R}$)[14] and the *Rosa26-loxP-STOP-loxP-eYFP* or *R26R-eYFP* ($Rosa26^{R26R-eYFP/R26R-eYFP}$) mice[15]. Induction of recombination, resulting in activation of *lacZ* or *eYFP*, was achieved by a single injection of 4-hydroxytamoxifen (4′OHT) into pregnant females (200–250 mg per kg) at different days of gestation. Such treatment did not compromise embryonic development and did not induce abortion ($n > 40$ litters), but often had an adverse effect on delivery of the pups, which was routinely assisted by caesarean section. 4′OHT itself did not cause any obvious effects on the haematopoietic system, consistent with the complete absence of oestrogen receptors on haematopoietic progenitors in the mouse embryo[16].

First, we analysed the extent of nonspecific cell labelling in our tracing system. Activation of *lacZ* by means of 4′OHT was easily observed, as shown in the example of E9.5–E10.5 embryos induced by a single dose of the drug at E7.5 (Supplementary Fig. 4A). No

staining was detected in control $Runx1^{wt/wt}:Rosa26^{R26R/wt}$ embryos exposed to 4′OHT ($n > 20$) nor $Runx1^{Cre/wt}:Rosa26^{R26R/wt}$ compound embryos not exposed to the drug ($n > 50$, Supplementary Fig. 3A). The peripheral blood of adult control $Runx1^{Cre/wt}:Rosa26^{R26R/wt}$ and $Runx1^{Cre/wt}:Rosa26^{R26R-eYFP/wt}$ mice ($n > 20$) lacked any β-gal⁺ or eYFP⁺ cells, respectively (Supplementary Fig. 4B), which is consistent with the lack of background recombination activity of the MER-Cre-MER fusion[17].

For detailed evaluation of the system, we injected pregnant females of the $Runx1^{Cre/wt} \times Rosa26^{R26R/R26R}$ cross with 4′OHT at E6.5, E7.5, E8.5 and E9.5. Embryos were dissected and analysed at E11.5–E12.0 by X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) staining (Fig. 1). Induction of recombination at E6.5 resulted in very few or no β-gal⁺ blood cells in E11.5 compound $Runx1^{Cre/wt}:Rosa26^{R26R/wt}$ embryos ($n = 16$), although some yolk sacs contained labelled clusters of visceral endodermal and endothelial cells as well as a few primitive blood cells (2 out of 16; data not shown). By E7.5, expression of *Runx1* is strongly upregulated in the proximal yolk sac. 4′OHT-induced recombination at this stage results in the appearance of large numbers of β-gal⁺ blood cells in the circulation of E9.5–E11.5 embryos (Fig. 1, inset I; see also Supplementary Fig. 4A). In all embryos examined ($n = 25$) we observed the accumulation of labelled cells in hepatic sinusoids (Fig. 1, inset II). Most of the labelled cells were yolk-sac-derived primitive erythroblasts, because definitive erythrocytes are not detected before E12 (refs 18, 19). Endothelial labelling in the yolk sac was extremely limited; only a few endothelial LacZ⁺ cells were found in large vessels continuing into the vitelline artery and vein (2 out of 25; data not shown), whereas the endothelium of the umbilical cord was labelled in most cases (16 out of 25).

At about E8.5 *Runx1* expression in primitive haematopoietic cells is downregulated. Accordingly, tamoxifen induction at this stage results in the appearance of fewer circulatory (and liver) β-gal⁺ cells in E11.5–E12.0 embryos ($n = 17$) (Fig. 1). In a notable number of cases (6 out of 17), the endothelium of the umbilical cord (but not of the yolk sac) was also labelled (data not shown). No labelling was detected in non-haematopoietic tissues in all E8.5-induced embryos examined (Fig. 1). By E9.5, *Runx1* expression extends from haematopoietic cells in the yolk sac, embryo circulation, AGM region and fetal liver, to non-haematopoietic tissues such as the olfactory epithelium, spinal ganglia and maxillary processes, as well as external genitalia primordium (Fig. 1). Induction of recombination at E9.5 always resulted in the labelling of these non-haematopoietic tissues and often (8 out of 19) in intensive labelling of the dorsal aorta endothelium and subjacent mesenchymal cells in E11.5–E12.0 embryos (Fig. 1, inset V).

Induction at E6.5 and E8.5 failed to label E7.5 and E9.5 Runx1⁺ cells, respectively, indicating that cell labelling is restricted to a relatively short period of time (<24 h). The short labelling time indicates rapid removal of 4′OHT from the uterus, consistent with pharmacological studies indicating that the drug is rapidly accumulated by a number of internal tissues; this process decreases the drug's concentration in the uterus[20,21].

We conducted a more detailed study of the time course of cell labelling upon 4′OHT administration at E7.5. Substantial variations of developmental stages at the same time points have been observed in gastrulating mouse embryos[22], which creates a range of starting stages in our marking experiments. The time-course studies were designed to investigate whether these variations lead to intraembryonic cell labelling, and to assess the time range of active recombination in the *Rosa26* locus. The first signs of β-gal expression were seen in some embryos just 4 h after injection of 4′OHT (Fig. 2a). After 6 h, the proportion of stained embryos as well as the number of β-gal⁺ cells in the blood-island region increased (Fig. 2b). All embryos that did not sustain recombination in the *Rosa26* locus at this time point were lagging developmentally and were possibly not yet expressing *Runx1* at a sufficient level. However, after 12 h all induced embryos displayed a ring of β-gal⁺ yolk sac cells (Fig. 2c). Despite noticeable
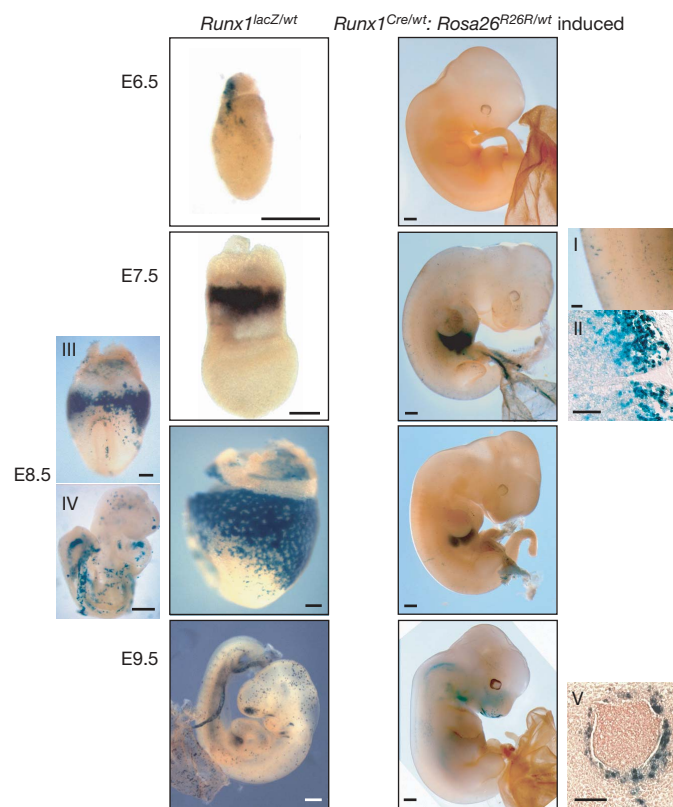
Figure 1 | Embryonic expression of *Runx1* and analysis of inducible *Runx1*-dependent cell tagging. Left column: representative β-gal staining of $Runx1^{lacZ/wt}$ embryos at the designated age. Right column: representative (four independent experiments) staining of E11.5–E12.0 $Runx1^{Cre/wt}:Rosa26^{R26R/wt}$ embryos activated at the indicated age by a single injection of 4′OHT. Insets: I, E7.5-tagged cells in the blood circulation of an E11.5–E12 embryo, higher magnification; II, accumulation of E7.5-tagged blood cells in hepatic sinusoids of a E11.5–E12.0 embryo; III, β-gal/Runx1⁺ yolk sac cells enter the embryo proper at E8.25–E8.5; IV, the same cells in the embryo circulation at E8.5; V, about half of E9.5-activated embryos displayed intensive staining around the dorsal aorta at E11.5–E12.0. Scale bars represent 150 μm in images of E6.5–E8.5 embryos and insets I–IV; 0.5 mm in images of E9.5 and E12.5 embryos; 50 μm in inset V.

embryo stage variation (Fig. 2f), all β-gal activity was observed exclusively in the proximal band of yolk sac cells adjacent to extra-embryonic ectoderm. The actual genetic recombination in the *Rosa26* locus outpaced the appearance of β-gal[+] cells (Fig. 2e and Supplementary Fig. 1e). The β-gal[+] cellular rings in the yolk sac at 12 h were narrower in all induced embryos compared with same-stage *Runx1*[lacZ/wt] embryos (Fig. 2f). However, although better stained, these rings were not broader than those of embryos at 4–6 h after injection. Thus, only a fraction of Runx1[+] yolk sac cells becomes labelled by Cre-recombination at E7.5, and this fraction does not increase with time. This indicates that within 6 h of a single 4′OHT injection the process of cell tagging in *Runx1*-expressing embryos is practically over. Moreover, activation at the advanced E8.0 time point still showed a complete lack of intraembryonic cell labelling (Supplementary Fig. 5), demonstrating that even the concepti that advanced beyond the reported variation range at the E7.5 time point[22] are specifically tagged within their yolk sac component.

The yolk sac contributes to fetal lymphoid cells

Next, we analysed the haematopoietic progeny of yolk sac cells labelled at E7.5 by activation of the *lacZ* or *eYFP* reporter genes. At E9.5–E10.5 the embryos activated at E7.5 showed extensive and consistent staining of primitive Ter119[+] blood cells in the circulation (contribution = 24.0 ± 2.48%, n = 13), whereas contribution to the CD45[+] embryo blood population was lower (9.7 ± 3.8%, n = 13) (Fig. 3a). We also observed a substantial proportion of the progeny of tagged cells (8.9 ± 1.4%, n = 6) in the E9.5 yolk sac c-Kit[+]CD34[+] population, known to be enriched for early newborn liver-repopulating cells[5]. At E12.5, when primitive haematopoiesis yields to fetal liver haematopoiesis, the yolk-sac-derived cells are found in the c-Kit[+]Ter119[+] definitive erythroid progenitor population as well as in myeloid cells of the fetal liver (Fig. 3b). The contribution of the cells derived from the blood-island region to E12.5 yolk sac VE-cadherin[+]CD45[−]Ter119[−] endothelial cells[23] was extremely low (0.09 ± 0.05%, n = 6) (Fig. 3b),

confirming our β-gal-staining data described above. Together, these results suggest that in the E7.5 yolk sac the endothelial and haematopoietic lineages are almost entirely, if not completely, segregated. Further analysis of E16.5 fetal liver and thymus demonstrated a variable contribution of tagged yolk sac cells towards the erythromyeloid and lymphoid lineages (0.1–14%, n = 21 from 5 litters) (Fig. 3c). These data show that the yolk sac can give rise to both fetal myeloid and lymphoid progenitors.

To follow the anatomical distribution of blood-island cells labelled at E7.5, embryos at E10.5 and E11.5 were sectioned after whole-mount LacZ staining. As shown in Fig. 4, cells derived from the E7.5 blood-island region were found incorporated into the vasculature of the umbilical vein and artery. In addition, some β-gal[+] cells apparently attached to the wall of the umbilical vein, forming cell clusters similar to endothelium-associated haematopoietic clusters[10]. Tagged yolk sac cells were also found in the endothelium of the E10.5 dorsal aorta. These observations indicate that the vascular endothelium of large embryonic vessels may represent a temporary niche for developing haematopoietic cells. Thus, we show that at least some cell clusters associated with the endothelium of umbilical cord vessels have their origins in the yolk sac.

Contribution of yolk sac cells to adult haematopoiesis

We then tested whether yolk sac cells tagged at E7.5 have a long-term contribution to adult haematopoiesis. Peripheral blood fluorescence-activated cell sorting (FACS) analysis of 9–12-month-old mice (n = 5, from 3 litters) revealed the contribution of yolk sac cells to the haematopoietic system of adult mice (Fig. 5a, c). In these mice, all major peripheral blood cell lineages as well as the population of c-Kit[+]Ter119[+] bone marrow erythroid progenitors had a yolk sac cell contribution in the range of 1–10% (Fig. 5). Furthermore, we analysed the direct contribution of yolk sac cells to the adult HSC population. As shown in Fig. 5b, the Lin[−]CD34[−]Sca1[+]c-Kit[+] bone marrow population—which is highly enriched with HSCs—comprised a
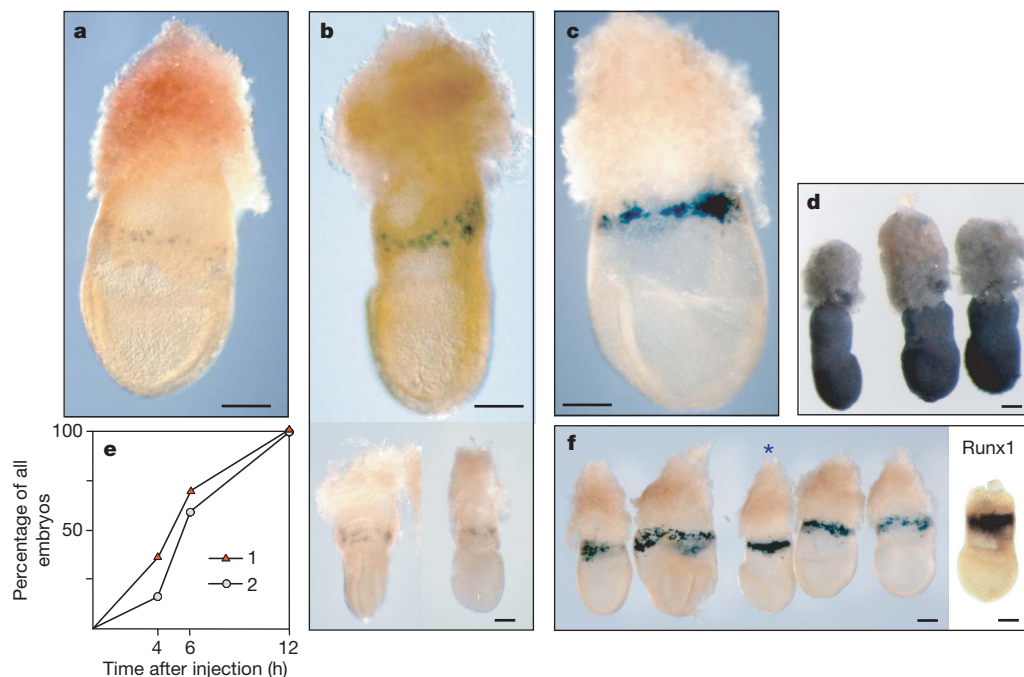
Figure 2 | Embryo cell tagging kinetics at E7.5. a–c, Concepti were fixed and stained whole mount with X-gal at 4 h (**a**), 6 h (**b**) and 12 h (**c**) after a single injection of 4′OHT. **d,** *Rosa26* expression pattern in E7.5 concepti reported by constitutively transcribed proviral β-geo[24]. **e,** Embryo cell tagging dynamics was measured as the percentage (*y* axis) of compound *Runx1*[Cre/wt]:*Rosa26*[R26R/wt] embryos with PCR-detectable recombination in the *Rosa26* locus (triangles) and the fraction of compound embryos

containing β-gal[+] cells (circles) against time after injection (*x* axis). Numbers of compound embryos for each time point: 4 h, n = 14 (three litters); 6 h, n = 10 (three litters); 12 h, n = 17 (five litters). **f,** Comparison of cell tagging in five embryos activated at 12 h; the *Runx1* expression pattern is shown. The asterisk denotes the activated embryo at the same developmental stage as the *Runx1*[lacZ/wt] embryo shown on the far right. Scale bars, 150 μm.
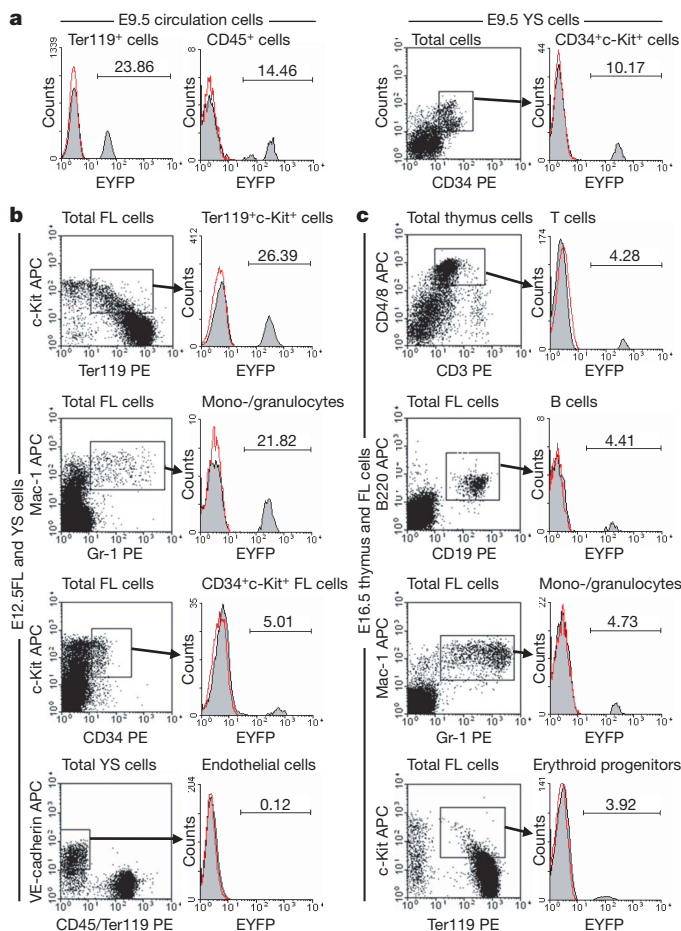
Figure 3 | The embryonic haematopoietic progeny of yolk sac blood-island cells. Red lines in the histograms show the control cell fluorescence of non-activated compound $Runx1^{Cre/wt}$:$Rosa26^{R26R/wt}$ embryos. Grey-filled histograms represent E7.5-activated mice. Numbers in the histograms indicate the percentage of cells in the marked population. The data are representatives of three independent activation experiments. **a,** Blood-island cell descendants in the circulation and in the $CD34^+$c-Kit^+ yolk sac population of E9.5 concepti. **b,** Progeny of E7.5-induced cells in the E12.5 fetal liver and yolk sac. Endothelial cells were defined as having a VE-cadherin$^+$CD45$^-$Ter119$^-$ phenotype. APC, allophycocyanin; FL, fetal liver; PE, phycoerythrin; YS, yolk sac. **c,** Contribution of E7.5 blood-island cells towards E16.5 T cells and major haematopoietic cell lineages of the fetal liver.
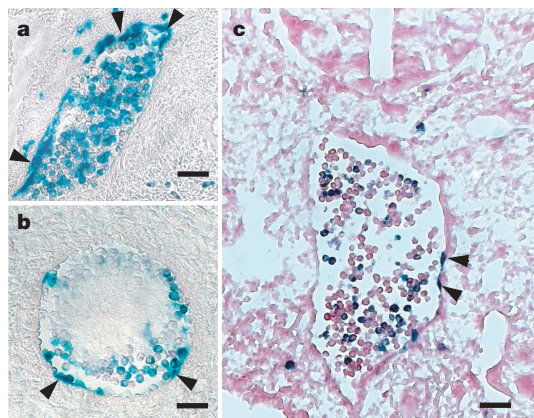
Figure 4 | Representative β-gal staining of E10.5 dorsal aorta and E11.5 umbilical artery and vein sections of embryos activated at E7.5 by a single injection of 4′OHT. a–c, Labelled cells (arrowheads) can be seen at E11.5 in the endothelium of the umbilical vein (**a**) and umbilical artery (**b**), and at E10.5 in the dorsal aorta (**c**). Scale bars, 50 μm.

significant proportion of yolk-sac-cell progeny (7.12 ± 3.12%, $n = 3$), similar to the levels observed in mature cells of the peripheral blood. These results demonstrate that at least some adult HSCs have an extra-embryonic origin.

Early segregation of the HSC lineage

We extended our analysis of the contribution to adult haematopoiesis to other embryonic stages. Analysis of mice activated at E9.5 ($n = 19$) revealed that a large fraction of every major adult blood cell lineage as well as the Lin$^-$CD34$^-$Sca1$^+$c-Kit$^+$ HSC population is represented by β-gal$^+$ descendants of tagged embryonic cells (data not shown). In adult mice of all ages, the contribution of labelled cells to peripheral blood was very high, reaching the relative level of the β-gal$^+$ cell fraction in mice with constitutively expressed *lacZ* under the control of the *Rosa26* promoter[24] (Supplementary Fig. 4B). In total, this virtually 100% contribution was detected in 7 out of 19 mice induced at E9.5. For this stage of *in utero* induction, Fig. 5c shows the representative level of peripheral blood contribution of the tagged cells in nine mice at the age of 9–12 months. Taking into consideration variation in embryonic age and effective concentration of 4′OHT in an individual conceptus, it is safe to conclude that already by E9.5 the definitive HSC lineage in an embryo is totally Runx1 positive. On the basis of these data and the function of Runx1 as a master regulator of early haematopoiesis, we propose that by E9.5 the HSC lineage is fully committed and completely segregated from Runx1$^-$ mesodermal precursors, *de novo* generation of the adult HSC lineage is completed, and thereafter new HSC recruitment is unlikely.

In agreement with our β-gal-staining data we detected no labelled cells in the haematopoietic system of adult mice (9–12-month-old mice) after 4′OHT induction at E6.5. Notably, cells labelled at E10.5—when *Runx1* is highly expressed in the AGM, umbilical cord, blood circulation[10,12] and in all HSCs[25]—contributed to adult haematopoiesis at a significantly lower level compared with cells induced at E9.5 (Fig. 5c). This radical drop in contribution indicates that MER-Cre-MER recombination efficiency is variable across different embryonic stages. After induction at E7.5 there was only partial recombination in the *Runx1*-expressing domain (Fig. 2f). In this regard, the observed lower contribution of yolk sac cells to adult haematopoiesis may not necessarily indicate that the yolk sac generates only a limited fraction of the adult HSC population.

Discussion

Our data support the idea of early segregation of the HSC lineage in the proximal yolk sac region. The yolk sac is typically considered to be a source of primitive blood cells and a restricted subset of definitive haematopoietic progenitors incapable of generating lymphoid cells; it is not seen as an autonomous site of HSC *de novo* generation[4,13]. Our non-invasive cell-tracing analysis directly demonstrates the migration of haematopoietic progenitors from the yolk sac to the fetal liver and thymus, and that yolk sac blood islands contain the precursors of adult HSCs. We also show that yolk sac cells colonize the luminal surface[26] of big vessels in the embryo and umbilical cord. It is conceivable that yolk-sac-derived pre-HSCs acquire functional HSC characteristics in the AGM/umbilical artery and umbilical vein region. However, the limits of our labelling system did not allow us to study the *de novo* generation of HSCs in this region, and so this remains an open question.

Our cell-tracing experiments were performed in mice with one *Runx1* allele supposedly inactivated by the *MER-Cre-MER* insert, similar to the inactivation of the *Runx1*^{lacZ} allele[12]. Although the *Runx1* heterozygotes were apparently normal, it remains to be addressed whether our observations are relevant to 'normal' haematopoiesis, because proposed haploinsufficiency of *Runx1* shifts the anatomy and dynamics of haematopoietic progenitor generation, possibly due to the increased production of haematopoietic progenitors in the AGM region[11]. However, a mild haploinsufficiency phenotype in knockout models may arise as a result of a DNA hitchhiking effect[27].
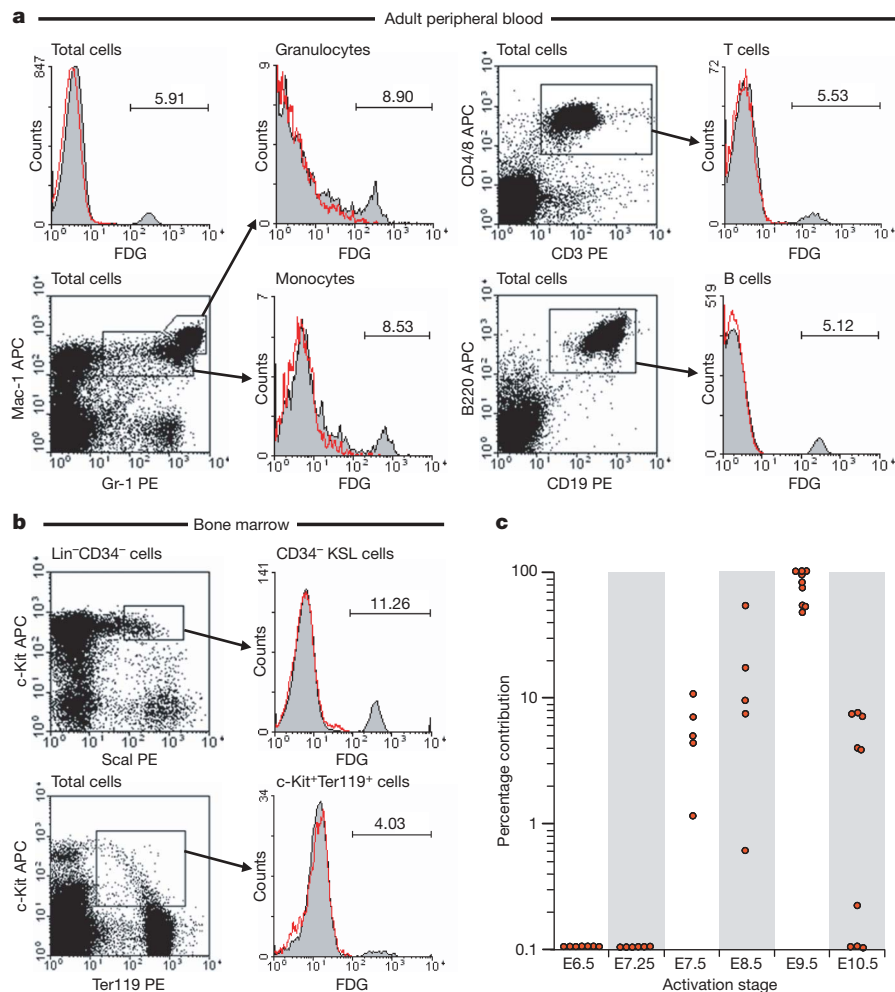
Figure 5 | The haematopoietic progeny in adult mice of yolk sac cells from the blood-island region. Red lines in the histograms show cell fluorescence of control, non-activated compound *Runx1^Cre/wt^:Rosa26^R26R/wt^* mice. Grey-filled histograms represent E7.5-activated mice. Numbers in the histograms indicate the percentage of cells in the marked population. **a,** Representative contribution of cells of the blood-island lineage labelled at E7.5 to major lineages from peripheral blood of adult mice. **b,** The progeny of cells labelled at E7.5 contribute significantly to the Lin⁻CD34⁻ Sca1⁺c-Kit⁺

(CD34⁻KSL) HSC population and erythroid progenitors from the bone marrow in adult mice (9–12 months old). The data are a representative FACS analysis of bone marrow cells from three mice. **c,** Contribution towards adult haematopoiesis of cells labelled at different developmental stages. The adjusted percentage of tagged white blood cells versus staining levels of mice constitutively expressing *lacZ* in the *Rosa26* locus[24] is plotted on the logarithmic *y* axis; embryonic stage of the pulse 4′OHT activation is shown on the *x* axis.

Mouse strains differ markedly in the number of haematopoietic stem/progenitor cells and have distinct allelic repertoires of genes controlling the frequencies of the progenitors[28]. There are three genes for interferon receptors located within 1.1 cM of the *Runx1* locus (MGI 3.43, Mouse Chromosome 16 Linkage Map, 62.2–64.0 cM; http://www.informatics.jax.org/), and these genes might be included in this repertoire, because interferons strongly affect haematopoietic progenitor proliferation[29,30]. In general terms, the early extra-embryonic origin of the HSC lineage represents an efficient ontogenetic process that markedly differs from the one proposed for intraembryonic sources. If the critical elements of the genetic programme controlling this process were evolutionarily preserved as a gene dosage compensation mechanism, it would suggest that the early embryo dose of Runx1 varies in normal mouse populations. Therefore, it is reasonable to assume that the yolk sac normally participates in establishing the adult haematopoietic system, but the extent of this participation remains to be determined.

METHODS

Mice. Generation of *Runx1-MER-Cre-MER* knock-in mice is described in Supplementary Methods. The knock-in allele was maintained on either a C57BL/6 or ICR outbred background. The Cre-reporter *R26R* [B6;129S4-Gt(ROSA)26Sor^tm1Sor^/J] mice were from the Jackson Laboratory and *R26R-*

eYFP mice were provided by F. Costantini. *ROSA26* [B6;129S-*Gt(ROSA)26Sor*/J] mice with constitutive expression of *β-geo* in the *Rosa26* locus were purchased from the Jackson Laboratory.

Induction of cell tagging and flow cytometry. For the timing of 4′OHT injections, the day of the copulation plug was regarded as 0.5 days post coitus. The detailed procedure for drug administration is described in Supplementary Methods. All antibodies used for the FACS analysis were purchased from BD PharMingen. Expression of β-gal was examined using fluorescein di-β-D-galactopyranoside (FDG; Molecular Probes). The FACS analysis procedures are described in Supplementary Methods.

Whole-mount X-gal staining, cryosectioning and PCR genotyping. Embryos from the timed matings were fixed and stained with X-gal at 37 °C overnight or for 4 days for the over-staining. The cryosectioning, PCR procedures and primers are described in Supplementary Methods.

1. Medvinsky, A. & Dzierzak, E. Definitive hematopoiesis is autonomously initiated by the AGM region. *Cell* **86,** 897–906 (1996).
2. Gekas, C., Dieterlen-Lievre, F., Orkin, S. H. & Mikolla, H. K. A. The placenta is a niche for hematopoietic stem cells. *Dev. Cell* **8,** 365–375 (2005).
3. Cumano, A., Dieterlen-Lievre, F. & Godin, I. Lymphoid potential, probed before circulation in mouse, is restricted to caudal intraembryonic splanchnopleura. *Cell* **86,** 907–916 (1996).

4. Cumano, A., Ferraz, J. C., Klaine, M., Di Santo, J. P. & Godin, I. Intraembryonic, but not yolk sac hematopoietic precursors, isolated before circulation, provide long-term multilineage reconstitution. *Immunity* **15**, 477–485 (2001).

5. Yoder, M. C. *et al.* Characterization of definitive lymphohematopoietic stem cells in the day 9 murine yolk sac. *Immunity* **7**, 335–344 (1997).

6. Weissman, I., Papaioannou, V. & Gardner, R. *Fetal Hematopoietic Origins of the Adult Hematolymphoid System* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1978).

7. Zhang, Y. *et al.* Inducible site-directed recombination in mouse embryonic stem cells. *Nucleic Acids Res.* **24**, 543–548 (1996).

8. Okuda, T., Van Deursen, J., Hiebert, S. W., Grosveld, G. & Downing, J. R. AML1, the target of multiple chromosomal translocations in human leukemia, is essential for normal fetal liver hematopoiesis. *Cell* **84**, 321–330 (1996).

9. Wang, Q. *et al.* Disruption of the *Cbfa2* gene causes necrosis and hemorrhaging in the central nervous system and blocks definitive hematopoiesis. *Proc. Natl Acad. Sci. USA* **93**, 3444–3449 (1996).

10. North, T. *et al. Cbfa2* is required for the formation of intra-aortic hematopoietic clusters. *Development* **126**, 2563–2575 (1999).

11. Cai, Z. *et al.* Haploinsufficiency of AML1 affects the temporal and spatial generation of hematopoietic stem cells in the mouse embryo. *Immunity* **13**, 423–431 (2000).

12. Samokhvalov, I. M. *et al.* Multifunctional reversible knockout/reporter system enabling fully functional reconstitution of the AML1/Runx1 locus and rescue of hematopoiesis. *Genesis* **44**, 115–121 (2006).

13. Palis, J., Robertson, S., Kennedy, M., Wall, C. & Keller, G. Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development* **126**, 5073–5083 (1999).

14. Soriano, P. Generalized *lacZ* expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).

15. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of *EYFP* and *ECFP* into the *ROSA26* locus. *BMC Dev. Biol.* **1**, 4 (2001).

16. Igarashi, H., Kouro, T., Yokota, T., Comp, P. C. & Kincade, P. W. Age and stage dependency of estrogen receptor expression by lymphocyte precursors. *Proc. Natl Acad. Sci. USA* **98**, 15131–15136 (2001).

17. Kellendonk, C. *et al.* Inducible site-specific recombination in the brain. *J. Mol. Biol.* **285**, 175–182 (1999).

18. Brotherton, T. W., Chui, D. H. K., Gauldie, J. & Patterson, M. Hemoglobin ontogeny during normal mouse fetal development. *Proc. Natl Acad. Sci. USA* **76**, 2853–2857 (1979).

19. Steiner, R. & Vogel, H. On the kinetics of erythroid cell differentiation in fetal mice: I. Microspectrophotometric determination of the hemoglobin content in erythroid cells during gestation. *J. Cell. Physiol.* **81**, 323–338 (1973).

20. Lien, E. A., Solheim, E. & Ueland, P. M. Distribution of tamoxifen and its metabolites in rat and human tissues during steady-state treatment. *Cancer Res.* **51**, 4837–4844 (1991).

21. Kisanga, E. R., Gjerde, J., Schjott, J., Mellgren, G. & Lien, E. A. Tamoxifen administration and metabolism in nude mice and nude rats. *J. Steroid Biochem. Mol. Biol.* **84**, 361–367 (2003).

22. Downs, K. M. & Davies, T. Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* **118**, 1255–1266 (1993).

23. Nishikawa, S.-I. *et al. In vitro* generation of lymphohematopoietic cells from endothelial cells purified from murine embryos. *Immunity* **8**, 761–769 (1998).

24. Mao, X., Fujiwara, Y. & Orkin, S. H. Improved reporter strain for monitoring Cre recombinase-mediated DNA excisions in mice. *Proc. Natl Acad. Sci. USA* **96**, 5037–5042 (1999).

25. North, T. E. *et al.* Runx1 expression marks long-term repopulating hematopoietic stem cells in the midgestation mouse embryo. *Immunity* **16**, 661–672 (2002).

26. Minot, C. S. Development of the blood, the vascular system and the spleen. In *Manual of Human Embryology* (eds Keibel, F. and Mall, F. P.) 498–534 (J. B. Lippincott, Philadelphia, 1912).

27. Gerlai, R. Gene-targeting studies of mammalian behavior: is it the mutation or the background genotype? *Trends Neurosci.* **19**, 177–181 (1996).

28. Müller-Sieburg, C. E. & Riblet, R. Genetic control of the frequency of hematopoietic stem cells in mice: mapping of a candidate locus to chromosome 1. *J. Exp. Med.* **183**, 1141–1150 (1996).

29. Young, H. A. *et al.* Bone marrow and thymus expression of interferon-γ results in severe B-cell lineage reduction, T-cell lineage alterations, and hematopoietic progenitor deficiencies. *Blood* **89**, 583–595 (1997).

30. Yu, J.-M. *et al.* Expression of interferon-γ by stromal cells inhibits murine long-term repopulating hematopoietic stem cell activity. *Exp. Hematol.* **27**, 895–903 (1999).

LETTERS

Chlorine isotope homogeneity of the mantle, crust and carbonaceous chondrites

Z. D. Sharp[1], J. D. Barnes[1], A. J. Brearley[1], M. Chaussidon[2], T. P. Fischer[1] & V. S. Kamenetsky[3]

Chlorine in the Earth is highly depleted relative to carbonaceous chondrites and solar abundances[1]. Knowledge of the Cl concentrations and distribution on Earth is essential for understanding the origin of these depletions. Large differences in the stable chlorine isotope ratios of meteoritic, mantle and crustal materials have been used as evidence for distinct reservoirs in the solar nebula[2] and to calculate the relative proportions of Cl in the mantle and crust[3]. Here we report that large isotopic differences do not exist, and that carbonaceous chondrites, mantle and crust all have the same $^{37}Cl/^{35}Cl$ ratios. We have further analysed crustal sediments from the early Archaean era to the Recent epoch and find no systematic isotopic variations with age, demonstrating that the mantle and crust have always had the same $\delta^{37}Cl$ value. The similarity of mantle, crust and carbonaceous chondrites establishes that there were no nebular reservoirs with distinct isotopic compositions, no isotopic fractionation during differentiation of the Earth and no late (post-core formation) Cl-bearing volatile additions to the crustal veneer with a unique isotopic composition.

Quantification of the Cl concentration and distribution on Earth would help to constrain models of planetary formation that are based on depletion due to nebular fractionation processes during formation of the inner planets, or due to early catastrophic degassing followed by expulsion during meteoritic bombardment[4]. In addition to Cl abundances alone, an independent constraint is provided by $^{37}Cl/^{35}Cl$ ratios. Distinct nebular reservoirs, or mixing between different reservoirs, could be identified by variations in the $^{37}Cl/^{35}Cl$ ratio of different materials. Using mass balance constraints, differences in the $\delta^{37}Cl$ value of the mantle, crust and carbonaceous chondrites could also be used to define the proportions of Cl in the crust and mantle reservoir if the chlorine isotope compositions of the different reservoirs were known.

Of all the important Cl-bearing terrestrial reservoirs, the crust is by far the best characterized for Cl isotopes. The oceans and evaporites make up the vast majority of the crustal Cl reservoir; the ocean has a constant $\delta^{37}Cl$ value of 0‰ (ref. 5) and evaporites are generally within 0.5‰ of the ocean value[6] (Fig. 1). Previous estimates of the isotopic composition of the mantle were made from analyses of pristine mid-ocean ridge basalt (MORB) glasses, using the thermal ionization mass spectrometry of CsCl. Values range from 3.0‰ to 7.2‰, and average 4.7‰ (refs 3, 7).

Only a few meteorites have been analysed for chlorine isotope ratios. Three analyses of bulk carbonaceous chondrites are between 2.7‰ and 4.0‰ (refs 3, 7). The positive $\delta^{37}Cl$ values of bulk carbonaceous chondrites are intermediate to those of mantle and crust, and were used to calculate the mantle/crust Cl ratio, with ~60% Cl residing in the degassed mantle[3]. The $\delta^{37}Cl$ values of halite and water-soluble chloride from two ordinary chondrites range from −2.8‰ to −1.4‰ (ref. 2). The negative values for the water-soluble fraction

were interpreted either as representing a second reservoir from the early Solar System or reflecting unknown large isotope fractionations linked to low-temperature fluid activity[2]. All meteorite data were measured by thermal ionization mass spectrometry of CsCl.

Details of all samples analysed and methods are given in the Supplementary Information. Five carbonaceous chondrites[8] were analysed for bulk chlorine isotope compositions using gas source mass spectrometry (Table 1). In addition, high-spatial-resolution spot analyses of sodalite grains from calcium aluminium inclusions (CAI) from Allende were made using a large radius ion microprobe. Allende sodalite has been shown to have ^{36}Cl and ^{26}Al anomalies, suggesting that its formation occurred very early in the history of the Solar System[9].

We analysed twelve MORB glasses to determine the $\delta^{37}Cl$ value of depleted MORB (Table 1). All are optically and chemically pristine, and appear to be free of seawater alteration. The samples cover a broad range of MORB types, coming from different ridges, and include low degrees of partial melting with high Cl content to depleted high degree of partial melting with low-Cl-content glasses.

Three mantle-derived samples in a continental setting were analysed. (1) Fresh carbonatite lavas from Oldoinyo Lengai, Tanzanian rift valley, collected during an active eruption in 2005. Rare gas and carbon and oxygen isotope ratios are consistent with a mantle source and lack of surficial alteration. (2) Massive halite nodules found in the surficially unaltered Siberian Udachnaya-East kimberlite[10]. (3) Djerfisherite ($K_6Na(Fe,Cu,Ni)_{24}S_{26}Cl$) from pegmatites from the Khibina massif, Kola peninsula, Russia. The Khibina massif is an enormous intrusive complex, presumably of mantle origin of Caledonian age (500–360 Myr ago).

Sediments covering a wide range of ages were analysed to determine whether secular variations in the $\delta^{37}Cl$ value of the crust exist. Twelve chert samples from the Archaean and Proterozoic eras and one from the Jurassic period were analysed. Other Precambrian bulk samples include an iron formation, modified evaporite (barite) and dolomite. Data from these samples and Phanerozoic halites (including the oldest-known salt deposit) are given in Table 2, and plotted in Fig. 2. An experiment was undertaken in order better to understand fractionation mechanisms that might be relevant to chlorine partitioning between phases in the nebular reservoir and, in particular, sodalite as a primitive Cl-bearing phase in chondrites. Nepheline and excess NaCl were reacted in a sealed platinum capsule at 825 °C for 48 h to form sodalite. The sodalite and excess salt in the reaction capsule were analysed for their $\delta^{37}Cl$ values using isotope ratio mass spectrometry (Supplementary Table SI-2) to determine the equilibrium fractionation between these two phases.

The average of all carbonaceous chondrites is 0.0 ± 0.7‰ (1σ) (all data are reported relative to the Standard Mean Ocean Chloride or SMOC[5]) (Table 1). (The Orgueil chondrite is a >2σ outlier when

excluded from the average and standard deviation. Although it is known to have undergone terrestrial alteration[11], we cannot see how this would change the $\delta^{37}Cl$ value, nor why the other CI chondrite, Ivuna, does not show the same ^{37}Cl enrichment. Orgueil is therefore included in the average.) Sodalite $\delta^{37}Cl$ values average $-1.3‰$ ($\pm0.6‰$, 1σ), statistically lighter than bulk meteorite data. The low values, compared to bulk chondrites, may represent a second nebular Cl reservoir, but can also be explained, and are indeed expected, in terms of equilibrium fractionation. The $1,000\ln\alpha_{sodalite - NaCl}$ value (where $\alpha_{a-b} = (\delta_a + 1,000)/(\delta_b + 1,000)$) from our experiment (see Supplementary Information) is $-0.3 \pm 0.1‰$ (1σ) at $825\,°C$, a fractionation that should increase with decreasing temperature. Fractionations of $-1‰$ at lower temperature are reasonable, and there is no need to call on multiple nebular reservoirs, although the data do not preclude it.

Our Phanerozoic crustal halite data average $0.0 \pm 0.3‰$, and show no correlation with age (Fig. 2). Bulk chert data (presumably from saline fluid inclusions) covering an age range of 3,500 to 150 Myr ago, average $-0.7 \pm 1.0‰$. The data have a lower average isotopic composition than for the halite, which may in part be explained by a small positive equilibrium fractionation between halite and aqueous solution[12]. More importantly, however, is that there is no correlation with age, as is seen also with the Proterozoic halite samples.

The $\delta^{37}Cl$ values of bulk carbonaceous chondrites, depleted oceanic mantle, subcontinental mantle and the crust are statistically indistinguishable (Table 1, Fig. 1). If the mantle and crust had different $\delta^{37}Cl$ values, then simple box modelling shows that there would be a measurable secular variation in the crust, given the enormous Cl flux between the two reservoirs[13]. That this is not the case further supports our contention that the crust and mantle have indistinguishable $\delta^{37}Cl$ values. The identical mantle and crustal $\delta^{37}Cl$ values suggest a common single Cl reservoir in the Earth, and

Figure 1 | $\delta^{37}Cl$ values of selected terrestrial and chondritic samples.
Samples were analysed from the crust, mantle and chondritic meteorites by gas source mass spectrometry (shown as diamonds). Crustal evaporites and the ocean have $\delta^{37}Cl$ values clustering tightly around 0‰ versus SMOC. Mantle samples include MORB (samples from the 9° N and 12° N regions of the East Pacific Rise, the Pacific–Antarctic rise crest 53°–57° S, Iceland and Macquarie Island, an above-sea level exposure of the Macquarie Ridge, representing the Miocene spreading of the Australia–Pacific plate boundary), and sub-continental mantle (halite nodules in dimondiferous kimberlite[10]; fresh carbonatite from Oldonyo Lengai[27] and an igneous djerfisherite ($K_6Na(Fe,Cu,Ni)_{24}S_{26}Cl$) from the Kola peninsula, Russia). Although the MORB samples may have suffered some seawater contamination, the low Cl/K and Cl/Nb ratios and the lack of correlation between Cl content and these ratios argues against significant seawater infiltration. The average for all mantle samples is $-0.1 \pm 0.4‰$ (1σ), with some scatter that may represent minor mantle heterogeneity. Bulk carbonaceous chondrite samples have more scatter, but still average $0.0 \pm 0.7‰$, indistinguishable from the terrestrial reservoir. Previous analyses made with solid source mass spectrometry (open circles, bottom of figure) have higher $\delta^{37}Cl$ values and are more variable. These data probably suffered from analytical artefacts and are not considered reliable (see Supplementary Information). The sodalite inclusions in a CAI from Allende was measured using a Cameca 1270 ion microprobe (triangles). The isotopically light sodalite may represent a second light nebular reservoir or may be explained in terms of equilibrium fractionation at low temperatures. All data are from this study, except for carbonaceous chondrites[3,7] and H3-6 chondrite, both analysed with thermal ionization mass spectrometry[2], and selected evaporites[6] and carbonatites[28] for which only samples with $\delta^{18}O$ values close to primary were considered. Analytical uncertainty (1σ) is equal to symbol size except where indicated.

Table 1 | $\delta^{37}Cl$ values of mantle and carbonaceous chondrite samples

Material	Sample ID	$\delta^{37}Cl$ (‰ versus SMOC)	Cl content (p.p.m.)
Carbonaceous chondrites			
CM chondrite	Murchison (*di*)	0.46, 0.15	n.d.
CI chondrite	Ivuna	−0.35	290
CI chondrite	Orgueil	1.21	603
C2 ungrouped chondrite	Tagish Lake	−0.71	n.d.
CV chondrite	Allende (*di*)	−0.39, −0.37	n.d.
Average bulk meteorite	Average 1σ	0.0 (0.7)	
Sodalite in Allende	CAI 4-2	−2.09	
Sodalite in Allende	CAI 4-3	−1.62	
Sodalite in Allende	CAI 4-4	−0.72	
Sodalite in Allende	CAI 4-5	−1.44	
Sodalite in Allende	CAI 4-7	−0.65	
Sodalite in Allende	CAI 4-8	−1.76	
Sodalite in Allende	CAI 4-9	−1.72	
Sodalite in Allende	CAI 2-1	−1.40	
Sodalite in Allende	CAI 2-2	−0.39	
Sodalite in Allende	CAI 2-3	−1.33	
Average sodalite in allende	Average 1σ	−1.3 (0.6)	
MORB glass			
9° N on-axis, East Pacific Rise	ALV 2358-4	−0.40	59
9° N off-axis, East Pacific Rise	ALV 2746-12	−0.24	78
9° N off-axis, East Pacific Rise	ALV 2772-1	0.27	63
9° N off-axis, East Pacific Rise	ALV 2772-2	0.05	59
10° N, East Pacific Rise	CHEPR	−0.09	86
12° N on-axis, East Pacific Rise	CH7-1	−1.04	327
12° N on-axis, East Pacific Rise	CH116-2	−1.03	214
Iceland, North Atlantic	DICE 13	0.36	160
Macquarie Ridge	47963 (*di*)	0.06	1,400
Pacific-Antarctic rise crest, 53–57° S	D64-2	0.37	66
Pacific-Antarctic rise crest, 53–57° S	D62-1A	−0.46	203
Pacific-Antarctic rise crest, 53–57° S	D17-1A	−0.16	67
Average MORB glass value	Average 1σ	−0.2 (0.5)	
Subcontinental mantle materials			
Djerfisherite, Kola Peninsula		−0.35	12,200
Carbonatite, Tanzanian rift valley	OL-D1 (*di*)	0.04	n.d.
Carbonatite, Tanzanian rift valley	OL-D3 (*di*)	−0.15	n.d.
Carbonatite, Tanzanian rift valley	OL-D2 (*di*)	0.00	n.d.
Halite from kimberlite, Siberia	UV-2-03 (*di*)	0.30, −0.03, −0.18	
Average subcontinental mantle value	Average 1σ	−0.03 (0.25)	

Multiple values denote repeats. See ref. 8 for a detailed discussion of chondrite classification. The Tagish Lake chondrite has an extremely high volatile content and is classified following ref. 16. Except for sodalite in Allende, samples were analysed in dual inlet (*di*) mode; all others were measured in continuous flow (see Supplementary Information). Sodalite was measured using the Cameca 1270 ion microprobe at CNRS, Nancy, France using conditions similar to ref. 17. A natural sodalite was used as a standard (calibrated using gas source mass spectrometry). The standard deviation of the sodalite standard analysed in the same session was 0.19‰ (1σ), suggesting that the scatter in Allende is not an analytical artefact (see Supplementary Information). Cl contents from the ALV and CH MORB samples from E. Hauri (Petrological Database, PETDB; http://www.petdb.org/) determined using ion microprobe.

Table 2 | $\delta^{37}Cl$ values of halite and fluid inclusion samples of various ages

Sample	Location	Age (Myr ago)	Cl content (p.p.m.)	$\delta^{37}Cl$ (‰)
Halite				
Wieliczka salt mine	Poland[18]	13		0.09
WIPP (Waste Isolation Pilot Plant)	New Mexico, USA[19]	250		−0.02, −0.15
Ghafeer-5	Oman[20]	543		−0.24, −0.37
Minha-1	Oman[20]	544		0.14, 0.03
Bulk sediment samples				
Chert, Onverwacht Group	South Africa[21]	3,400	23.	0.77, 0.42
North Pole barite	West Australia[22]	~3,450	n.d.	2.20, 2.52
Kona dolomite	Michigan[23]	~2,200	n.d.	0.16, −0.39
Holy Cross Mountains chert	Poland[24]	~155	14	−0.51
Isua iron formation	Greenland[25]	3,850	15	−0.78

Geologic/tectonic unit	Sample location	Sample number	Age (Myr ago)	Cl content (p.p.m.)	$\delta^{37}Cl$ (‰)
Chert samples from the Precambrian Paleobiology Research Group sample suite[26]					
Bitter Springs Formation/Amadeus basin	Northern Territory, Australia	1,315	850	627	−1.35
Siyeh Formation/Belt basin	Montana, USA	1,026	1,450	15	−1.01, −0.30
Gunflint iron formation	South Ontario, Canada	1,284	1,878	67	−1.68
Towers Formation/Pilbara block	West Australia	1,437	2,200	25	1.04
Fig Tree group/ Barberton greenschist belt	Transvaal, South Africa	2,569	2,200	28	−0.41, 0.37
Gamohaan Formation/Kaapvaal craton	Transvaal, South Africa	2,515	2,500	21	−0.21
Maddina basalt/Hamersley basin	West Australia	2,033	2,700	25	−3.16
Tumbiana Formation/Hamersley basin	West Australia	2,095	2,700	15	−0.24
Hooggenoeg Formation/Barberton greenstone belt	Transvaal, South Africa	1,386	3,450	68	−0.83
Apex basalt/Pilbara block	West Australia	1,998	3,500	34	−0.57, −0.36

Halites measured in dual inlet, all others measured in continuous flow.

that early mantle degassing[14] of Cl was not accompanied by isotopic fractionation.

The similar $\delta^{37}Cl$ values of terrestrial reservoirs and carbonaceous chondrites requires that there should be little to no fractionation associated with volatile loss during the early stages of planetary accretion or with volatile loss due to later impact heating and ejection[4]. Further, no Cl isotope fractionation occurred between the crust and mantle. The question of the core is unresolved, but if a significant fraction of Earth's chlorine resides in the core[15], then fractionation between the core and the bulk silicate Earth must also be nearly zero for the bulk Earth and carbonaceous chondrites to have the same $\delta^{37}Cl$ value. The discrepancy between the present data set and previously published data are explained by analytical artefacts of the latter. The thermal ionization mass spectrometry analyses of CsCl resulted in artificially high $\delta^{37}Cl$ values (see Supplementary Methods).

The similarity of all major Cl reservoirs measured in this study supports the idea of isotopic homogeneity at all stages of planetary formation. Equally important, the new data eliminate some unreasonable

constraints on nebular heterogeneity and mantle–crust fractionation that were required by the previous data. Whether catastrophic or continual in nature[14], Cl transfer from the mantle to crust occurred without measurable isotopic fractionation. Chondritic averages are identical to those in the bulk Earth, but the range of $\delta^{37}Cl$ values of chondritic reservoirs is well outside analytical uncertainty. The spread can be explained solely by low-temperature isotopic fractionation, although further work is warranted to identify possible variations in nebular material.

1. Wänke, H. & Dreibus, G. Chemical compilation and accretion history of terrestrial planets. *Phil. Trans. R. Soc. Lond. A* **325**, 545–557 (1988).
2. Bridges, J. C., Banks, D. A., Smith, M. & Grady, M. M. Halite and stable chlorine isotopes in the Zag H3-6 breccia. *Meteorit. Planet. Sci.* **39**, 657–666 (2004).
3. Magenheim, A. J., Spivack, A. J., Michael, P. J. & Gieskes, J. M. Chlorine stable-isotope composition of the oceanic-crust: Implications for Earth's distribution of chlorine. *Earth Planet. Sci. Lett.* **131**, 427–432 (1995).
4. Kargel, J. S. Earth's volatility trend and the case of the missing halogens. *Geol. Soc. Am. Prog. Abstr.* **29**, 190 (1997).
5. Kaufmann, R., Long, A., Bentley, H. & Davis, S. Natural chlorine isotope variations. *Nature* **309**, 338–340 (1984).
6. Barnes, J. D. *Tectonic and Metamorphic Implications of High Chlorine Contents in Serpentinites.* 1–161, PhD thesis, Univ. New Mexico (2006).
7. Magenheim, A. J., Spivack, A. J., Volpe, C. & Ransom, B. Precise determination of stable chlorine isotopic ratios in low-concentration natural samples. *Geochim. Cosmochim. Acta* **58**, 3117–3121 (1994).
8. Brearley, A. J. & Jones, R. H. in *Planetary Materials* (ed. Papike, J. J.), **3**, 1–398 (Mineral. Soc. Am., Washington DC, 1998).
9. Hsu, W., Guan, Y., Leshin, L. A., Ushikubo, T. & Wasserburg, G. J. A late episode of irradiation in the early Solar System: Evidence from extinct ^{36}Cl and ^{26}Al in meteorites. *Astrophys. J.* **640**, 525–529 (2006).
10. Kamenetsky, M. B., Sharygin, V. V., Faure, K. & Golovin, A. V. Chloride and carbonate immiscible liquids at the closure of the kimberlite magma evolution (Udachnaya-East kimberlite, Siberia). *Chemical Geol.* **237**, 384–400 (2007).
11. Gounelle, M. & Zolensky, M. E. A terrestrial origin for sulfate veins in CI1 chondrites. *Meteorit. Planet. Sci.* **36**, 1321–1329 (2001).
12. Eggenkamp, H. G. M., Kreulen, R. & Koster van Groos, A. F. Chloride stable isotope fractionation in evaporites. *Geochim. Cosmochim. Acta* **59**, 5169–5175 (1995).
13. Sharp, Z. D. & Barnes, J. D. Water-soluble chlorides in massive seafloor serpentinites: a source of chloride in subduction zones. *Earth Planet. Sci. Lett.* **226**, 243–254 (2004).
14. Rubey, W. W. Geologic history of sea water: an attempt to state the problem. *Bull. Geol. Soc. Am.* **62**, 1111–1148 (1951).
15. McDonough, W. F. in *Treatise on Geochemistry* (eds Holland, H. D. and Turrekian, K. K.) 547–568 (Elsevier, Amsterdam, 2004).
16. Mittlefehldt, D. W. Geochemistry of the ungrouped carbonaceous chondrite Tagish Lake, the anomalous CM chondrite Bells, and comparison with CI and CM chondrites. *Meteorit. Planet. Sci.* **37**, 703–712 (2002).
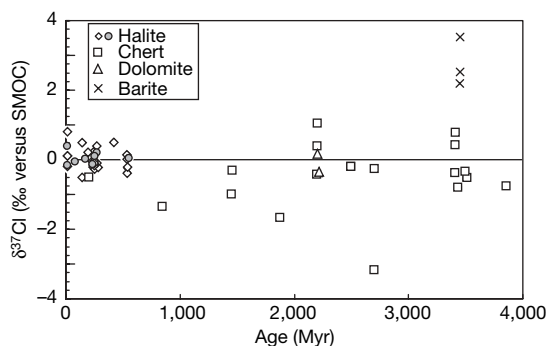
Figure 2 | Measured $\delta^{37}Cl$ values of crustal sediments plotted as a function of age. Phanerozoic halite data plotted as diamonds (from this study and ref. 6) and as circles (mean $\delta^{37}Cl$ values of halite-facies salt from ref. 29). Older sediments include chert, barite and dolomite. There is no correlation with age for either the halite or the chert data, although the latter have an average $\delta^{37}Cl$ value that is 0.7‰ less than the halite. The high, heterogeneous values for the barite may be related to diagenesis and exchange with a non-marine reservoir. Analytical uncertainties (1σ) are ±0.12‰ (1σ) for halite and ±0.3‰ (1σ) for all others.

17. Layne, G. D., Godon, A., Webster, J. D. & Bach, W. Secondary ion mass spectrometry for the determination of δ^{37}Cl: Part I. Ion microprobe analysis of glasses and fluids. *Chem. Geol.* **207,** 277–289 (2004).

18. Eastoe, C. J. & Peryt, T. Stable chlorine isotope evidence for non-marine chloride in Badenian evaporites, Carpathian mountain region. *Terra Nova* **11,** 118–123 (1999).

19. Renne, P. R., Sharp, W. D., Montañez, I. P., Becker, T. A. & Zierenberg, R. A. ^{40}Ar/^{39}Ar dating of Late Permian evaporites, southeastern New Mexico, USA. *Earth Planet. Sci. Lett.* **193,** 539–547 (2001).

20. Amthor, J. A. *et al.* Extinction of Cloudina and Namacalathus at the Precambrian-Cambrian boundary in Oman. *Geology* **31,** 431–434 (2003).

21. Lowe, D. R. & Knauth, L. P. Sedimentology of the Onverwacht Group (3.4 billion years),Transvaal, South Africa, and its bearing on the characteristics and evolution of the Early Earth. *J. Geol.* **85,** 695–723 (1977).

22. Perry, E. C., Hickman, A. H. & Barnes, I. L. Archean sedimentary barite, Pilbara goldfield, western Australia. *Geol. Soc. Am. Prog. Abstr.* **7,** 1226 (1975).

23. Okakangas, R. W., Morey, G. B. & Southwick, D. L. Paleoproterozoic basin development and sedimentation in the Lake Superior region, North America. *Sedim. Geol.* **141,** 319–341 (2001).

24. Sharp, Z. D., Durakiewicz, T., Migaszewski, Z. M. & Atudorei, V. N. Antiphase hydrogen and oxygen isotope periodicity in chert nodules. *Geochim. Cosmochim. Acta* **66,** 2865–2873 (2002).

25. Nutman, A. P., McGregor, V. R., Friend, C. R. L., Bennett, V. C. & Kinny, P. D. The Itsaq Gneiss Complex of southern West Greenland; the world's most extensive record of early crustal evolution (3900–3600 Ma). *Precambr. Res.* **78,** 1–39 (1996).

26. Moore, T. B. & Schopf, J. W. in *The Proterozoic Biosphere: A Multidisciplinary Study* (eds Schopf, J. W. & Klein, C.) 603–693 (Cambridge Univ. Press, Cambridge, 1992).

27. Burnard, P., Marty, B., Fischer, T. & Hilton, D. Noble gases in carbonatite magmatism: Oldonyo Lengai. *Eos* (Fall Meet. Suppl.), **87** (52), abstr. V23C–0645 (2006).

28. Eggenkamp, H. G. M. & Koster van Groos, A. F. Chlorine stable isotopes in carbonatites: evidence for isotopic heterogeneity in the mantle. *Chem. Geol.* **140,** 137–143 (1997).

29. Eastoe, C. J., Peryt, T. M., Petrychenko, O. Y. & Geisler-Cussey, D. Stable chlorine isotopes in Phanerozoic evaporites. *Appl. Geochem.* **22,** 575–588 (2007).

LETTERS

Weighing of biomolecules, single cells and single nanoparticles in fluid

Thomas P. Burg[1]*, Michel Godin[1]*, Scott M. Knudsen[1], Wenjiang Shen[3], Greg Carlson[3], John S. Foster[3], Ken Babcock[3,4] & Scott R. Manalis[1,2]

Nanomechanical resonators enable the measurement of mass with extraordinary sensitivity[1–7]. Previously, samples as light as 7 zepto-grams (1 zg = 10^{-21} g) have been weighed in vacuum, and proton-level resolution seems to be within reach[8]. Resolving small mass changes requires the resonator to be light and to ring at a very pure tone—that is, with a high quality factor[9]. In solution, viscosity severely degrades both of these characteristics, thus preventing many applications in nanotechnology and the life sciences where fluid is required[10]. Although the resonant structure can be designed to minimize viscous loss, resolution is still substantially degraded when compared to measurements made in air or vacuum[11–14]. An entirely different approach eliminates viscous damping by placing the solution inside a hollow resonator that is surrounded by vacuum[15,16]. Here we demonstrate that suspended microchannel resonators can weigh single nanoparticles, single bacterial cells and sub-monolayers of adsorbed proteins in water with sub-femtogram resolution (1 Hz bandwidth). Central to these results is our observation that viscous loss due to the fluid is negligible compared to the intrinsic damping of our silicon crystal resonator. The combination of the low resonator mass (100 ng) and high quality factor (15,000) enables an improvement in mass resolution of six orders of magnitude over a high-end commercial quartz crystal microbalance[17]. This gives access to intriguing applications, such as mass-based flow cytometry, the direct detection of pathogens, or the non-optical sizing and mass density measurement of colloidal particles.

The resonance frequency of a suspended microfluidic channel with micrometre-thin walls and a comparably thin fluid layer is highly sensitive to the presence of molecules or particles whose mass density differs from that of the solution (Fig. 1). This enables experiments that extend conceptually beyond the measurement of bulk fluid density, which has been demonstrated previously[18,19]. As the ratio of surface area to volume of our suspended microchannel is very large (10^4 cm^{-1}), surface adsorption is an effective mechanism for bio-molecular mass sensing (Fig. 1b)[15].

In another measurement mode (Fig. 1c), particles suspended in the solution flow through the resonator, and the resulting frequency shift depends on the mass and position of the particles. For dilute suspensions, this measurement yields a series of well separated peaks whose heights are directly proportional to the mass excess of each particle in solution.

Changes in mass inside the channel translate into shifts in the resonance frequency, f, according to[20]:

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m^* + \alpha\,\Delta m}} \tag{1}$$

The spring constant of the resonator is denoted by k, m^* is the
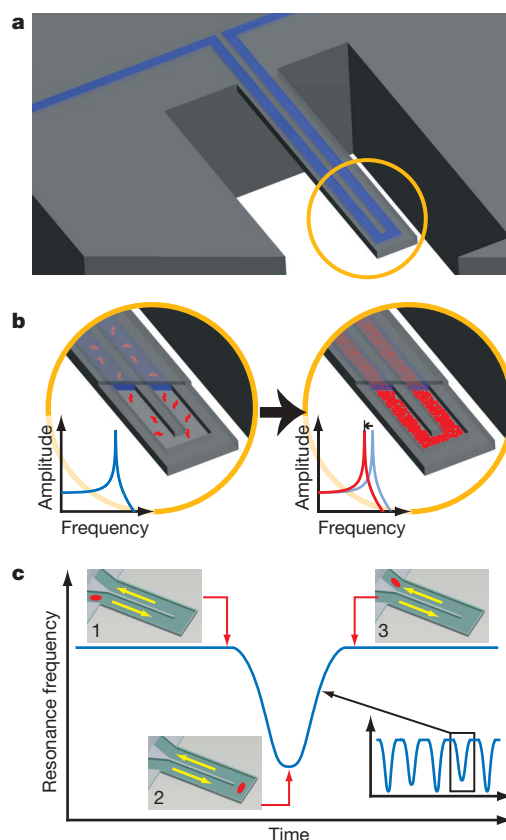
Figure 1 | Illustration of two mass measurement modes enabled by a fluid-filled microcantilever. a, A suspended microchannel translates mass changes into changes in resonance frequency. Fluid continuously flows through the channel and delivers biomolecules, cells or synthetic particles. Sub-femtogram mass resolution is attained by shrinking the wall and fluid layer thickness to the micrometre scale and by packaging the cantilever under high vacuum. **b,** While bound and unbound molecules both increase the mass of the channel, species that bind to the channel wall accumulate inside the device, and, as a result, their number can greatly exceed the number of free molecules in solution. This enables specific detection by way of immobilized receptors. **c,** In another measurement mode, particles flow through the cantilever without binding to the surface, and the observed signal depends on the position of particles along the channel (insets 1–3). The exact mass excess of a particle can be quantified by the peak frequency shift induced at the apex.

[1]Department of Biological Engineering, [2]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [3]Innovative Micro Technology, Santa Barbara, California 93117, USA. [4]Affinity Biosensors, Santa Barbara, California 93117, USA.
*These authors contributed equally to this work.

effective mass, and α is a numerical constant that depends on the geometric localization of the added mass Δm. For our cantilever-shaped resonators, α takes on a value of ~ 0.24 for changes in solution density and for uniform adsorption of molecules to the surface, and $\alpha \approx 1$ when a particle in transit is positioned at the apex.

Our fabrication process is based on the creation of buried channels in silicon-on-insulator wafers, followed by wafer thinning and dry etching to form suspended microchannels with walls $2–3\,\mu m$ thick and a $3\,\mu m$ fluid layer. Two hundred devices are fabricated and vacuum-packaged on a six-inch wafer, with yields exceeding 80%. A 'getter' layer prevents slow degradation of the on-chip vacuum due to outgassing. Integrated under each cantilever is an electrostatic drive electrode (Fig. 2a), and the cantilever vibration is detected optically. Further detail regarding the device fabrication is provided in Supplementary Information.

Dry resonators possess quality factors up to 15,000 and, as shown by the frequency response plot in Fig. 2b, no increase in damping occurs after filling the device with water. In comparison, conventional microresonators submersed in fluid have quality factors of up to ~ 150 (refs 11–14). Owing to the high quality factor, a gain controlled oscillator circuit with our device as the frequency determining element exhibits a frequency stability of 10 p.p.b. (parts per billion) in a 1 Hz bandwidth, even when the suspended microchannel is subjected to continuous flow. As the oscillator continuously tracks the resonance frequency, the dynamic range exceeds seven orders of

magnitude for short measurements (of the order of seconds with a 1 Hz bandwidth). Over extended periods of time, drift becomes the limiting factor, and at higher bandwidths frequency resolution is limited by phase noise.

Perhaps the most widely used mass-sensing method today is the quartz crystal microbalance, with a noise floor in solution of $\sim 1\,ng\,cm^{-2}$ and a total mass resolution of $\sim 1\,ng$ (ref. 17). Microcantilevers in solution have been used to resolve mass changes of $\sim 1\,\mu g\,cm^{-2}$ ($\sim 1\,ng$ total mass)[13] and other micromachined resonators optimized for fluid operation have achieved $10\,ng\,cm^{-2}$ (ref. 12). In the context of biomolecular detection, surface plasmon resonance (SPR) is widely regarded as the 'gold standard'. Although the detection in SPR is based on refractive index, the signal is approximately proportional to mass, and laboratory instruments attain a resolution equivalent to $\sim 0.05\,ng\,cm^{-2}$ ($\sim 1\,pg$ total mass)[21]. Using the suspended microchannel resonator, mass that is evenly adsorbed to the inner surfaces can be resolved to $0.01\,ng\,cm^{-2}$ (1 fg total). A unique aspect of our device is the ability to measure the mass of objects that are not attached to the surface. Sensitivity is highest near the cantilever tip, enabling the measurement of particle masses with a precision of 300 attograms ($1\,ag = 10^{-18}\,g$). If the frequency measurement is refined to approach the thermomechanical noise floor of our device, it should be possible to resolve mass changes as low as 1 ag, and shrinking the wall thickness and/or channel size to sub-micrometre dimensions could further improve sensitivity. A more detailed discussion of mass sensing with suspended microchannel resonators in relationship to other methods is provided in Supplementary Information.

To demonstrate the surface-based detection of proteins, we measured the binding of goat anti-mouse immunoglobulin-γ (IgG) molecules to anti-goat IgG antibodies that were immobilized on the channel walls (Fig. 3). Biotinylated antibodies were linked to the surface via a layer of Neutravidin bound to poly(ethyleneglycol)-biotin grafted poly-L-lysine (PLL-PEG-biotin). Before adsorbing the PLL-PEG-biotin, the surface was cleaned with a mixture of acetic acid and hydrogen peroxide (1:1). This mixture or, alternatively, sulphuric acid and hydrogen peroxide ('piranha solution') may be repeatedly injected into the all-glass and silicon device, thus enabling cleaning and multiple re-use of the chips.

The exact mass of the different layers can be quantified by the difference in resonance frequency before and after each injection, as shown in Fig. 3a. Directly before injections, we filled the inlet bypass with $14\,\mu l$ of sample at low pressure while the outlet bypass and the resonator were continuously rinsed with phosphate buffered saline (PBS) (Fig. 3a left inset). Pressure on the inlet bypass was subsequently turned on, forcing the sample through the resonator at a flow rate of $\sim 100\,nl\,min^{-1}$ (Fig. 3a right inset). This method of sample introduction ensures a sharp concentration profile and minimizes reagent consumption.

We injected goat anti-mouse IgG at concentrations ranging from $0.7\,nM$ ($100\,ng\,ml^{-1}$) to $0.7\,\mu M$ ($100\,\mu g\,ml^{-1}$) into the functionalized resonator, and observed the binding in real time by monitoring resonance frequency (Fig. 3b). All samples were prepared in PBS containing 0.01% Tween 20 and $1\,mg\,ml^{-1}$ bovine serum albumin (BSA) as a carrier protein. The signal caused by the solution density difference between the sample and the running buffer was subtracted off-line, as detailed in Supplementary Information.

Appropriate surface functionalization is crucial for biomolecular recognition with high specificity and sensitivity, and a number of specialized coatings that are compatible with silicon substrates have been developed for this purpose[22]. Assuming monolayer coverage (1–2 $pmol\,cm^{-2}$) of active antibodies with a dissociation constant of 1 nM, we expect an ultimate limit of detection of the order of $1\,pM$ for a 30 kDa analyte. In comparison, enzyme-linked immunosorbent assays (ELISA) have detection limits down to 0.1 pM. However, ELISA assays do not provide a real-time quantitative readout of antigen binding. An additional comparison can be made to microcantilever surface stress
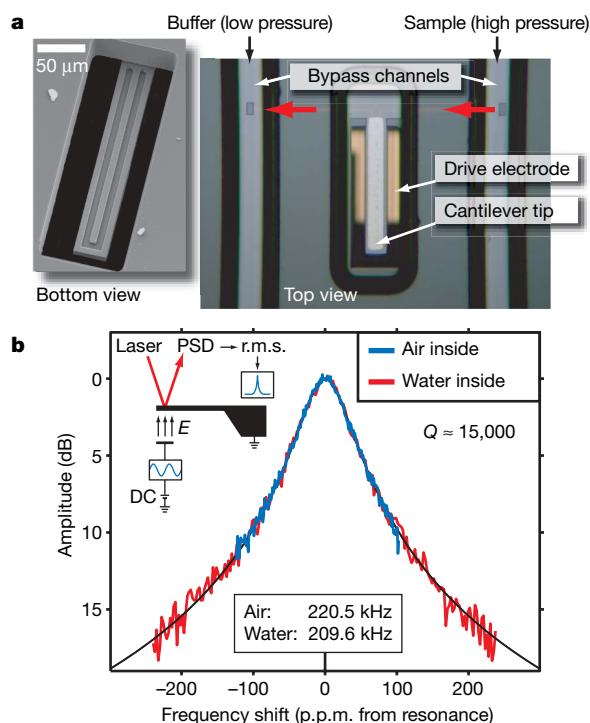
Figure 2 | Micrographs and frequency response of a suspended microchannel resonator. a, The $200 \times 33 \times 7\,\mu m$ (length $\times$ width $\times$ thickness) microcantilever containing a $3 \times 8\,\mu m$ (height $\times$ width) channel is suspended in a vacuum cavity (optical micrograph, right). Microfluidic bypass channels ($30 \times 100\,\mu m$, height $\times$ width) are connected to the inlet and the outlet of the suspended channel, and enable the quick exchange of samples by pressure driven flow (red arrows). The electron micrograph (left) shows a bottom view of a cantilever that has been intentionally etched open to visualize the fluidic conduit inside. **b,** Frequency response plots of a cantilever before (blue) and after (red) filling with water reveal different resonance frequencies but indistinguishable quality factors. To measure the frequency response, we monitored the vibration amplitude with a laser and a position sensitive photodetector (PSD) while the cantilever was being driven electrostatically at different frequencies (inset; E denotes the electric field, and 'DC' represents a bias voltage of $\sim 60\,V$).

measurements, which have achieved detection limits of 1 pM to 1 nM for protein binding[23,24]. However, owing to the complex dependence of surface stress on molecular binding, sensitivity is highly assay dependent.

The limit of detection by mass could be improved through mass enhancement by gold labelled secondary antibodies. For instance, the binding of a single 100 nm Au nanoparticle near the apex of the cantilever can clearly be resolved. Although the advantage of label-free detection is lost in this approach, mass-labelling would enable non-optical detection in a continuous-flow microfluidic device with femtomolar sensitivity. In addition, differential sensing and temperature control can be implemented to improve the detection of small signals in the presence of thermal drift, non-specific binding, or bulk density differences.

In applications where large fluid volumes need to be processed, the utility of our approach is limited by the high flow resistance of the channel (the flow rate at 10 bar is ~1 μl min^{-1}). On the other hand, the suspended microchannel resonator enables the analysis of precious samples with minimal reagent consumption: at a flow rate of 100 nl min^{-1}, the 10 pl detector volume is swept more than 100 times per second while the large surface-to-volume ratio of the thin channel provides for a high capture efficiency.

To illustrate the flow-through detection scheme, we measured the mass of bacteria and synthetic micro- and nanoparticles in transit through the suspended microchannel. Carefully matching the pressure at the inlet and the outlet of the resonator enabled us to reduce the flow rate to less than 10 pl s^{-1}. A low flow rate enables higher-resolution frequency measurements by increasing the transit time of the particle through the device. Several hundred particles can be weighed individually in a few minutes, producing a histogram of particle masses (Fig. 4a). One element of uncertainty is the exact lateral position of the particles at the time they pass the cantilever tip: particles travelling along the outer sidewall of the 8-μm-wide channel induce an 8% larger frequency shift than particles following the centreline of the channel. This variability could be reduced to below 1% by hydrodynamic focusing or by using a different resonator design, thereby enabling precise weighing—or, if the density is known, sizing—of sub-micrometre particles. In cases where the
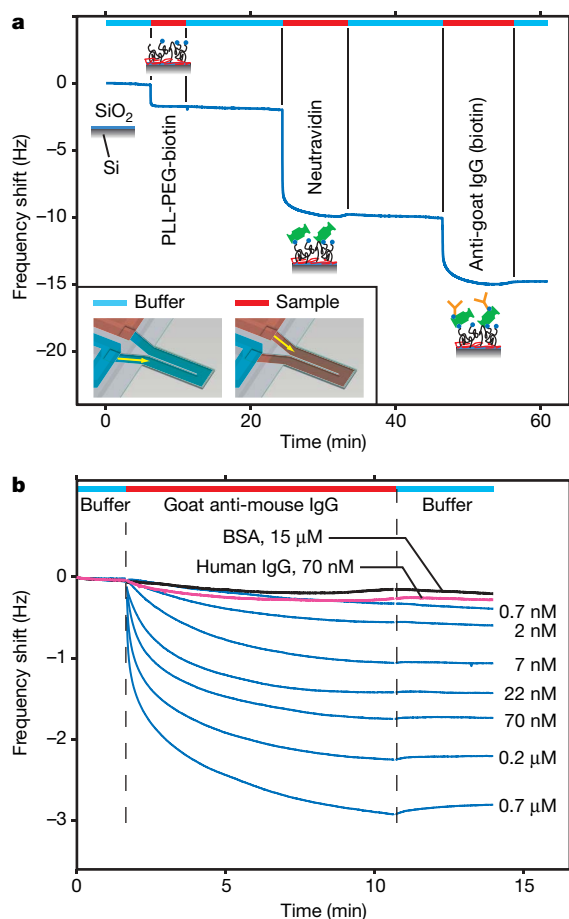
Figure 3 | Resonance frequency shifts caused by accumulation of proteins inside the cantilever. a, Antibodies to goat IgG are immobilized on the native silicon dioxide surface of the suspended microchannel resonator in three steps, as illustrated by the pictograms along the frequency trace: electrostatic adsorption of poly(ethyleneglycol)-biotin grafted poly-L-lysine (PLL-PEG-biotin, 1 mg ml^{-1}), binding of Neutravidin (0.5 mg ml^{-1}), and attachment of biotinylated antibodies (0.5 mg ml^{-1}) to the Neutravidin. The mass increase at each step can be followed in real time (1 Hz ≈ 300 fmol cm^{-2} for 150 kDa IgG molecules). One bypass channel was continuously rinsed with phosphate buffered saline (PBS) while the other bypass contained the sample. During the injection interval (red), the bypass holding the sample was pressurized, and during rinse cycles (blue), the pressure difference was reversed, as shown in the inset at the bottom of the panel. **b,** Goat anti-mouse IgG was injected at concentrations from 0.7 nM to 0.7 μM (blue traces). Between measurements, the surface was regenerated by injecting 200 mM glycine (HCl; pH 2.5), which dissociated the analyte while preserving the activity of the antibodies. Control injections with no IgG (black) or human IgG (red) showed very low levels of non-specific binding.
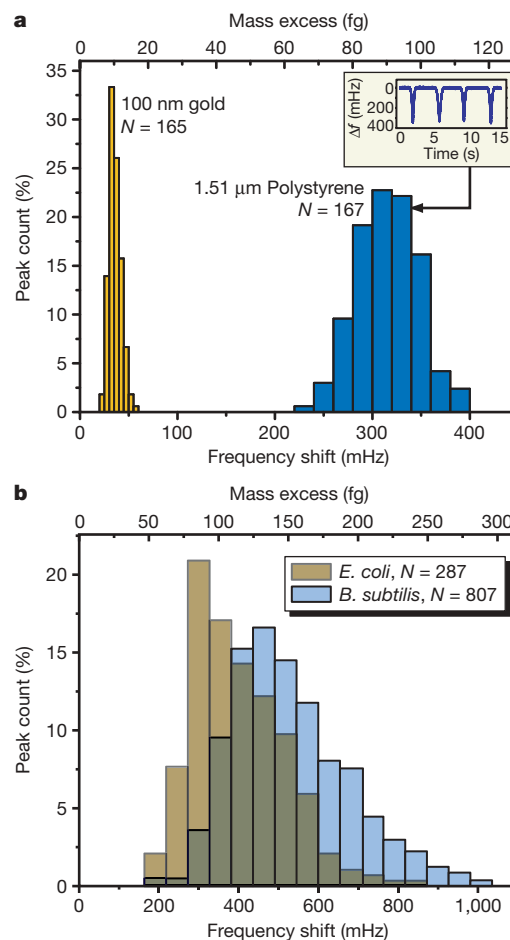
Figure 4 | Histograms of peak frequency shifts caused by particles and bacteria flowing through the resonator. a, Synthetic particles of known size and density were measured to calibrate the mass sensitivity of the device. Gold nanoparticles (100 ± 8 nm) weighing 10 fg more than the water they displace produced a mean frequency shift of 36 mHz with a standard deviation of 6 mHz. On a different device, we measured a frequency shift of 310 ± 30 mHz for polystyrene microspheres (1.51 ± 0.01 μm) with 90.1 fg mass excess. **b,** The masses of *E. coli* and *B. subtilis* in PBS were measured by passing the bacteria through the resonator and collecting peak height histograms.

density is unknown, this value may be obtained by conducting the measurement in two different carrier solutions.

An interesting application enabled by flow-through resonant mass sensing is the weighing of individual live bacteria. The mass histograms depicted in Fig. 4b reveal the distribution of masses within two bacterial cultures. The mass of *Escherichia coli* was measured to be 110 ± 30 fg in excess of the displaced buffer, while the mass of *Bacillus subtilis* is 150 ± 40 fg. This is consistent with the distinct sizes of *E. coli* ($0.8 \times 2\,\mu m$ rod) and *B. subtilis* ($1 \times 2.5\,\mu m$ rod) and reported mass density values for bacteria. The distribution asymmetry is consistent with previously reported bacterial size distributions, and is related to the bacterial growth and division process[25].

Transient particle measurements require no chemical modification of the resonator and are not affected by sensor drift or bulk density changes. Furthermore, a large number of assays can be run on the same device without requiring surface regeneration. Nanoparticle–biomolecule conjugates enable a variety of interesting protocols for highly specific and extremely sensitive detection of molecules, viruses and cells[26,27]. Mass sensing with resonating microchannels provides an attractive readout method for such colloidal biochemical assays, owing to its ability to measure single particle masses with high resolution using inexpensive instrumentation. At present, the low flow rate of $10–100$ pl s^{-1} that is required for femtogram resolution limits throughput to $1–10$ particles per second. However, using arrays of suspended microchannels could substantially increase throughput by measuring particles in parallel. Another intriguing possibility is the integration of such arrays with microfluidic systems made from elastomers, which can provide pumps and valves to deliver fluids to the detector with great precision[28,29].

The size of our suspended channels is appropriate for the characterization of nanoparticles and bacteria, but too small for the passage of most eukaryotic cells. Future device generations will address this limitation, and we envisage that flow-through resonant mass sensing could ultimately be configured for counting specific cells in a fashion similar to flow cytometry. For example, CD4 cells, which are routinely counted to monitor AIDS, could be labelled with nanoparticles functionalized with CD4-specific antibodies to distinguish them from other cells present in a blood sample. This method for CD4 counting could be a cheap solution in developing countries where flow cytometry is not available or too expensive[30]. By the same principle, mass labelling would enable the specific detection of pathogens such as various bacteria, viruses and toxins.

1. Thundat, T., Wachter, E. A., Sharp, S. L. & Warmack, R. J. Detection of mercury vapor using resonating microcantilevers. *Appl. Phys. Lett.* **66,** 1695–1697 (1995).
2. Lange, D., Hagleitner, C., Hierlemann, A., Brand, O. & Baltes, H. Complementary metal oxide semiconductor cantilever arrays on a single chip: Mass-sensitive detection of volatile organic compounds. *Anal. Chem.* **74,** 3084–3095 (2002).
3. Lavrik, N. V. & Datskos, P. G. Femtogram mass detection using photothermally actuated nanomechanical resonators. *Appl. Phys. Lett.* **82,** 2697–2699 (2003).
4. Ono, T., Li, X., Miyashita, H. & Esashi, M. Mass sensing of adsorbed molecules in sub-picogram sample with ultrathin silicon resonator. *Rev. Sci. Instrum.* **74,** 1240–1243 (2003).
5. Gupta, A., Akin, D. & Bashir, R. Single virus particle mass detection using microresonators with nanoscale thickness. *Appl. Phys. Lett.* **84,** 1976–1978 (2004).
6. Ilic, B. *et al.* Attogram detection using nanoelectromechanical oscillators. *J. Appl. Phys.* **95,** 3694–3703 (2004).
7. Forsen, E. *et al.* Ultrasensitive mass sensor fully integrated with complementary metal-oxide-semiconductor circuitry. *Appl. Phys. Lett.* **87,** 043507 (2005).
8. Yang, Y. T., Callegari, C., Feng, X. L., Ekinci, K. L. & Roukes, M. L. Zeptogram-scale nanomechanical mass sensing. *Nano Lett.* **6,** 583–586 (2006).
9. Ekinci, K. L., Yang, Y. T. & Roukes, M. L. Ultimate limits to inertial mass sensing based upon nanoelectromechanical systems. *J. Appl. Phys.* **95,** 2682–2689 (2004).
10. Lucklum, R. & Hauptmann, P. Acoustic microsensors—the challenge behind microgravimetry. *Anal. Bioanal. Chem.* **384,** 667–682 (2006).
11. Weinberg, M. S., Dube, C. E., Petrovich, A. & Zapata, A. M. Fluid damping in resonant flexural plate wave device. *J. Microelectromech. Syst.* **12,** 567–576 (2003).
12. Zhang, H. & Kim, E. S. Micromachined acoustic resonant mass sensor. *J. Microelectromech. Syst.* **14,** 699–706 (2005).
13. Braun, T. *et al.* Micromechanical mass sensors for biomolecular detection in a physiological environment. *Phys. Rev. E* **72,** 031907 (2005).
14. Pang, W. *et al.* Femtogram mass sensing platform based on lateral extensional mode piezoelectric resonator. *Appl. Phys. Lett.* **88,** 243503 (2006).
15. Burg, T. P. & Manalis, S. R. Suspended microchannel resonators for biomolecular detection. *Appl. Phys. Lett.* **83,** 2698–2700 (2003).
16. Burg, T. P. *et al.* Vacuum-packaged suspended microchannel resonant mass sensor for biomolecular detection. *J. Microelectromech. Syst.* **15,** 1466–1476 (2006).
17. Q-Sense. Model D3000 specifications. ⟨http://www.q-sense.com/viewArticle.asp?ID=31⟩.
18. Enoksson, P., Stemme, G. & Stemme, E. Silicon tube structures for a fluid-density sensor. *Sens. Actuators A* **54,** 558–562 (1996).
19. Westberg, D., Paul, O., Andersson, G. & Baltes, H. A CMOS-compatible fluid density sensor. *J. Micromech. Microeng.* **7,** 253–255 (1997).
20. Sarid, D. *Scanning Force Microscopy: With Applications to Electric, Magnetic, and Atomic Forces* (Oxford Univ. Press, USA, 1994).
21. Myszka, D. G. Analysis of small-molecule interactions using Biacore S51 technology. *Anal. Biochem.* **329,** 316–323 (2004).
22. Angenendt, P., Glokler, J., Sobek, J., Lehrach, H. & Cahill, D. J. Next generation of protein microarray support materials: Evaluation for protein and antibody microarray applications. *J. Chromatogr. A* **1009,** 97–104 (2003).
23. Wu, G. H. *et al.* Bioassay of prostate-specific antigen (PSA) using microcantilevers. *Nature Biotechnol.* **19,** 856–860 (2001).
24. Backmann, N. *et al.* A label-free immunosensor array using single-chain antibody fragments. *Proc. Natl Acad. Sci. USA* **102,** 14587–14592 (2005).
25. Akerlund, T., Nordstrom, K. & Bernander, R. Analysis of cell size and DNA content in exponentially growing and stationary-phase batch cultures of *Escherichia coli. J. Bacteriol.* **177,** 6791–6797 (1995).
26. Nam, J. M., Thaxton, C. S. & Mirkin, C. A. Nanoparticle-based bio-bar codes for the ultrasensitive detection of proteins. *Science* **301,** 1884–1886 (2003).
27. Agrawal, A., Zhang, C. Y., Byassee, T., Tripp, R. A. & Nie, S. M. Counting single native biomolecules and intact viruses with color-coded nanoparticles. *Anal. Chem.* **78,** 1061–1070 (2006).
28. Thorsen, T., Maerkl, S. J. & Quake, S. R. Microfluidic large-scale integration. *Science* **298,** 580–584 (2002).
29. Whitesides, G. M. The origins and the future of microfluidics. *Nature* **442,** 368–373 (2006).
30. Yager, P. *et al.* Microfluidic diagnostic technologies for global public health. *Nature* **442,** 412–418 (2006).

LETTERS

Effect of natural iron fertilization on carbon sequestration in the Southern Ocean

Stéphane Blain[1], Bernard Quéguiner[1], Leanne Armand[1], Sauveur Belviso[2], Bruno Bombled[2], Laurent Bopp[2], Andrew Bowie[3,4], Christian Brunet[5], Corina Brussaard[6], François Carlotti[1], Urania Christaki[7], Antoine Corbière[5], Isabelle Durand[8], Frederike Ebersbach[3], Jean-Luc Fuda[9], Nicole Garcia[1], Loes Gerringa[6], Brian Griffiths[10], Catherine Guigue[11], Christophe Guillerm[12], Stéphanie Jacquet[13], Catherine Jeandel[14], Patrick Laan[6], Dominique Lefèvre[11], Claire Lo Monaco[5], Andrea Malits[15], Julie Mosseri[1], Ingrid Obernosterer[16], Young-Hyang Park[8], Marc Picheral[15], Philippe Pondaven[17], Thomas Remenyi[3], Valérie Sandroni[1], Géraldine Sarthou[17], Nicolas Savoye[13,18], Lionel Scouarnec[12], Marc Souhaut[14], Doris Thuiller[5], Klaas Timmermans[6], Thomas Trull[3,10], Julia Uitz[15], Pieter van Beek[14], Marcel Veldhuis[6], Dorothée Vincent[7], Eric Viollier[19], Lilita Vong[1] & Thibaut Wagener[15]

The availability of iron limits primary productivity and the associated uptake of carbon over large areas of the ocean. Iron thus plays an important role in the carbon cycle, and changes in its supply to the surface ocean may have had a significant effect on atmospheric carbon dioxide concentrations over glacial–interglacial cycles[1–5]. To date, the role of iron in carbon cycling has largely been assessed using short-term iron-addition experiments[6,7]. It is difficult, however, to reliably assess the magnitude of carbon export to the ocean interior using such methods, and the short observational periods preclude extrapolation of the results to longer timescales[8]. Here we report observations of a phytoplankton bloom induced by natural iron fertilization—an approach that offers the opportunity to overcome some of the limitations of short-term experiments. We found that a large phytoplankton bloom over the Kerguelen plateau in the Southern Ocean was sustained by the supply of iron and major nutrients to surface waters from iron-rich deep water below. The efficiency of fertilization, defined as the ratio of the carbon export to the amount of iron supplied, was at least ten times higher than previous estimates from short-term blooms induced by iron-addition experiments[7]. This result sheds new light on the effect of long-term fertilization by iron and macronutrients on carbon sequestration, suggesting that changes in iron supply from below—as invoked in some palaeoclimatic[9,10] and future climate change scenarios[11]—may have a more significant effect on atmospheric carbon dioxide concentrations than previously thought.

The Southern Ocean plays a major role in the climate system, and is recognized as the oceanic body most sensitive to climate change[12,13]. Iron fertilization of its surface waters during glacial times by enhanced dust deposition is a scenario (known as the 'iron hypothesis'[1]) proposed to explain lower atmospheric CO_2 during colder climates. Different versions of this scenario[2,3], as well as the magnitude of the CO_2 drawdown induced by iron fertilization[4,5], are still debated, but the important role of iron in carbon cycling is no longer in question. The proximate control of biological productivity by iron in the Southern Ocean has been unequivocally demonstrated by iron addition experiments[14–16]. All experiments show an enhancement of primary production associated with a shift in the structure of the phytoplankton community[7]. However, the short observational periods, as well as other intrinsic limits and artefacts of the small scale fertilization technique, have prevented a clear assessment of carbon export and preclude extrapolation to longer timescales.

Although the Southern Ocean is the largest high-nutrient low-chlorophyll (HNLC) region of the global ocean, natural phytoplankton blooms do occur in the vicinity of many Southern Ocean islands, as first reported in ref. 17 and subsequently characterized in detail by satellite images[18,19]. The largest bloom is observed around Kerguelen Island and the adjacent plateau to its southeast (Fig. 1a, b). The bloom has two main features: (1) a narrow plume that extends northeast of the island and north of the Polar Front that shows high mesoscale and temporal variability[20], and (2) a larger bloom ($\sim$45,000 km^2) southeast of the island and south of the Polar Front which is remarkably constrained to the bathymetry of the plateau. In 2004–05, this larger bloom began in early November, achieved high phytoplankton biomass ($\sim$3 µg chlorophyll a per litre) in December and January and then collapsed in late February (Fig. 1c).

The KEOPS (Kerguelen ocean and plateau compared study; 19 January to 13 February 2005) cruise carried out a survey inside and outside this bloom (Fig. 1b), with particular focus on two contrasting stations: A3 (50° 38′ S, 72° 05′ E) and C11 (51° 39′ S, 78° 00′ E).

[1]Laboratoire d'Océanographie et de Biogéochimie, Centre Océanologique de Marseille, CNRS, Université de la Méditerranée, campus de Luminy, case 901, 13288 Marseille Cedex 09, France. [2]IPSL/Laboratoire des Sciences du Climat et de l'Environnement, CEN de Saclay, Bât. 701 l'Orme des Merisiers, 91191 Gif-sur-Yvette, France. [3]Antarctic Climate and Ecosystems CRC, Hobart, Tasmania 7001, Australia. [4]ACROSS, School of Chemistry, University of Tasmania, Hobart, Tasmania 7001, Australia. [5]LOCEAN-IPSL, UMR 7159, CNRS, Université P. et M. Curie, Case 100, 4 place Jussieu, 75252 Paris Cedex 5, France. [6]Royal Netherlands Institute for Sea Research (NIOZ), PO Box 59, 1790 AB Den Burg, Texel, The Netherlands. [7]FRE ELICO, Université du Littoral Côte d'Opale, Maison de la Recherche en Environnement Naturel (MREN), 32 avenue Foch, 62930 Wimereux, France. [8]USM402/LOCEAN, Département des Milieux et Peuplements Marins, Muséum National d'Histoire Naturelle, 43 rue Cuvier, F-75231 Paris Cedex 05, France. [9]Centre Océanologique de Marseille, Campus de Luminy, 13288 Marseille Cedex 09, France. [10]CSIRO Division of Marine and Atmospheric Research, GPO Box 1538, Hobart, Tasmania 7001, Australia. [11]LMGEM UMR CNRS 617, Campus de Luminy, case 901, 13288 Marseille Cedex 09, France. [12]DT INSU CNRS, Bât. IPEV BP 74, Technopole Brest Iroise, 29280 Plouzané, France. [13]Department of Analytical and Environmental Chemistry, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium. [14]LEGOS (CNRS/CNES/IRD/UPS), Observatoire Midi-Pyrénées, 14 avenue Edouard Belin, 31400 Toulouse, France. [15]Laboratoire d'Océanographie de Villefranche, Quai de La Darse, BP 8, 06238 Villefranche-sur-Mer, France. [16]Université Pierre et Marie Curie-Paris 6, UMR7621, CNRS, F66650 Banyuls-sur-mer, France. [17]UMR 6539/LEMAR/IUEM, Technopole Brest Iroise, Place Nicolas Copernic, 29280 Plouzané, France. [18]Observatoire Aquitain des Sciences de l'Univers, UMR CNRS 5805 EPOC, Station Marine d'Arcachon, 2 rue du Pr. Jolyet, 33120 Arcachon, France. [19]Laboratoire de Géochimie des Eaux, UMP IPGP 7154, Université Denis Diderot Paris 7, 2 place Jussieu, 75251 Paris Cedex 05, France.

Measurements of the partial pressure of CO_2 (p_{CO_2}) in surface waters indicate that the bloom was an important sink for CO_2 (Fig. 1d), with a mean p_{CO_2} drawdown of $58 \pm 11 \mu atm$, which was ~2–3 times higher than that observed following Southern Ocean iron addition experiments, probably reflecting the longer duration of the Kerguelen bloom. Silicic acid was almost depleted (1–2 μM), while nitrate concentrations remained relatively high (23 μM) (Supplementary Fig. 1).

Measurements of dissolved iron (DFe) revealed that concentrations in the surface mixed layer were low on and off the plateau (0.090 ± 0.034 nM; mean $\pm$ s.d., $n = 49$) and typical of surface waters of the open Southern Ocean[15,21]. But vertical profiles of DFe revealed a major difference between on- and off-plateau stations below 150 m (Fig. 2b). Above the plateau, DFe concentrations increased with depth and reached a mean maximum of ~0.35 nM (range 0.19–0.51 nM) at 500 m, close to the bottom. In contrast, off the plateau (bottom > 1,500 m) mean DFe concentrations were ~0.18 nM at 600 m. Organic ligand concentrations (L) were sufficient to keep the supplied DFe soluble ($L/DFe > 2.4$ for all stations). The low concentrations of DFe measured in surface waters both on and off the plateau are known to be close to rate-limiting for many phytoplankton from bottle[22] and mesoscale iron addition experiments[14–16]. Thus it appears that phytoplankton production depletes the iron to rate-limiting levels both off and on the plateau, with the higher biomass over the plateau induced by higher iron supply from below.

Availability of iron in surface waters over the plateau appears to be enhanced by the greater winter stock and by a higher ongoing supply from increased vertical mixing and the steeper DFe gradient. Between 100 and 200 m, the vertical diffusivity ($K_z = (3.3 \pm 3.3) \times 10^{-4} \, m^2 \, s^{-1}$, see Table 1) was substantially higher than the most recent estimate of

K_z in the open Southern Ocean ($(0.11 \pm 0.2) \times 10^{-4} \, m^2 \, s^{-1}$)[23]. This enhancement of vertical mixing was probably due to internal wave activity (Fig. 2c, d), which also appears to enhance mixing at the C11 site off the plateau ($K_z = (3.2 \pm 2.3) \times 10^{-4} \, m^2 \, s^{-1}$, see Table 1). Combining the K_z estimates with the DFe gradients (Table 1) suggests a nearly eightfold higher DFe flux into surface waters on the plateau ($31 \, nmol \, m^{-2} \, d^{-1}$) than off the plateau ($4 \, nmol \, m^{-2} \, d^{-1}$). For comparison, DFe fluxes around $3 \, nmol \, m^{-2} \, d^{-1}$ were found at the SOIREE and FeCycle HNLC sites[23,24].

The net DFe requirements of phytoplankton ($208 \pm 77 \, nmol \, m^{-2} \, d^{-1}$)—defined as the difference between the total uptake and regeneration rates of DFe, and estimated from our ^{55}Fe experiments carried out at the beginning (19 January) and at the end (12 February) of the survey—are not balanced by the vertical supply of DFe ($31 \, nmol \, m^{-2} \, d^{-1}$). This suggests an additional supply from, for example, ongoing depletion of the winter surface stock or from dissolution of lithogenic particulate Fe (PFe_{litho}). The total supply cannot exceed the demand, because this would lead to elevated iron concentrations in surface waters that would be in contradiction with our observations. Between 19 January and 12 February, the steady depletion of a winter stock by 0.06 nM or the application of a daily dissolution rate of ~2.5% of the observed PFe_{litho} at A3 on 24 January (~0.1 nM based on particulate Al and a crustal Fe/Al ratio) is sufficient to balance the demand. The accurate determination of these fluxes is out of reach of the present analytical capabilities, but the values are within realistic ranges[11] and are upper limits because the additional supply is most probably a combination of both processes. Constraining the balance of the demand with supply suggests an additional DFe input of $177 \, nmol \, m^{-2} \, d^{-1}$ at A3 (Table 1). Including this possible supply suggests an excess of DFe supply over
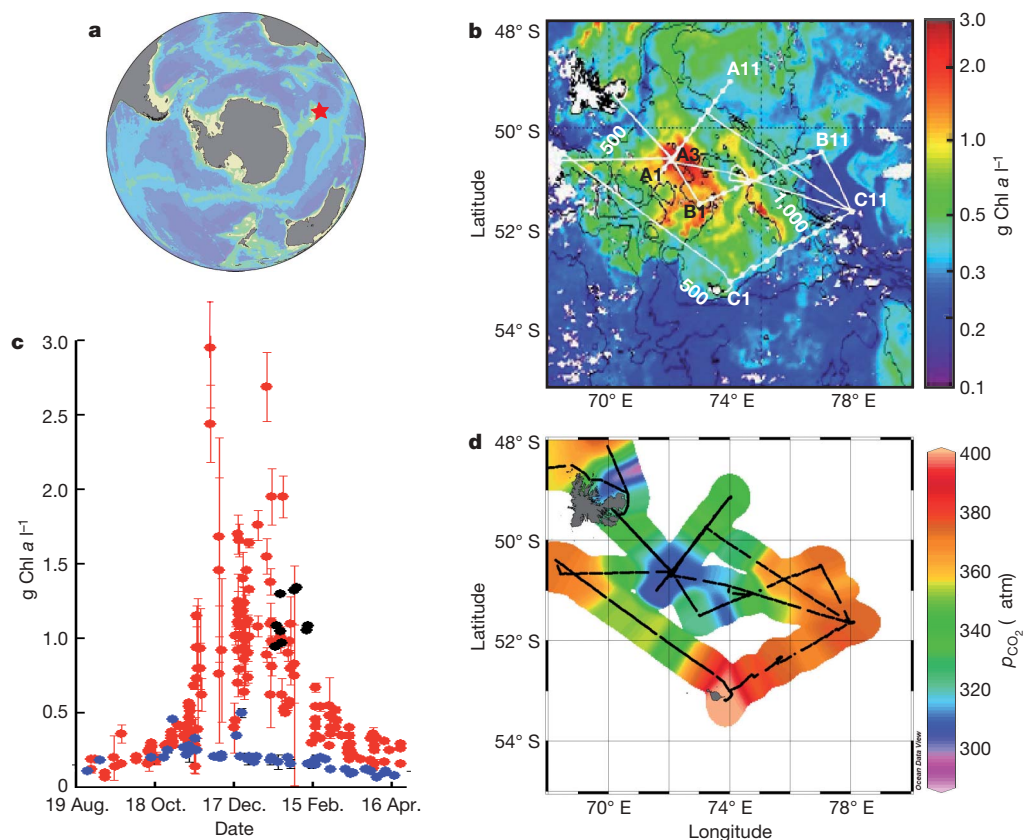
Figure 1 | Location of the study sites, temporal evolution of the bloom and surface water properties. a, Location of the Kerguelen plateau (red star) in the Southern Ocean. **b,** Satellite image of the bloom during the KEOPS cruise. The track of the cruise (white line), the position of the stations (white dots) and bathymetry (black lines) are shown. Chl *a*, chlorophyll *a*. **c,** Time series of satellite-determined Chl *a* at stations within the bloom (red dots) and at C11 (blue dots). (MODIS results provided by CSIRO Marine Research, Hobart.) Error bars, ± 1 s.d. calculated from the individual passes of the satellite. Black dots denote high-performance liquid chromatography measurements in surface waters during the cruise. **d,** Surface p_{CO_2} measured by underway sampling.
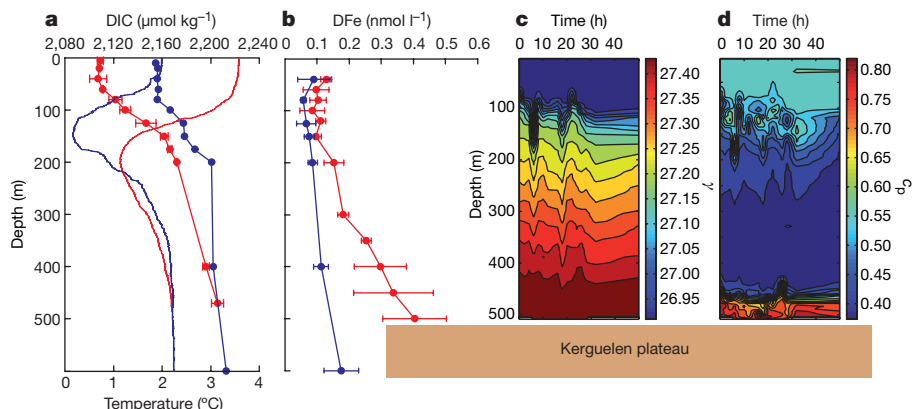
Figure 2 | Iron fertilization above the plateau. a, Mean profiles ($n = 3$) of DIC concentrations at A3 (red circles) and at C11 (blue circles). Red and blue lines denote typical temperature profiles at A3 and C11, respectively. Error bars, ± 1 s.d. **b,** Mean profiles ($n = 3$) of DFe concentrations at A3 (red circles) and outside the plateau (blue circles). Stations outside the plateau are C11, C9, B11, B9, A11 and A9. Error bars, ± 1 s.d. Above the Kerguelen plateau, close to the bottom, a large variability of the vertical DFe gradients among the different profiles was observed, resulting in the large s.d. of the mean DFe value. **c,** Time series of vertical profiles of γ at station A3 (γ = water density $- 10^3$, where water density is in $kg\, m^{-3}$). **d,** Time series of vertical profiles of particle attenuation at 660 nm (c_p) at station A3. A cast was performed every two hours. The brown block labelled Kerguelen plateau represents the bottom at station A3.

the plateau, in comparison to surrounding waters, of $204 \pm 77\, nmol\, m^{-2}\, d^{-1}$. This is a maximum estimate of the excess DFe supply, in that it assumes no additional supply to balance the demand at C11 as the demand was not determined at that location (Table 1).

Carbon dioxide uptake fuelled by the enhanced iron supply is clear from the lowered p_{CO_2} of surface waters (Fig. 1d), but how much carbon is exported from surface waters and sequestered at depth? We estimated carbon export using the deficit of ^{234}Th activity in the water column, and the ratio of particulate organic carbon (POC) to ^{234}Th (Fig. 3; Table 1).

The POC export indicated by the ^{234}Th technique in the bloom over the plateau was approximately twice as large as in the HNLC waters off the plateau (Table 1), but still within the range of POC fluxes from the mixed layer estimated in the same way for blooms in the Polar Frontal Zone[25]. At A3, the POC export flux represented about 30% of primary production. The mean excess of carbon export at A3 compared to C11 was 10.8 ± 4.9 and $14.2 \pm 7.9\, mmol\, m^{-2}\, d^{-1}$ at 100 and 200 m, respectively. This can be compared to the results of SOFEX, the only Southern Ocean iron addition experiment that has measured enhanced POC export[26]. The excess measured at 100 m was $7.1\, mmol\, m^{-2}\, d^{-1}$, and extrapolated to 250 m was $3\, mmol\, m^{-2}\, d^{-1}$.

On the basis of the excess of POC export at 200 m and the excess of DFe supply, we obtain a sequestration efficiency (that is, the excess of POC export divided by the excess of DFe supply) of $70{,}000 \pm$

Table 1 | DFe and carbon budget

Station	DFe		Carbon	
	Bloom (A3)	HNLC (C11)	Bloom (A3)	HNLC (C11)
Short-term fluxes*				
Vertical diffusivity ($10^{-4}\, m^2\, s^{-1}$)†	3.2	2.4	3.0	3.8
Vertical gradient ($mmol\, m^{-4}$)‡	11.2×10^{-7}	2.0×10^{-7}	0.56	0.32
Vertical supply ($mmol\, m^{-2}\, d^{-1}$)§	3.1×10^{-5}	0.4×10^{-5}	14.6	10.6
Unaccounted supply ($mmol\, m^{-2}\, d^{-1}$)	17.7×10^{-5} ‖	ND‖	28¶	−2.7¶
Th-derived POC export ($mmol\, m^{-2}\, d^{-1}$) at 100 m#			23.0	12.2
Th-derived POC export ($mmol\, m^{-2}\, d^{-1}$) at 200 m#			24.5	10.3
Seasonal budget*				
Winter concentration ($mmol\, m^{-3}$)⊗	1.53×10^{-4}	8.6×10^{-5}	2,230	2,237
Summer concentration ($mmol\, m^{-3}$)**	8.6×10^{-5}	7.2×10^{-5}	2,167	2,214
Winter stock utilization ($mmol\, m^{-2}$)††	4.7×10^{-3}	9.5×10^{-4}	4,410	1,564
Vertical supply ($mmol\, m^{-2}$)‡‡	1.4×10^{-3}	1.9×10^{-4}	657	477
Air–sea supply ($mmol\, m^{-2}$)‡‡			1,260	−121
POC accumulation in the ML ($mmol\, m^{-2}$)§§			670	190
DOC accumulation in the ML ($mmol\, m^{-2}$)§§			610	ND
POC export below the ML ($mmol\, m^{-2}$)‖‖			5,047	1,730

HNLC, high-nutrient low-chlorophyll; ND, not determined; POC, particulate organic carbon; DOC, dissolved organic carbon; ML, mixed layer.
* See also Supplementary Table 1.
† Mean values corresponding to the depth stratum where the gradient is observed (150–200 m for DFe, and 80–150 m for DIC).
‡ Vertical gradients were determined using the linear part of the vertical profiles (Fig. 2a, b) corresponding to the depth stratum defined in the previous footnote.
§ Vertical supply was calculated by multiplying the vertical gradient with the mean vertical diffusivity coefficient converted to $m^2\, d^{-1}$.
‖ This term is the supply of DFe needed to balance the mean net DFe demand ($208\, nmol\, m^{-2}\, d^{-1}$) calculated as the difference between the total uptake rate ($423\, nmol\, m^{-2}\, d^{-1}$) and the regeneration rate ($215\, nmol\, m^{-2}\, d^{-1}$) measured during the cruise. This term was not determined at C11. See text for details.
¶ This term is the mean gas exchange flux calculated from two different parameterizations[31,32] and the wind speed measured during the cruise.
\# See text and Methods for details.
⊗ Winter concentration (C_{winter}) was taken from the concentration measured at the depth of the temperature minimum characterizing the remnant winter water (Fig. 2a).
** Summer concentration (C_{summer}) was the mean value measured within the mixed layer during the KEOPS survey.
†† Seasonal apparent consumption was calculated using the equation ($C_{winter} - C_{summer}$)MLD. The mean mixed layer depth (MLD) was $70 \pm 20\, m$ and $68 \pm 13\, m$ at A3 and C11, respectively.
‡‡ Vertical supply and air–sea supply were calculated assuming linear build-up of the gradients over 90 days.
§§ The accumulation of POC and DOC was the difference between the stock in summer and in winter. (no DOC data are available at C11).
‖‖ POC export = DIC winter stock utilization + DIC vertical supply + DIC air–sea supply − POC accumulation − DOC accumulation. In a similar way to the POC export derived from Th fluxes, we assume that the carbon export at 200 m was roughly the same as just below the MLD.

46,000 mol mol^{-1} above the Kerguelen plateau. To take into account the possible variability in the C export and DFe supply as the season progresses, we also calculated the sequestration efficiency of the natural fertilization based on the seasonal budgets (Table 1). This yields a much higher value of 668,000 mol mol^{-1} (the ratios between the C export and the DFe supply were 9.9×10^5 and 1.5×10^6 at A3 and C11, respectively). It is evident from the individual terms of the budgets (Supplementary Table 1) that the uncertainty on the efficiency estimate is large, but we have investigated three possible biases, as follows. (1) The vertical diffusivity is difficult to constrain. However, the efficiency ratio is only weakly sensitive to its magnitude because the dissolved inorganic carbon (DIC)/DFe ratio of the resupply is similar to that of the export. (2) Our late-summer observations may under-estimate the iron supply, but even the extreme assumption that DFe concentrations over the plateau were much higher earlier in the year (for example, that the 500 m near-bottom DFe value of 0.40 nM at A3 characterized waters at 200 m throughout the rest of the year) still leads to a seasonal sequestration efficiency of 149,000 mol mol^{-1}. (3) The seasonal budget does not include an additional supply from the dissolution of PFe$_{litho}$. If we applied, over the whole season, the maximum PFe$_{litho}$ dissolution flux determined in the short-term budget, this also leads to a high efficiency ratio of 159,000 mol mol^{-1}. The high efficiencies resulting from the short and seasonal budgets are consistent with our shipboard observations at A3 of C:Fe uptake rate ratios (200,000 ± 118,000 mol mol^{-1}) and C:Fe contents of phytoplankton cells (227,000 ± 5,000 mol mol^{-1}).

Our different approaches to estimating the sequestration efficiency give results that range over an order of magnitude, yet all the estimates are at least 10 times higher than the mean efficiency of ~4,300 mol mol^{-1} estimated from SOFEX and other mesoscale iron addition experiments in the Southern Ocean[7]. Our estimates do not include horizontal transport of DIC and DFe, but mixing with waters outside the plateau would tend to supply DIC and remove DFe from the plateau and thus to increase the sequestration efficiency. Our ratio might be biased if a strong decoupling between carbon and iron cycling existed, but our measurements show that, in the bloom, similar percentages of the fixed C (40%) and Fe (50%) were

remineralized as DIC and DFe, respectively. Thus the conclusion that natural iron fertilization over the Kerguelen plateau has a high carbon sequestration efficiency is quite robust.

The higher carbon sequestration efficiency of the natural bloom in comparison to mesoscale iron addition experiments derives from differences in both terms of the sequestration efficiency ratio. The excess of carbon export was higher during the KEOPS bloom than SOFEX, probably owing to its later stage of development (Supplementary Materials), and because during SOFEX the end of the bloom was not reached[26]. The addition of DFe occurs slowly and continuously during natural fertilization, whereas purposeful additions of large amounts of iron within a short period lead to the loss of most (80–95%) of the added DFe during mesoscale enrichment experiments[27].

The occurrence of the Kerguelen bloom is contingent on persistent iron fertilization, but nevertheless its duration is due to the concomitant supply of macronutrients from surrounding waters and from below. The KEOPS results are tightly linked to the mode of iron supply, which is different from dust deposition or purposeful additions. Our estimate of carbon sequestration efficiency is important to the evaluation of palaeo hypotheses[9,10] and future climate change scenarios[11] linking iron supply from below with ocean productivity and its influence on atmospheric carbon dioxide. The enrichment of deep water by terrigenous inputs[10] or by the indirect effect of dust deposition[9] has been proposed. The complex interplay between the iron and carbon cycles[28] precludes a direct extrapolation of how much CO_2 will be really removed from the atmosphere following the fertilization of the Southern Ocean from below, but clearly the natural system is extremely sensitive to iron, far more so than suggested by mesoscale iron addition experiments. However, for the reasons presented above, we emphasize that the high sequestration efficiency determined in the Kerguelen bloom should not be taken as an indication that controversial geoengineering CO_2 mitigation proposals[29,30] will be able to obtain high efficiencies.

METHODS SUMMARY

Composite images from the satellite-borne MODIS (Moderate Resolution Imaging Spectroradiometer) and MERIS (Medium Resolution Imaging Spectrometer Instrument), processed by ACRI Co. and delivered daily to the RV *Marion Dufresne*, were used to locate the stations inside and outside the bloom. The vessel's seawater supply was used to monitor surface waters. Water column properties were measured *in situ* by instruments mounted on a rosette frame, and discrete samples were collected using Niskin bottles. *In situ* pumps were used to collect particles. All the samples were analysed applying standard protocols. Special attention was paid to Fe sampling and measurements, using trace metal clean techniques.

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

Figure 3 | Carbon export at A3 and C11. a, Profiles of ^{234}Th activity at A3 (red lines) and C11 (blue lines). The date of the visits (in 2005) are as follows. At A3; 25 January (circles), 3 February (triangles), 12 February (squares): at C11; 26 January (circles), 05 February (triangles). The dashed line denotes the mean profile of ^{238}U activity. **b**, POC fluxes at A3 and C11. The POC fluxes were derived from ^{234}Th fluxes using a non-steady-state method and POC:^{234}Th ratios (see Methods). The contribution of the vertical diffusivity to the fluxes was negligible at 200 m and around 10% at 100 m, at A3 and C11. The non-steady-state contribution was less than 4% at 100 m. At 200 m it was 34% and 29% at A3 and C11, respectively. Error bars, ±1 s.d. of the total flux. At both depths, POC fluxes were substantially higher inside the bloom compared to outside.

1. Martin, J. H. Glacial-interglacial CO_2 change: The iron hypothesis. *Paleoceanography* **5**, 1–13 (1990).
2. Brzezinsky, M. A. *et al.* A switch from Si(OH)$_4$ to NO$_3^-$ depletion in the glacial Southern Ocean. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL014349 (2002).
3. Sigman, D. M. & Boyle, E. A. Glacial/interglacial variations in atmospheric carbon dioxide. *Science* **407**, 859–869 (2000).
4. Bopp, L., Kohfeld, K. E., Le Quéré, C. & Aumont, O. Dust impact on marine biota and atmospheric CO_2 during glacial periods. *Paleoceanography* **18**, doi:10.1029/2002PA000810 (2003).
5. Watson, A. J., Bakker, D. C. E., Ridgwell, A. J., Boyd, P. W. & Law, C. Effect of iron supply on Southern Ocean CO_2 uptake and implications for glacial atmospheric CO_2. *Nature* **407**, 730–733 (2000).
6. Boyd, P. W. *et al.* Mesoscale iron enrichment experiments 1993–2005: Synthesis and future directions. *Science* **315**, 612–617 (2007).
7. De Baar, H. J. W. *et al.* Synthesis of iron fertilization experiments: from the iron age in the age of enlightenment. *J. Geophys. Res.* **110**, doi:10.1029/2004JC002601 (2005).
8. Boyd, P. W., Jackson, G. A. & Waite, A. M. Are mesoscale perturbation experiments in polar waters prone to physical artefacts? Evidence from algal aggregation modelling studies. *Geophys. Res. Lett.* **20**, doi:10.1029/2001GL014210 (2002).
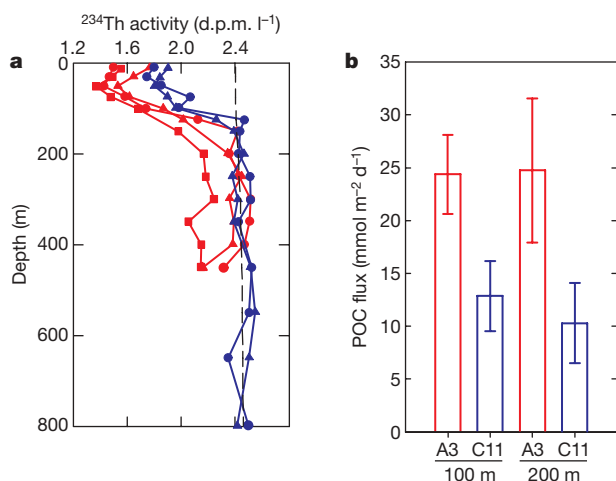
1073

9. Ridgwell, A. J. & Watson, A. Feedback between aeolian dust, climate, and atmospheric CO_2 in glacial time. *Paleoceanogr.* **17**, 1059, doi:10.1029/2001PA000729 (2002).
10. Latimer, J. C. & Filipelli, G. M. Terrigenous input and paleoproductivity in the Southern Ocean. *Paleoceanography* **16**, 627–643 (2001).
11. Jickells, T. D. *et al.* Global iron connections between desert dust, ocean biogeochemistry, and climate. *Science* **308**, 67–71 (2005).
12. Sarmiento, J. L., Hughes, C. W., Stouffer, R. J. & Manabe, S. Simulated response of the ocean carbon cycle to anthropogenic climate warming. *Nature* **393**, 245–249 (1998).
13. Marinov, I., Gnanadesikan, A., Toggweiler, J. R. & Sarmiento, J. L. The Southern Ocean biogeochemical divide. *Nature* **441**, 964–967 (2006).
14. Coale, K. H. *et al.* Southern Ocean iron enrichment experiment: Carbon cycling in high- and low-Si waters. *Science* **304**, 408–414 (2004).
15. Boyd, P. W. *et al.* A mesoscale phytoplankton bloom in the polar Southern Ocean stimulated by iron fertilization. *Nature* **407**, 695–702 (2000).
16. Gervais, F., Riebesell, U. & Gorbunov, M. Y. Change in primary productivity and chlorophyll a response to iron fertilization in the Southern Polar Frontal Zone. *Limnol. Oceanogr.* **47**, 1324–1335 (2002).
17. Hart, T. J. Phytoplankton periodicity in Antarctic surface water. *Discov. Rep.* **VIII**, 1–268 (1942).
18. Sullivan, C. W., Arrigo, K. R., McClain, C. R., Comiso, J. C. & Firestone, J. Distribution of phytoplankton blooms in the Southern Ocean. *Science* **262**, 1832–1837 (1993).
19. Tyrell, T. *et al.* Effect of seafloor depth and phytoplankton blooms in high nitrate low chlorophyll (HNLC) regions. *J. Geophys. Res.* **110**, doi:10.1029/2005JG000041 (2005).
20. Blain, S. *et al.* A biogeochemical study of the island mass effect in the context of the iron hypothesis: Kerguelen Islands, Southern Ocean. *Deep-sea Res. I* **48**, 163–187 (2001).
21. Measures, C. I. & Vink, S. Dissolved Fe in the upper waters of the Pacific sector of the Southern Ocean. *Deep-sea Res. II* **48**, 3913–3941 (2001).
22. Timmermans, K. R. *et al.* Growth rates of large and small Southern Ocean diatoms in relation to availability of iron in natural seawater. *Limnol. Oceanogr.* **46**, 260–266 (2001).
23. Law, C., Abraham, E. R., Watson, A. & Liddicoat, M. Vertical eddy diffusion and nutrient supply to the surface mixed layer of the Antarctic Circumpolar Current. *J. Geophys. Res.* **108**, doi:10.1029/2002JC001604 (2003).
24. Boyd, P. W. *et al.* FeCycle: attempting an iron biogeochemical budget from a mesoscale SF_6 tracer experiment in unperturbated low iron waters. *Glob. Biochem. Cycles* **19**, doi:10.1029/2005GB002494 (2005).
25. Rutgers Van Der Loeff, M. M., Buesseler, K. O., Bathmann, U., Hense, I. & Andrews, J. Comparison of carbon and opal export rates between summer and spring bloom in the region of the Antarctic Polar Front, SE Atlantic. *Deep-sea Res. II* **49**, 3849–3869 (2002).
26. Buesseler, K. O., Andrews, J. E., Pike, S. M. & Charette, M. A. The effects of iron fertilization on carbon sequestration in the Southern Ocean. *Science* **304**, 414–417 (2004).
27. Bowie, A. R. *et al.* The fate of added iron during a mesoscale fertilisation experiment in the Southern Ocean. *Deep-sea Res. II* **48**, 2703–2743 (2001).
28. Gnanadesikan, A., Sarmiento, J. L. & Slater, R. D. Effects of patchy ocean fertilization on atmospheric carbon dioxide and biological production. *Glob. Biochem. Cycles* **17**, doi:10.1029/2002GB001940 (2003).
29. Buesseler, K. O. & Boyd, P. W. Will ocean fertilization work? *Science* **300**, 67–68 (2003).
30. Chisholm, S. W., Falkowski, P. G. & Cullen, J. J. Dis-crediting ocean fertilization. *Science* **294**, 309–310 (2001).
31. Wanninkhof, R. H. & McGillis, W. R. A cubic relationship between air-sea exchange and wind speed. *Geophys. Res. Lett.* **26**, 1889–1892 (1999).
32. Nightingale, P. *et al. In situ* elevation of the sea-air gas exchange parameterisations using novel conservative and volatile tracers. *Glob. Biochem. Cycles* **14**, 373–387 (2000).

METHODS

Selection of stations and hydrographic sampling. Composite images of MODIS and MERIS processed by ACRI Co. and delivered daily to the RV *Marion Dufresne* were used to locate the stations inside and outside the bloom. The vessel seawater supply (intake at a depth of 5 m) was used to continuously monitor p_{CO_2}, DIC, temperature, salinity and fluorescence[33]. Vertical sampling of the water column from the surface to the bottom was performed using 22 acid-cleaned 12-l Niskin bottles mounted on a CTD rosette system with a lowered acoustic Doppler current profiler (LADCP) and a C star transmissiometer (WETLABS). Clean trace metal samples were obtained using ten 12-l Go-Flo bottles mounted on a 2,500-m Kevlar line and trigged by Teflon messengers. *In situ* pumps were used to collect particles. The vertical diffusivity, K_z, was estimated from the vertical density profiles and from corresponding Thorpe scales[34,35].

Biogeochemical parameters. Dissolved iron[36] and organic ligands[37] present in filtered ($<0.2\,\mu m$) sea water were measured on board in a clean laboratory. The accuracy of DFe determinations was checked using the new sub-nanomolar Fe standard sampled and analysed during the SAFE cruise[38]. Samples from Go-Flo bottles were filtered ($>0.2\,\mu m$) for particulate trace metal determinations[39]. The crustal ratio $PFe/PAl = 0.2$ (ref. 40) was used to estimate PFe_{lith}. Discrete water samples taken from Niskin bottles were analysed for macronutrients[41], DIC[33], DOC, POC[41], and U/Th β-counting[42]. Rates of production were determined using 24-h deck incubations with simulated *in situ* light levels, using ^{13}C (ref. 41). Uptake and regeneration rates of iron were determined using ^{55}Fe and a biogenic budget[27]. The net iron demand was calculated as the difference between the total biogenic demand and supply. Gross community production and community respiration were concurrently determined by O_2 and DIC evolution using 24-h incubations with simulated light levels[43]. ^{234}Th export fluxes at 100 m and 200 m depth were estimated using a non-steady-state (NSS) one-dimensional model accounting for vertical diffusion (equations in ref. 44). ^{234}Th fluxes at A3 were 2,999 and 5,217 d.p.m. $m^{-2} d^{-1}$ at 100 and 200 m, respectively, and at C11, 2,003 and 2,281 d.p.m. $m^{-2} d^{-1}$ at 100 and 200 m, respectively. Conversion to POC export was based on POC/^{234}Th (C/Th) ratios of 5–210-μm particles from large volume filtrations at 100 m depth (7.7 and 6.1 μmol d.p.m.$^{-1}$ at A3 and C11, respectively). Estimates at 200 m depth were based on C/Th from 1–335-μm particles from free-drifting sediment trap collections at site A3 (4.7 μmol d.p.m.$^{-1}$) and on C/Th ratios from 5–210-μm particles from large volume filtrations at 130 m depth at C11 (4.5 μmol d.p.m.$^{-1}$).

33. Jabaud, A., Metzl, N., Brunet, C., Poisson, A. & Schauer, B. Interannual variability of the carbon dioxide system in the Southern Indian Ocean (20°S-60°S): the impact of a warm anomaly in austral summer. *Glob. Biochem. Cycles* **18**, doi:10.1029/2002GB002017 (2004).
34. Osborn, T. R. Estimate of vertical diffusion from dissipation measurements. *J. Phys. Oceanogr.* **10**, 83–89 (1980).
35. Dillon, T. M. Vertical overturns: a comparison of Thorpe and Ozmidov length scales. *J. Geophys. Res.* **85**, 9601–9613 (1982).
36. Sarthou, G. *et al.* Atmospheric iron deposition and sea-surface dissolved iron concentrations in the eastern Atlantic Ocean. *Deep-Sea Res. I* **50**, 1339–1352 (2003).
37. Croot, P. & Johansson, M. Determination of iron speciation by cathodic stripping voltammetry in seawater using the competing ligand 2-(2-thiazolylazo-*p*-cresol (TAC). *Electroanalysis* **12**, 565–576 (2000).
38. Johnson, K. S. *et al.* Sampling and analysis of Fe: The SAFE iron intercomparison cruise. *Eos Trans. AGU* **87**(36), Ocean Sci. Meet. Suppl. (2006).
39. Cullen, J. T. & Sherrel, M. Techniques for determination of trace metal in small samples of size-fractionated particulate matter: phytoplankton metals off central California. *Mar. Chem.* **67**, 233–247 (1999).
40. Wedepohl, K. H. The composition of continental crust. *Geochim. Cosmochim. Acta* **59**, 1217–1232 (1995).
41. Fernandez, C., Raimbault, P., Caniaux, Y., Garcia, N. & Rimmelin, P. An estimation of annual new production and carbon budget in the northeast Atlantic Ocean during 2001. *J. Geophys. Res.* **110**, doi:10.1029/2004JC002621 (2005).
42. Pike, S. M., Buesseler, K. O., Andrews, J. & Savoye, N. Quantification of ^{234}Th recovery in small volume of sea water samples by inductively coupled plasma-mass spectrometry. *J. Radioanal. Nucl. Chem.* **263**, 355–360 (2005).
43. Robinson, C. & Williams, P. J. I. Plankton net community production and dark respiration in the Arabian Sea during September 1994. *Deep-Sea Res. II* **46**, 745–765 (1999).
44. Savoye, N. *et al.* ^{234}Th sorption and export models in the water column: a review. *Mar. Chem.* **100**, 234–249 (2006).

The role of fluids in lower-crustal earthquakes near continental rifts

Martin Reyners[1], Donna Eberhart-Phillips[1,2] & Graham Stuart[3]

The occurrence of earthquakes in the lower crust near continental rifts has long been puzzling, as the lower crust is generally thought to be too hot for brittle failure to occur[1,2]. Such anomalous events have usually been explained in terms of the lower crust being cooler than normal[3,4]. But if the lower crust is indeed cold enough to produce earthquakes, then the uppermost mantle beneath it should also be cold enough[2], and yet uppermost mantle earthquakes are not observed[5]. Numerous lower-crustal earthquakes occur near the southwestern termination of the Taupo Volcanic Zone (TVZ), an active continental rift in New Zealand[6]. Here we present three-dimensional tomographic imaging of seismic velocities and seismic attenuation in this region using data from a dense seismograph deployment[7]. We find that crustal earthquakes accurately relocated with our three-dimensional seismic velocity model form a continuous band along the rift, deepening from mostly less than 10 km in the central TVZ to depths of 30–40 km in the lower crust, 30 km southwest of the termination of the volcanic zone. These earthquakes often occur in swarms, suggesting fluid movement in critically loaded fault zones[8]. Seismic velocities within the band are also consistent with the presence of fluids, and the deepening seismicity parallels the boundary between high seismic attenuation (interpreted as partial melt) within the central TVZ and low seismic attenuation in the crust to the southwest. This linking of upper and lower-crustal seismicity and crustal structure allows us to propose a common explanation for all the seismicity, involving the weakening of faults on the periphery of an otherwise dry, mafic crust by hot fluids, including those exsolved from underlying melt. Such fluids may generally be an important driver of lower-crustal seismicity near continental rifts.

The TVZ is an active continental rift in the central part of North Island, New Zealand (Fig. 1). It is the product of back-arc spreading within the Hikurangi subduction zone, with the top of the subducted Pacific plate being about 85 km deep beneath the TVZ. The modern central TVZ is the most frequently active and productive silicic volcanic system on Earth, erupting rhyolite at $\sim 0.28 \, \mathrm{m}^3 \, \mathrm{s}^{-1}$ for at least the past 0.34 million years (ref. 9). The average heat flux from the central $6{,}000 \, \mathrm{km}^2$ of the TVZ is very high at $700 \, \mathrm{mW} \, \mathrm{m}^{-2}$ (ref. 10). Also, the southern limit of the TVZ has been migrating southwestwards with time, as indicated by the timing of both recent volcanism and associated surface faulting in the southern TVZ[11].

A recent six-month dense deployment of portable seismographs in central North Island with a station spacing of 15–20 km has provided an opportunity to study the relationship between the TVZ and the lower-crustal seismicity near its southwestern termination. This deployment has allowed detailed three-dimensional tomographic imaging of seismic velocities (compressional P-wave velocity v_P; and the ratio of v_P to the shear S-wave velocity v_P/v_S) throughout

the crust and upper mantle[7]. Here we additionally model the three-dimensional seismic attenuation structure, by determining the quality factor of P-waves Q_P (which is inversely related to attenuation), using the t^* inversion technique of ref. 12. We have included a limited frequency dependence for Q_P, with Q_P being frequency-independent above f_0, and having a frequency dependence of $(f/f_0)^{0.5}$ below f_0. We have taken f_0 to be 10 Hz, so the Q_P determined is equivalent to a Q_P at $\sim 10 \, \mathrm{Hz}$.

This three-dimensional imaging has revealed that in the central TVZ the crust is $\sim 35 \, \mathrm{km}$ thick, with the lower crust greatly modified through intrusion and underplating. This crustal structure is similar to that deduced from two-dimensional modelling of explosive and earthquake sources[13]. Southwest of the termination of the TVZ at Mt Ruapehu the crust thickens by $\sim 10 \, \mathrm{km}$, and is interpreted to be composed of Cretaceous schist and remnant oceanic crust[12]. In the mantle wedge, v_P is lowest ($7.4 \, \mathrm{km \, s}^{-1}$) and v_P/v_S is highest (1.87) at 65 km depth, immediately west of the Taupo caldera. This region is
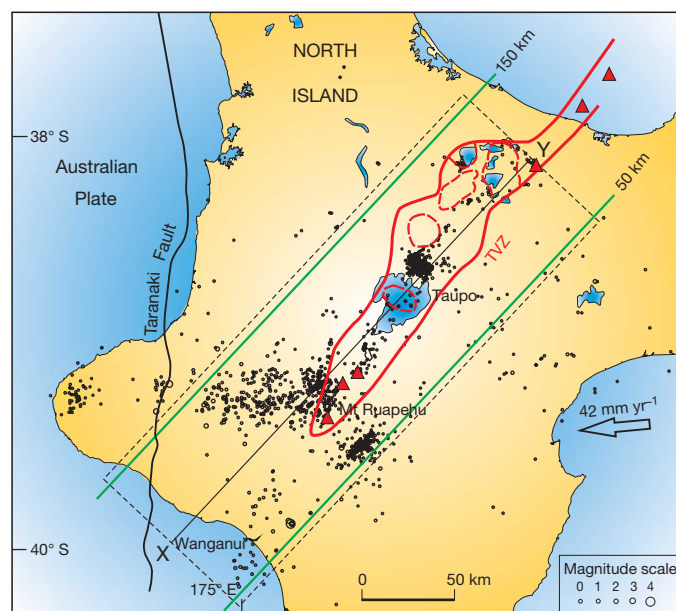
Figure 1 | Tectonic setting of the central North Island of New Zealand. The TVZ is outlined in red, with rhyolite-dominant caldera volcanoes shown dashed and andesite-dominant cone volcanoes shown as triangles[9]. The arrow indicates the velocity of the subducted Pacific plate relative to the overlying Australian plate, and green lines show the depth of the top of the Pacific plate. Circles are earthquakes within the Australian plate relocated using the three-dimensional seismic velocity model of ref. 7. The dashed rectangle encloses earthquakes included in the depth section XY shown in Fig. 2.

[1]GNS Science, PO Box 30 368, Lower Hutt 5040, New Zealand. [2]Geology Department, University of California Davis, Davis, California 95616-8605, USA. [3]School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK.

best interpreted as a significant volume of partial melt, produced by the reaction of fluid released by dehydration of the subducted plate with the convecting mantle wedge. There is no significant low-v_P zone in the mantle wedge southwest of the TVZ, suggesting that the thicker crust there has choked off mantle corner flow[14]. This lack of corner flow is a plausible explanation of the cessation of volcanism to the southwest. The abrupt cessation of corner flow southwest of Mt Ruapehu is likely to promote lateral fluid flow in this region, as suggested by S-wave splitting results[15]. In particular, we would expect the corner flow in the northeast to entrain fluid from the southwest, leading to an anomalous concentration of partial melt, because fluid from a large volume of mantle to the southwest is available to flux melt. This lateral fluid flow is a plausible explanation for the exceptional magmatic productivity of the central TVZ[7].

We have used the three-dimensional seismic velocity model to relocate all earthquakes in the overlying Australian plate during the duration of the seismograph deployment, a total of 1,618 events (Fig. 1). A depth section of those earthquakes within 50 km of the axis of the TVZ is shown in Fig. 2. These events have median standard errors of 0.5 km in both epicentre and depth. Within the TVZ, the earthquakes are mostly less than 10 km deep. At the southwest termination of the TVZ at Mt Ruapehu, the earthquakes form a continuous ~10-km-thick band which dips to the southwest before flattening out in the 30–40 km depth range. The relationship of this seismicity distribution to seismic properties in the crust and uppermost mantle is illustrated in Fig. 3.

Within the TVZ, seismicity in the upper crust is associated with low v_P/v_S (<1.70). Similar low v_P/v_S ratios have been observed beneath active volcanoes in northeastern Japan, and this can be explained by the presence of water[16,17]. The region exhibits vigorous geothermal activity, which is known to be associated with low v_P/v_S (ref. 18). In the lower crust of the TVZ, v_P/v_S increases rapidly and Q_P is low (<300). Using the methodology of ref. 19, Reyners et al.[7] have interpreted the low-v_P feature centred near 18 km depth as a region of texturally equilibrated partial melt. This interpretation is consistent with the low Q_P observed in this region. Seismic attenuation is a strong function of temperature[20], with a lower Q_P indicating a higher temperature. Hence the mid- and lower crust of the TVZ can be interpreted as heavily intruded and underplated, largely because of the large flux of melt from mantle corner flow in this region (evident in the high v_P/v_S and low Q_P in the mantle wedge in Fig. 3). We interpret upper-crustal seismicity within the TVZ to be promoted by fluids exsolved from such intrusions, by means of cyclic fluid pore pressure increases which weaken faults[21]. The earthquakes mostly occur in swarms, suggesting fluid movement in critically loaded fault zones[8]. While most fluid at shallow depth may be of meteoric origin[10], the high $^3He/^4He$ ratios associated with the TVZ[22] suggest that some fluid has been exsolved from melt.

At the southwestern termination of the TVZ near Mt Ruapehu, the low v_P/v_S associated with shallow seismicity in the central TVZ dips to the southwest in concert with the seismicity, and the dipping seismicity band parallels the boundary between low Q_P (<300) within the central TVZ and high Q_P (>600) in the crust to the southwest (Fig. 3). Earthquakes comprising the dipping band of seismicity often occur in swarms[23], again suggesting the influence of fluids. By analogy with the seismicity and seismic properties in the central TVZ, we also interpret earthquakes in the dipping seismicity band as having been promoted by fluids, including those exsolved from the magmatic intrusions at the southwest margin of the TVZ. The continuation of sporadic earthquake activity in the lower crust further to the southwest under the Wanganui basin (Fig. 2) likewise suggests the influence of fluids. These earthquakes often occur as swarms[24], and the lower crust in which they occur is reflective[25]. Both these observations are consistent with the presence of fluids. A Q_P of 500–600 in the uppermost mantle in this region (Fig. 3) precludes the existence of connected melt. Rather, fluids in the lower crust beneath the Wanganui basin are most probably the result of a well-hydrated mantle wedge, caused by dehydration of the thicker-than-normal subducted crust in this region[7].

How robust is our interpretation of seismicity in the lower crust being triggered by fluids? The strength of the crust is governed by strain rate, temperature and rock composition, as well as pore fluid pressure[26]. Geodetically measured strain rates at the southern termination of the TVZ are ~1×10^{-15} s^{-1}, significantly lower than strain rates within the TVZ itself[27]. The fact that the seismicity continues southwestwards, despite the reduction in strain rate, suggests that regional strain rate is not a major factor in the generation of the seismicity. There are no heat flow measurements in the region of the lower-crustal earthquakes, but on the basis of those nearby, the heat flow is probably about 60 mW m^{-2} (ref. 6). Such a heat flow would suggest that the lower crust would be ductile and hence aseismic. We must thus turn to rock composition to explain the lower-crustal seismicity, and it has previously been proposed that the lower crust is drier and more mafic than usual[6]. The high Q_P (>600) seen in the lower crust to the southwest of the TVZ (Fig. 3) would be consistent with a dry, mafic crust. Yet such a change in rock composition does not appear to be reflected in the seismicity distribution. Seismicity parallels the dipping boundary between low Q_P (<300) within the central TVZ and this high-Q_P (>600) lower crust to the southwest, rather than being restricted only to the high-Q_P lower crust.

High fluid pore pressures thus seem the most plausible explanation of the seismicity distribution. Our preferred model for the seismicity is that it is promoted by hot fluids (including those exsolved from underlying melt) weakening faults at the periphery of an otherwise dry, mafic crust. The volume proportion of these hot fluids is likely to be small, insofar as seismologically determined properties
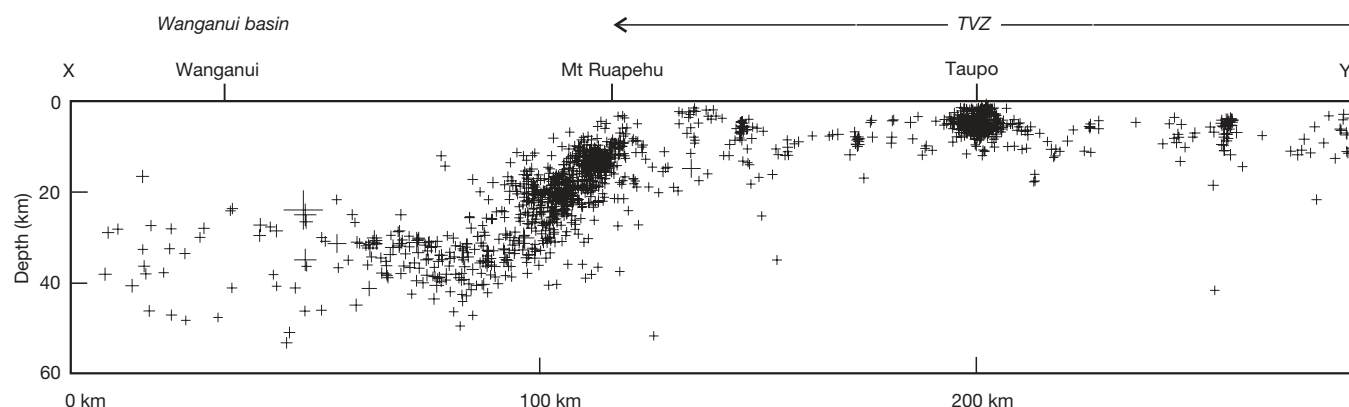

Figure 2 | Depth section of relocated earthquakes in the Australian plate along the axis of the TVZ. Events shown are within 50 km of the line XY shown in Fig. 1. 'Plus' symbol size is scaled to magnitude, with the smallest magnitude 0.9 and the largest magnitude 4.9.

still suggest a dry lower crust. This model is similar to that recently proposed for brittle failure in exhumed lower-crustal rocks, based on petrological and isotopic data[28].

If the lower crust at the southwestern termination of the TVZ is seismic, why are there so few earthquakes in the upper crust? A possible explanation suggested by the high Q_P of the lower crust is that this crust is particularly competent and impervious, restricting fluid flow to a fault-fracture mesh at its periphery. Both Q_P and v_P anisotropy studies[29] suggest that the crust is part of the competent Rakaia/Haast Schist geological terrane. Further south, this rheologically strong terrane is in contact with the subducted Pacific plate, and may contribute to the strong plate coupling in the southern North Island[12]. A competent Sierran block has similarly been suggested as a factor in producing the dipping band of lower-crustal earthquakes

Figure 3 | Depth sections of v_P, v_P/v_S and Q_P (which is inversely related to attenuation) along the axis of the TVZ. Earthquakes ('plus' symbols) in both plates within 20 km of the line XY in Fig. 1 are shown, together with the top surface of the subducted Pacific plate (dashed). Small dots are nodes used in the tomographic inversions. Regions of low resolution are masked but contoured (shown white for v_P and Q_P, and green for v_P/v_S).

beneath the western Sierra Nevada range[30]. The influence of the Rakaia/Haast Schist terrane is further demonstrated by the distribution of lower-crustal earthquakes normal to the TVZ. These extend westwards only as far as the Taranaki fault[6], which marks the western limit of thicker crust associated with this terrane[7]. The ~100 km northwest–southeast extent of the lower-crustal seismicity is much wider than the 30–50 km width of the TVZ. This reflects the wide distribution of low v_P and low Q_P in the uppermost mantle immediately to the northeast[7,29], indicative of elevated temperatures, fluids and melt in this region of active corner flow in the mantle wedge. The TVZ is only the surface expression of an extensive region of mantle corner flow that is progressively eroding the northeastern edge of the Rakaia/Haast Schist terrane. Previous erosion of this terrane further northeast provides an explanation of why lower-crustal earthquakes do not occur along the northwest and southeast edges of the TVZ.

We speculate that the same processes that we observe at the southwestern margin of the TVZ may also apply to lower-crustal earthquakes on the margins of other continental rifts. Earthquake swarm activity in the lower crust has been identified at several rifts, suggesting the possibility of fluid involvement (see Supplementary Information). Our seismicity and tomography results suggest that for these lower-crustal events to occur, not only fluids are needed, but also a rheologically strong lower crust which restricts these fluids to its base.

1.   Ruff, L. J. in *Rheology and Deformation of the Lithosphere at Continental Margins* (eds Karner, G. D., Taylor, B., Driscoll, N. W. & Kohlstedt, D. L.) 138–165 (Columbia Univ. Press, New York, 2004).
2.   Chen, W.-P. & Molnar, P. Focal depths of intracontinental and intraplate earthquakes and their implications for the thermal and mechanical properties of the lithosphere. *J. Geophys. Res.* **88,** 4183–4214 (1983).
3.   Doser, D. I. & Yarwood, D. R. Deep crustal earthquakes associated with continental rifts. *Tectonophysics* **229,** 123–131 (1994).
4.   Shudofsky, G. N., Cloetingh, S., Stein, S. & Wortel, R. Unusually deep earthquakes in East Africa: constraints on the thermo-mechanical structure of a continental rift system. *Geophys. Res. Lett.* **14,** 741–744 (1987).
5.   McKenzie, D., Jackson, J. & Priestley, K. Thermal structure of oceanic and continental lithosphere. *Earth Planet. Sci. Lett.* **233,** 337–349 (2005).
6.   Sherburn, S. & White, R. S. Crustal seismicity in Taranaki, New Zealand using accurate hypocentres from a dense network. *Geophys. J. Int.* **162,** 494–506 (2005).
7.   Reyners, M., Eberhart-Phillips, D., Stuart, G. & Nishimura, Y. Imaging subduction from the trench to 300 km depth beneath the central North Island, New Zealand, with Vp and Vp/Vs. *Geophys. J. Int.* **165,** 565–583 (2006).
8.   Hainzl, S. & Fischer, T. Indications of successively triggered rupture growth underlying the 2000 earthquake swarm in Vogtland/NW Bohemia. *J. Geophys. Res.* **107,** doi:10.1029/2002JB001865 (2002).
9.   Wilson, C. J. N. *et al.* Volcanic and structural evolution of Taupo Volcanic Zone, New Zealand: a review. *J. Volcanol. Geotherm. Res* **68,** 1–28 (1995).
10. Bibby, H. M., Caldwell, T. G., Davey, F. J. & Webb, T. H. Geophysical evidence on the structure of the Taupo Volcanic Zone and its hydrothermal circulation. *J. Volcanol. Geotherm. Res.* **68,** 29–58 (1995).
11. Villamor, P. & Berryman, K. R. Evolution of the southern termination of the Taupo Rift, New Zealand. *NZ J. Geol. Geophys.* **49,** 23–37 (2006).
12. Eberhart-Phillips, D., Reyners, M., Chadwick, M. & Chiu, J.-M. Crustal heterogeneity and subduction processes: 3-D Vp, Vp/Vs and Q in the southern North Island, New Zealand. *Geophys. J. Int.* **162,** 270–288 (2005).
13. Harrison, A. & White, R. S. Lithospheric structure of an active backarc basin: the Taupo Volcanic Zone, New Zealand. *Geophys. J. Int.* **167,** 968–990 (2006).
14. Kincaid, C. & Sacks, I. S. Thermal and dynamical evolution of the upper mantle in subduction zones. *J. Geophys. Res.* **102,** 12,295–12,315 (1997).
15. Audoine, E., Savage, M. K. & Gledhill, K. Anisotropic structure under a back-arc spreading region, the Taupo Volcanic Zone, New Zealand. *J. Geophys. Res.* **109,** doi:10.1029/2003JB002932 (2004).
16. Nakajima, J., Matsuzawa, T., Hasegawa, A. & Zhao, D. Three-dimensional structure of Vp, Vs, and Vp/Vs beneath northeastern Japan: implications for arc magmatism and fluids. *J. Geophys. Res.* **106,** 21,843–21,857 (2001).
17. Nakajima, J., Matsuzawa, T., Hasegawa, A. & Zhao, D. Seismic imaging of arc magma and fluids under the central part of northeastern Japan. *Tectonophysics* **341,** 1–17 (2001).
18. Chatterjee, S. N., Pitt, A. M. & Iyer, H. M. Vp/Vs ratios in the Yellowstone National Park region, Wyoming. *J. Volcanol. Geotherm. Res* **26,** 213–230 (1985).
19. Takei, Y. Effect of pore geometry on Vp/Vs: from equilibrium geometry to crack. *J. Geophys. Res.* **107,** doi:10.1029/2001JB000522 (2002).

20. Faul, U. H. & Jackson, I. The seismological signature of temperature and grain size variations in the upper mantle. *Earth Planet. Sci. Lett.* **234**, 119–134 (2005).
21. Sibson, R. H. Fluid involvement in normal faulting. *J. Geodynam.* **29**, 469–499 (2000).
22. Hulston, J. R. & Lupton, J. E. Helium isotope studies of geothermal fields in the Taupo Volcanic Zone, New Zealand. *J. Volcanol. Geotherm. Res* **74**, 297–321 (1996).
23. Hayes, G., Reyners, M. & Stuart, G. The Waiouru, New Zealand, earthquake swarm: persistent mid crustal activity near an active volcano. *Geophys. Res. Lett.* **31**, doi:10.1029/2004GL020709 (2004).
24. Garrick, R. A. & Gibowicz, S. J. Continuous swarm-like seismicity: the Wanganui, New Zealand, earthquakes. *Geophys. J. R. Astron. Soc.* **75**, 493–512 (1983).
25. Stern, T. A. & Davey, F. J. in *Origin and Evolution of Sedimentary Basins and their Energy and Mineral Resources* (ed. Price, R. A.) 73–85 (Geophys. Monogr. Ser. 48, American Geophysical Union, Washington DC, 1989).
26. Kohlstedt, D. L., Evans, B. & Mackwell, S. J. Strength of the lithosphere: constraints imposed by laboratory experiments. *J. Geophys. Res.* **100**, 17587–17602 (1995).
27. Beavan, J., Ellis, S., Wallace, L. & Denys, P. in *Tectonics of a Continental Transform Plate Boundary: the South Island, New Zealand* (Geophys. Monogr. Ser., American Geophysical Union, Washington DC, in the press).
28. Comacho, A., Lee, J. K. W., Hensen, B. J. & Braun, J. Short-lived orogenic cycles and the eclogitization of cold crust by spasmodic hot fluids. *Nature* **435**, 1191–1196 (2005).
29. Eberhart-Phillips, D., Reyners, M., Chadwick, M. & Stuart, G. 3-D distribution of anisotropy and attenuation in the Hikurangi Subduction zone, New Zealand. *Eos* (Fall Meet. Suppl.) **87** (52), abstr. T21G–02 (2006).
30. Wong, I. G. & Savage, W. U. Deep intraplate seismicity in the western Sierra Nevada, central California. *Bull. Seismol. Soc. Am.* **73**, 797–812 (1983).

Author Contributions M.R. and G.S. planned and undertook the seismograph deployment, M.R. coordinated the data analysis, and D.E.-P. took the lead in the tomographic inversions. All authors contributed to the writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.R. (m.reyners@gns.cri.nz).

LETTERS

Resource-use efficiency and plant invasion in low-resource systems

Jennifer L. Funk[1] & Peter M. Vitousek[1]

No species can maximize growth, reproduction and competitive ability across all environments, so the success of invasive species is habitat-dependent. Nutrient-rich habitats often experience more invasion than resource-poor habitats[1–4], a pattern consistent with traits generally associated with successful invaders (high growth rates, early reproduction and many offspring[5–8]). However, invaders do colonize resource-poor environments, and the mechanisms that allow their success in these systems are poorly understood. Traits associated with resource conservation are widespread among species adapted to resource-poor environments[9–11], and invasive species may succeed in low-resource environments by employing resource conservation traits such as high resource-use efficiency (RUE; carbon assimilation per unit of resource). We investigated RUE in invasive and native species from three habitats in Hawaii where light, water or nutrient availability was limiting to plant growth. Here we show that across multiple growth forms and broad taxonomic diversity invasive species were generally more efficient than native species at using limiting resources on short timescales and were similarly efficient when RUE measures were integrated over leaf lifespans. Our data challenge the idea that native species generally outperform invasive species under conditions of low resource availability[3], and suggest that managing resource levels is not always an effective strategy for invasive species control.

Invasive species represent one of the most serious threats to biodiversity and ecosystem function worldwide[12], and understanding the mechanisms by which invasive species outperform native species is crucial to controlling their spread. One mechanism that has received considerable recent attention is the fluctuating resource hypothesis[13,14], which proposes that invasion is facilitated by high resource availability resulting from disturbance or low resource uptake by the native plant community. This hypothesis suggests that resource-demanding invasive species are likely to enter low-resource habitats following a disturbance that increases resource availability, such as clear cutting, soil disturbance or fire. However, in the absence of continued disturbance, it is unclear whether invasive species will persist[15], particularly if introduced species experience strong competition from native species that are adapted to low-resource conditions[16]. The idea that the performance of invasive species will be suppressed under low-resource conditions is a crucial component of invasive species control programmes and native ecosystem restoration strategies that manipulate resource availability to promote the growth of native species[3,17,18]. Nevertheless, invasions into resource-poor habitats occur and we know relatively little about the mechanisms involved.

To outperform native species and persevere in a low-resource environment, invasive species must actively increase resource availability (for example, N_2-fixation or positive feedbacks through rapid litter turnover and decomposition[19]), promote continued disturbance that increases resource availability (for example, fire), be better

at acquiring limiting resources or be more efficient at using limiting resources. Relatively few invaders introduce 'novel' traits to a native community (for example, N_2-fixation or high flammability). More commonly, invasive species differ quantitatively in traits already present in existing native species[19,20]. Plants adapted to low-resource environments generally possess traits associated with resource conservation (a long leaf lifespan, high concentrations of defence compounds, low tissue-nutrient content or thicker leaves), which results in reduced rates of growth but maximizes RUE[9–11]. The few ecologically and taxonomically appropriate comparisons of RUE in invasive and native species in resource-poor systems show mixed patterns of RUE in invaders relative to co-occurring natives[21,22]. The paucity of data on RUE in invasive and native species, and the potential impact of RUE on invasive species control strategies warrant a thorough examination of RUE as a potential mechanism to explain the success of invasive species in low-resource systems.

We compared leaf-level physiological traits associated with RUE in 19 pairs of phylogenetically related invasive and native species from three habitats in Hawaii where light, water or nutrient availability was limiting to plant growth. Trait variation among species often reflects phylogenetic relationships in that more closely related taxa share similar trait values. Thus, we employed a phylogenetic comparative design to minimize trait differences associated with comparing unrelated species and disparate life forms[23,24]. To identify generalizations in patterns of RUE, we surveyed a broad array of plant groups, including ferns, C_3 and C_4 grasses, herbs, shrubs and trees (Supplementary Table 1).

Across all habitats, invasive species showed higher rates of carbon assimilation relative to native species (Fig. 1a). Higher assimilation rates for invaders corresponded with higher light-use efficiency (Fig. 1b), instantaneous nitrogen-use efficiency (PNUE) and instantaneous energy-use efficiency (PEUE) in light- and nutrient-limited systems (Fig. 2a, b). Instantaneous water-use efficiency (WUE) was not significantly different among invasive and native species in any habitat (Fig. 2c). Collectively, these instantaneous measures of RUE support the idea that invasive species can outperform native species in low-resource environments.

Species adapted to resource-limited environments often have slower leaf turnover and create thicker leaves to minimize nutrient and leaf loss[11]. Retaining leaves for a longer time period maximizes carbon assimilation per unit resource invested in leaf construction over the lifespan of the leaf[9,10]. Thus, RUE integrated over leaf lifespan more accurately depicts species differences in resource use over longer timescales[25]. Higher leaf lifespan in native relative to invasive species (see Supplementary Table 2) compensated for higher instantaneous RUE in invasive species and resulted in similar integrated PNUE and PEUE (Fig. 2d, e). In one habitat (PNUE in N-limited habitat; Fig. 2d), a measure of integrated RUE was significantly higher for natives relative to invaders. Leaf $\delta^{13}C$, which is often used

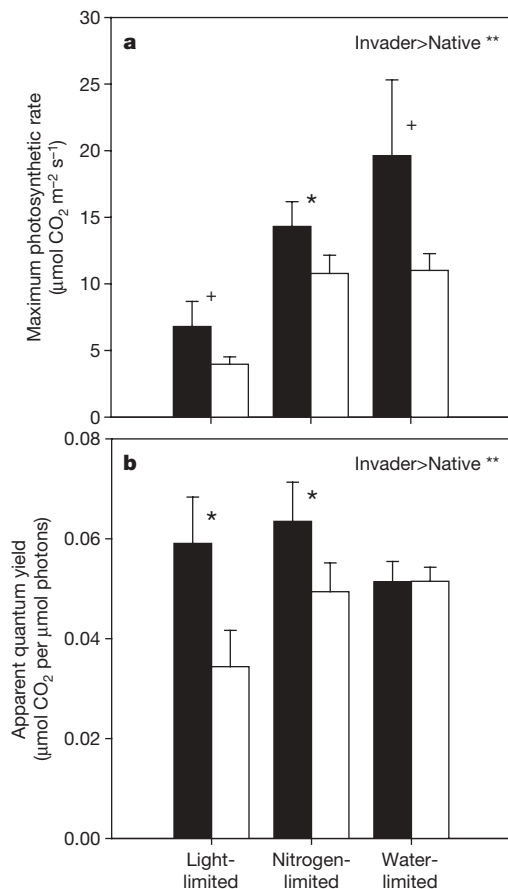[1]Department of Biological Sciences, Stanford University, Stanford, California 94305-5020, USA.

Figure 1 | Photosynthetic rates and light-use efficiency (apparent quantum yield) for 19 phylogenetically related pairs of invasive and native plant species from three habitats in Hawaii. Data are means and standard error for invasive (black bars) and native (open bars) species. Data for each trait were analysed using paired *t*-tests across all 19 pairs and also within each habitat ($+$ denotes $P < 0.1$, $*$ denotes $P < 0.05$ and $**$ denotes $P < 0.01$). The number of paired comparisons was 4, 10 and 5 for light-, nitrogen- and water-limited habitats, respectively.

Figure 2 | Instantaneous and time-integrated measures of nitrogen-, energy- and water-use efficiency for phylogenetically related pairs of invasive and native plant species from three habitats in Hawaii. **a, d**, Photosynthetic nitrogen-use efficiency; **b, e**, photosynthetic energy-use efficiency; and **c, f**, water-use efficiency. Data and symbols as in Fig. 1 (NS, no significant differences across habitats). Instantaneous PNUE, PEUE and WUE measures include data from 19, 13 and 19 pairs, respectively. Integrated PNUE, PEUE and WUE include data from 12, 13 and 17 pairs, respectively.

to integrate WUE over leaf lifespan (see Supplementary Methods), was similar among native and invasive species (Fig. 2f). Overall, the time-integrated RUE data suggest that invaders have no long-term advantage or disadvantage under conditions of continued low resources.

Our results suggest that high RUE is a plausible mechanism for plant invasion and persistence in low-resource systems; however, the importance of RUE to invasive species success will vary across habitats and timescales of observation. Specifically, instantaneous and integrated measures of RUE portray two different scenarios for invasive species success. The instantaneous measures of RUE suggest that invasive species can outperform natives on short (for example, seasonal) timescales. In contrast, our time-integrated RUE measures indicate that invaders are not at a disadvantage on longer timescales (for example, multiple seasons) and may persist under conditions of continued low-resource availability. These results contradict the general paradigm that invasive species allocate resources to growth and reproduction at the expense of resource conservation[5–8]—an idea resulting from studies conducted in predominantly disturbed or resource-rich environments. However, the potentially different conclusions derived from instantaneous versus integrated measures of RUE highlight the need to resolve their potential to influence plant community dynamics for various timescales and phases of invasion (for example, establishment and persistence).

Our finding that native species did not have appreciably higher instantaneous or integrated RUE relative to invaders indicates that

lowering resource availability to prevent the spread of invaders[17,18] may not always be effective. Furthering our understanding of the mechanisms governing the interactions among invasive and native species across resource gradients is crucial to developing successful management and restoration strategies. For example, future research should examine how RUE may act synergistically with other plant traits to promote invasiveness (such as positive feedbacks to enhance resource availability, formation of monospecific stands through vegetative reproduction or seed bank saturation). In addition, more data are needed on how patterns of resource acquisition and use vary spatially and temporally (for example, in response to resource pulses) and how they interact with processes occurring on larger spatial scales (such as propagule dispersal and herbivory[26]) to influence patterns of invasive species establishment, spread and persistence.

METHODS

Species and site description. Species comparisons were selected on the basis of co-occurring of phylogenetically related invasive and native species of a similar growth form and at a given site (similar light, precipitation, elevation and soil substrate age within a one-mile radius). In total, there were three congeneric, fourteen confamilial and two within-order comparisons (Supplementary Table 1). Sites were grouped based on light and water availability into three habitat types. 'Nitrogen-limited' sites occurred throughout Hawaii Volcanoes National Park on the windward side of the island of Hawaii and were characterized by high light-levels ($>600\ \mu mol\ photon\ m^{-2}\ s^{-1}$) and high precipitation ($>1,000\ mm\ yr^{-1}$). 'Light-limited' sites occurred in closed-canopy forests within the national park and were characterized by low light-levels ($<600\ \mu mol\ photon\ m^{-2}\ s^{-1}$) and high precipitation ($>1,000\ mm\ yr^{-1}$). 'Water-limited' sites occurred on the leeward side of Hawaii in the Puu Waa Waa ranch and

the Palamanui Preserve. These sites were characterized by high light-levels ($>600\,\mu mol$ photon $m^{-2}\,s^{-1}$) and low precipitation ($<500\,mm\,yr^{-1}$). All sites were characterized by young, nitrogen-poor volcanic soils ranging in age from 100 to 10,000 years (ref. 27).

Measurements. In June and July of 2004, two recently initiated leaves on five plants per species were tagged and monitored every two to four weeks for eighteen months to determine leaf longevity. Leaf lifespan was estimated from leaf number, leaf birth rate and plastochron interval[28]. In September and October of 2004, gas exchange measurements were conducted on one recently mature leaf per plant (five plants per species). Photosynthetic rates were measured with an LI-6400 portable photosynthesis system (LI-COR). Ambient CO_2 concentration was maintained at 400 µl per litre and relative humidity was maintained between 40 and 80%. Light response curves were conducted by varying light level between 0 and 2,000 µmol photon $m^{-2}\,s^{-1}$, while leaf temperature was held constant.

After each gas exchange measurement, leaves were clipped and area was measured using a scanner and imaging program (Beta 4.0.2, Scion Image). Leaves were then dried at 65 °C and weighed to determine leaf mass per area. Leaf material was ground and analysed for leaf N content with a Costech ECS 4010 elemental analyser (Costech Analytical Technologies). $\delta^{13}C$ was measured with an elemental analyser (Costech ECS 4010) interfaced to an isotope ratio mass spectrometer (Finnigan Delta V Advantage, Finnigan MAT). Cost of leaf construction was calculated as previously described[29], substituting total N for organic N (ref. 30). Heat of combustion was determined from 150–200 mg pellets of dry, ground leaf tissue using a Parr 1425 Semimicro bomb calorimeter (Parr Instrument Company). Equations for RUE measures can be found in the Supplementary Methods.

A mixed-model nested ANOVA was conducted to evaluate RUE differences among native and invasive species, using 'pair' as a random effect and 'invasive nested within pair' as a fixed effect[24]. All analyses were highly significant ($P < 0.001$), but conveyed nothing about directional trends across native and invasive species. Paired t-tests controlled for relatedness and were used to assess directional differences in RUE traits (data presented in Figs 1 and 2). Data that violated the ANOVA assumptions of normality and homogeneity of variance were rank transformed. All analyses were performed in JMP 5.1.2 (SAS Institute).

1. Huenneke, L. F., Hamburg, S. P., Koide, R., Mooney, H. A. & Vitousek, P. M. Effects of soil resources on plant invasion and community structure in California serpentine grassland. *Ecology* **71**, 478–491 (1990).

2. Burke, M. J. W. & Grime, J. P. An experimental study of plant community invasibility. *Ecology* **77**, 776–790 (1996).

3. Daehler, C. C. Performance comparisons of co-occurring native and alien invasive plants: implications for conservation and restoration. *Annu. Rev. Ecol. Evol. Syst.* **34**, 183–211 (2003).

4. Gross, K. L., Mittelbach, G. G. & Reynolds, H. L. Grassland invasibility and diversity: responses to nutrients, seed input, and disturbance. *Ecology* **86**, 476–486 (2005).

5. Pysek, P., Prach, D. & Smilauer, P. in *Plant Invasions: General Aspects and Special Problems* (eds Pysek, P., Prach, D., Rejmanek, M. & Wade, M.) 39–66 (SPB Academic Publishing, Amsterdam, 1995).

6. Rejmanek, M. & Richardson, D. M. What attributes make some plant species more invasive? *Ecology* **77**, 1655–1661 (1996).

7. Grotkopp, E., Rejmanek, M. & Rost, T. L. Toward a causal explanation of plant invasiveness: seedling growth and life-history strategies of 29 pine (*Pinus*) species. *Am. Nat.* **159**, 396–419 (2002).

8. Hamilton, M. A. *et al.* Life-history correlates of plant invasiveness at regional and continental scales. *Ecol. Lett.* **8**, 1066–1074 (2005).

9. Chapin, F. S. III. The mineral nutrition of wild plants. *Annu. Rev. Ecol. Syst.* **11**, 233–260 (1980).

10. Vitousek, P. Nutrient cycling and nutrient use efficiency. *Am. Nat.* **119**, 553–572 (1982).

11. Coley, P., Bryant, J. & Chapin, F. S. III. Resource availability and plant antiherbivore defense. *Science* **230**, 895–899 (1985).

12. Millennium Ecosystem Assessment. *Ecosystems and Human Well-Being: Biodiversity Synthesis* (World Resources Institute, Washington, DC, 2005).

13. Davis, M. A., Grime, J. P. & Thompson, K. Fluctuating resources in plant communities: a general theory of invasibility. *J. Ecol.* **88**, 528–534 (2000).

14. Blumenthal, D. Interrelated causes of plant invasion. *Science* **310**, 243–244 (2005).

15. D'Antonio, C. M., Dudley, T. L. & Mack, M. in *Ecosystems of Disturbed Ground* (ed. Walker, L. R.) 413–452 (Elsevier, The Netherlands, 1999).

16. Tilman, D. Niche tradeoffs, neutrality, and community structure: a stochastic theory of resource competition, invasion, and community assembly. *Proc. Natl Acad. Sci. USA* **101**, 10854–10861 (2004).

17. Blumenthal, D. M., Jordan, N. R. & Russelle, M. P. Soil carbon addition controls weeds and facilitates prairie restoration. *Ecol. Appl.* **13**, 605–615 (2003).

18. Suding, K. N., LeJeune, K. D. & Seastedt, T. R. Competitive impacts and responses of an invasive weed: dependencies on nitrogen and phosphorus availability. *Oecologia* **141**, 526–535 (2004).

19. Ehrenfeld, J. G. Effects of exotic plant invasions on soil nutrient cycling processes. *Ecosystems* **6**, 503–523 (2003).

20. Vitousek, P. M. Biological invasions and ecosystem processes: towards an integration of population biology and ecosystem studies. *Oikos* **57**, 7–13 (1990).

21. Baruch, Z. & Goldstein, G. Leaf construction cost, nutrient concentration, and net CO_2 assimilation of native and invasive species in Hawaii. *Oecologia* **121**, 183–192 (1999).

22. Reed, H. E., Seasteadt, T. R. & Blair, J. M. Ecological consequences of C_4 grass invasion of a C_4 grassland: a dilemma for management. *Ecol. Appl.* **15**, 1560–1569 (2005).

23. Agrawal, A. A. & Kotanen, P. M. Herbivores and the success of exotic plants: a phylogenetically controlled experiment. *Ecol. Lett.* **6**, 712–715 (2003).

24. Burns, J. H. Relatedness and environment affect traits associated with invasive and noninvasive introduced Commelinaceae. *Ecol. Appl.* **16**, 1367–1376 (2006).

25. Reich, P. B., Walters, M. B. & Ellsworth, D. S. Leaf life-span in relation to leaf, plant, and stand characteristics among diverse ecosystems. *Ecol. Monogr.* **62**, 365–392 (1992).

26. Mitchell, C. E. *et al.* Biotic interactions and plant invasions. *Ecol. Lett.* **9**, 726–740 (2006).

27. Vitousek, P. M., Walker, L. R., Whiteaker, L. D. & Matson, P. A. Nutrient limitations to plant growth during primary succession in Hawaii Volcanoes National Park. *Biogeochemistry* **23**, 197–215 (1993).

28. Reich, P. B., Uhl, C., Walters, M. B., Prugh, L. & Ellsworth, D. S. Leaf demography and phenology in Amazonian rain forest: a census of 40,000 leaves of 23 tree species. *Ecol. Monogr.* **74**, 3–23 (2004).

29. Williams, K., Percival, F., Merino, J. & Mooney, H. A. Estimation of tissue construction cost from heat of combustion and organic nitrogen content. *Plant Cell Environ.* **10**, 725–734 (1987).

30. Nagel, J. M. *et al.* Atmospheric CO_2 enrichment alters energy assimilation, investment and allocation in *Xanthium strumarium*. *New Phytol.* **166**, 513–523 (2005).

LETTERS

How swifts control their glide performance with morphing wings

D. Lentink[1], U. K. Müller[1], E. J. Stamhuis[2], R. de Kat[3], W. van Gestel[1], L. L. M. Veldhuis[3], P. Henningsson[4], A. Hedenström[4], J. J. Videler[2,5] & J. L. van Leeuwen[1]

Gliding birds continually change the shape and size of their wings[1–6], presumably to exploit the profound effect of wing morphology on aerodynamic performance[7–9]. That birds should adjust wing sweep to suit glide speed has been predicted qualitatively by analytical glide models[2,10], which extrapolated the wing's performance envelope from aerodynamic theory. Here we describe the aerodynamic and structural performance of actual swift wings, as measured in a wind tunnel, and on this basis build a semi-empirical glide model. By measuring inside and outside swifts' behavioural envelope, we show that choosing the most suitable sweep can halve sink speed or triple turning rate. Extended wings are superior for slow glides and turns; swept wings are superior for fast glides and turns. This superiority is due to better aerodynamic performance—with the exception of fast turns. Swept wings are less effective at generating lift while turning at high speeds, but can bear the extreme loads. Finally, our glide model predicts that cost-effective gliding occurs at speeds of 8–10 m s^{-1}, whereas agility-related figures of merit peak at 15–25 m s^{-1}. In fact, swifts spend the night ('roost') in flight at 8–10 m s^{-1} (ref. 11), thus our model can explain this choice for a resting behaviour[11,12]. Morphing not only adjusts birds' wing performance to the task at hand, but could also control the flight of future aircraft[7].

Bird wings lend themselves to morphing because they have an articulated skeleton under muscular control, and because the changing overlap between feathers allows continuous changes in wing shape and wing size. Gliding birds sweep their hand-wings back at high flight speeds[1–5,13], and spread their wings in turns[4]. To test whether a bird's chosen wing geometry maximizes its flight performance, biologists have focused on gliding flight[1–5,13], during which changes in wing geometry are not related to wing beat. Aerodynamic forces have been inferred from the behaviour of freely gliding birds[1,2,5,13]; lift and drag have also been measured directly on single bird wings fixed in one shape[14,15]. These approaches provide no information on morphing outside the bird's behavioural envelope, and must be supplemented with aerodynamic theory in order to predict wing aerodynamic performance and bird glide performance[2,16,17].

Rather than estimating how wing geometry affects wing performance, we measured it in a wind tunnel. We chose the common swift (*Apus apus*), which spends most of its life on the wing, foraging, courting, migrating and even roosting[11,12,18,19], and has a gliding repertoire to suit: soaring, gliding and 'flap-gliding'. Flap-gliding birds alternate flapping and gliding at 1–2 s intervals[11,19], matching the speeds of flapping and gliding episodes[20]. With speed approximately constant, glides can be approximated as 'equilibrium gliding', which encompasses turns and straight glides (turns with infinite radius)

(Fig. 1; Methods). During this steady state, flight performance can be deduced from four readily measured parameters: aerodynamic lift, drag, body mass, and flight velocity (Fig. 1b). These determine the swift's glide path, conventionally described by glide angle, turning radius, and bank angle[21]. Glide path and velocity determine bird glide performance.

Aerodynamic force is proportional to force coefficient × wing area × square of glide speed[21]. Swifts control force coefficient by altering wing shape, angle of attack, and speed. Increasing sweep angle from 5° (fully extended) to 50° (Fig. 2a) decreases wing area and shape (that is, aspect ratio) by roughly one-third (Fig. 2b, c). We quantified how variable sweep affects wing aerodynamics by measuring
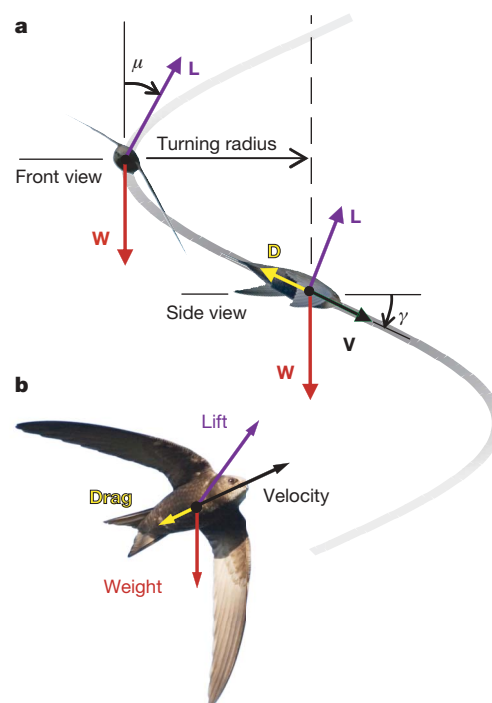
Figure 1 | Equilibrium gliding along a helical path. a, Turning swifts glide at a constant glide speed, whereas glide velocity (V) changes direction along a helical path (grey ribbon) inclined downward at glide angle γ. To turn without sideslip, swifts incline sideways at bank angle μ. Glide angle is determined by the $\cos\mu$ component of lift divided by drag, while the $\sin\mu$ component of lift provides the centripetal force required for turning. **L,** lift; **D,** drag; **W,** weight. **b,** Main forces acting on a swift gliding at a given velocity.

[1]Experimental Zoology Group, Wageningen University, 6709 PG Wageningen, The Netherlands. [2]Department of Marine Biology, Groningen University, 9750 AA Haren, The Netherlands. [3]Department of Aerospace Engineering, Delft University of Technology, 2629 HS Delft, The Netherlands. [4]Department of Theoretical Ecology, Lund University, Ecology Building, SE-223 62 Lund, Sweden. [5]Institute of Biology, Leiden University, 2300 RA Leiden, The Netherlands.
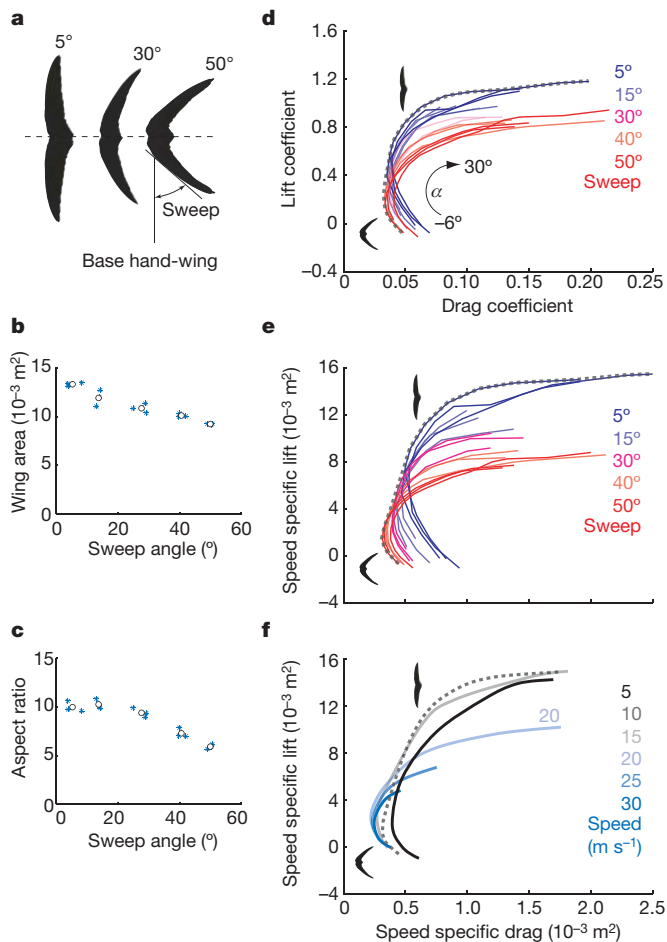
Figure 2 | Morphing swift wings can generate higher lift and lower drag than wings with a fixed geometry. a–c, Adjusting sweep angle (**a**) alters wing area (**b**) and aspect ratio (**c**). Blue stars, individual wings; open circles, average per sweep. **d,** The enveloping polar (grey dashed line) spans a wider range of lift and drag coefficients than the polar for any one sweep (fine coloured lines) (angles of attack α, $-6°$ to $+30°$; glide speed, 10 m s^{-1}). **e,** The combined effect of wing shape and size (speed-specific lift and drag[22]) further enlarges the enveloping polar. **f,** Increasing glide speed shifts enveloping polars to the left, and decreases maximum speed-specific lift.

lift and drag on 15 wing pairs in the Delft low-turbulence wind tunnel (see Methods).

Our experiments show that variable sweep enlarges the aerodynamic performance envelope of swift wings. At a given glide speed, the 'polars' of lift versus drag coefficient (Fig. 2d) for individual sweep angles build up to a much wider enveloping polar: swept wings contribute low drag coefficients at low angles of attack; extended wings contribute high lift coefficients at high angles of attack. The effects of wing shape are 'amplified' by wing area (Fig. 2e plots the same data as polars of speed-specific lift and drag[22]; that is, lift coefficient $\times$ wing area versus drag coefficient $\times$ wing area). The decrease in wing area with increasing sweep further enlarges the enveloping polar for a given glide speed, further widening the performance gap between fixed-shape and morphing wings.

The enveloping polar changes with glide speed (Fig. 2f). With increasing speed, the polar at first maintains its shape and shifts to lower drag values[10,23] because drag coefficient scales with speed to a power less than two at low angles of attack, when flow separation is minimal[24]. Beyond 15 m s^{-1}, the enveloping polar breaks off at lower and lower speed-specific lift values because less swept wings break under the extreme loads; only the more swept wings are left to build up the enveloping polar.

To demonstrate how morphing wings can affect gliding, we translated the above measures of wings' aerodynamics into swift's flight-dynamics—our six figures of merit (see Supplementary equations). Three flight-cost related indices are: (1) glide distance (expressed as the maximum glide ratio[21]), (2) glide duration[21], and (3) turn angle for a given height loss. By maximizing distance or time spent gliding, birds reduce energy expenditure while foraging and roosting. Three indices are agility-related: pursuits and escapes require (4) fast turns (high angular velocity[21]) with (5) a high path curvature[21]; while (6) high horizontal speed (the horizontal component of glide velocity) helps to avoid drift in strong winds. Our discussion of performance maxima ignores combinations of sweep and glide speed that cause diving (glide angles $>45°$, Fig. 1).

Extended wings provide the best glide performance. Five of the six indices (Fig. 3a, b, d–f) reach an absolute maximum with extended wings—characteristic of gliders in general. The cost-related maxima occur between 8 and 15 m s^{-1}. At 10 m s^{-1}, within this optimal speed range, choosing extended over swept wings triples all three turning indices (Supplementary Fig. 2). Nevertheless, swifts sometimes choose higher glide speeds[11,18].

Figure 3 | Morphing improves glide performance of swifts.
a–f, Performance indices for a range of sweep angles (coloured lines) and flight speeds (x axis). Absolute performance maxima occur at low speeds and sweeps in all indices except horizontal velocity, **c.** In two of three straight-flight indices (top row), low sweep delivers superior performance at low speeds while high sweep is superior at high speeds. During turning (bottom row), no crossover from low to high sweep occurs during gliding. Solid lines, glide angle $<45°$; dot, glide angle $=45°$; no dot, wing fails before $45°$ is reached; faint lines, glide angle $>45°$.

During straight glides at higher-than-optimal speeds, high sweep improves aerodynamic wing performance. Consider Fig. 3a: at lower-than-optimal speeds (left of highest peak), extended wings deliver superior glide ratios. As speed increases beyond the optimum, the lines of constant sweep cross, and glide ratio is higher for swept wings. At 20 m s^{-1}, for example, a sweep angle of 50° yields a 70% improvement over extended wings, whereas at speeds below 10 m s^{-1}, extended wings improve glide ratio by as much as 50%. The second cost-related index, glide duration, behaves similarly (Fig. 3b). Unsurprisingly, 'horizontal speed' is the only index that peaks at high glide speeds (Fig. 3c). Although not sensitive to sweep angle at low glide speeds, horizontal speed increases with increasing sweep above 20 m s^{-1}. These results confirm predictions that swept wings improve glide performance at high speed[2,6].

Swept wings can bear higher loads during fast turns. Whereas, during straight gliding at constant speed, the wings bear a load necessarily equal to the bird's mass × 1g, centripetal acceleration increases the load during equilibrium turns. If the bird were to maximize aerodynamic wing performance, it should choose low sweep and low speed: the superior lift force of extended wings is desirable at any speed, in theory[16]. Consistent with this prediction, our measurement-based turning indices show no clear crossovers from low to high sweeps at glide angles below 45° (Fig. 3d–f). Dive performance (glide angle >45°) is severely limited during high-speed turns due to high loads. We measured loads of up to six times the bird's weight (Fig. 4 bottom), and observed two types of structural failure: one extended-wing specimen bent to the point of breaking at 15 m s^{-1}; another started vibrating violently at 15 and 20 m s^{-1}, which ultimately led to failure at the bone. Swept wings do not 'flutter', and they avoid static failure by bending and twisting under lift-loads (Fig. 4 top), which reduces the effective angle of attack at the hand-wing and thereby caps aerodynamic load. Such phenomena are not captured by theoretical or experimental studies using rigid wing models[25].

High sweep maximizes high-speed glide performance, but not by creating strong leading edge vortices (LEVs). LEVs have been observed over model swift wings, and have been proposed to boost lift[25]. Our flow visualizations on real wings confirm the presence
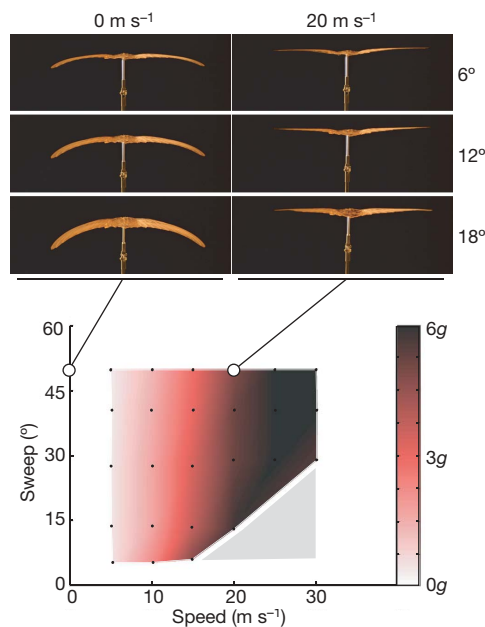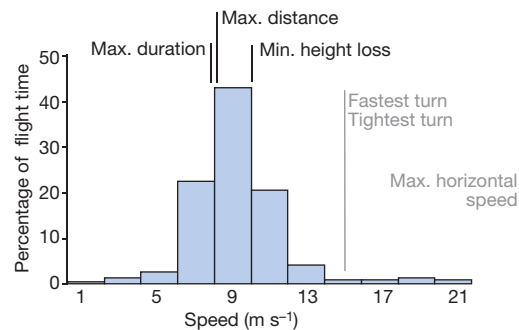
Figure 5 | Swifts roost at glide speeds that minimize energy expenditure. Five of six predicted performance maxima occur within the most commonly observed range of flight speeds during roosting[11]: two of three agility-related maxima (grey), and all cost-related maxima (black), which cluster around swifts' preferred roosting speed of 9 m s^{-1}.

of LEVs at high sweep angles (≥30°) (Supplementary Table 1). However, our force measurements at speeds of 5–30 m s^{-1} show that swept wings always generated less lift than extended wings. Extra lift from LEVs does not compensate for lift lost to the concomitant drop in wing area and aspect ratio (Fig. 2d, e), and to load-induced wing deformations at high speeds (Fig. 4 top). Sweep improves gliding by decreasing drag, rather than by increasing lift.

Our glide model predicts performance-maximizing glide speeds that agree with observations of swift behavioural choices. Glide speeds are readily observable in the field and therefore serve well to validate our semi-empirical glide model. Our model predicts different optimal gliding speeds for maximizing agility (15–25 m s^{-1}) versus cost-effectiveness (8–10 m s^{-1}). The only glide behaviour on which free-flight data have been published is roosting[11,12,18], for which we expect that flight-cost considerations outweigh agility[11,12]. Radar measurements[11] show that roosting swifts indeed flap-glide at speeds centred around 8–10 m s^{-1} (Fig. 5). This agreement of model predictions and field observations validates the analytical step from wing aerodynamic to bird glide performance.

The modification of glide performance achieved by morphing is comparable to the differences between bird species with widely different wing shapes and flight behaviours[15]. Swifts can adjust their wings' maximum lift coefficients between 0.8 and 1.1, which is similar to the full range from thrush (0.8) to nighthawk (1.15) found[15] for extended wings. Therefore, extended wing geometry alone might not be enough to properly evaluate bird gliding performance[8,15]. Birds with an aerial life style, such as swifts, face a wide range of tasks with sometimes conflicting performance goals. To match wing shape to the task at hand, morphing provides birds with a suite of wing geometries from which to choose.

METHODS

Animals. In spring 2005, we received from eight Dutch bird sanctuaries 35 adult swifts that had died after having been brought in. We selected 15 swifts on the basis of wing state and general state.

Wing preparation. We separated 15 wing pairs from the body at the shoulder joint, and manually extended them onto templates for five sweep angles (5°, 15°, 30°, 40°, 50°; Fig. 2a–c). Manually extending wings reliably reproduces wing shape during gliding[26]. Wing pairs were frozen, freeze-dried, then glued together to form a continuous wing surface. They were mounted onto the sting of the balance system and placed in a wind tunnel[27].

Wind tunnel tests. We used the Delft low-turbulence wind tunnel[27] with an octagonal test section of 1.80 × 1.25 m (turbulence levels ≤0.025% until 40 m s^{-1}). We designed a balance system with a resolution of 40,000 steps to measure lift and drag at air speeds between 5 and 30 m s^{-1} (Reynolds number, 12,000–77,000) with an accuracy of at least 3% (Ohaus SP402). We calibrated the balance with a 5 × 5 matrix of weights over the full lift and drag range to account for the system's small nonlinearity. Each force value was sampled at 5 Hz for 10 s at each angle of attack, α (0°→+30°→−6°→0°; Δα = 1.5°, precision < 0.5°). Data from the up- and down-leg of the α cycle were pooled, because hysteresis

Figure 4 | Morphing maintains wing structural integrity at high glide speeds. Bottom: during fast glides, swift wings experience considerable loads (expressed as multiples of g; grey, no measurements). Top: swept wings bend and twist under load to a lower angle of attack (rear view; sweep 50°; α =6°, 12°, 18°).

was negligible below stall. Measurements were corrected for aerodynamic forces of the sting, and for changes in the wings' centre of mass with α. To detect the LEV, we moved a tuft (hair from R.K.) along the wing[28].

Lift–drag polar of swifts. We built the total lift–drag polar of swifts from wing lift and drag, and body drag. Wing polars were built from force measurements across a range of glide speeds (5–30 m s^{-1}), sweeps (5°–50°) and α values ($-6°$ to 30°), excluding high values of α at which the wing stalls—stall was assumed to have occurred when mean lift flattened off and instantaneous lift suddenly became variable. We measured an average body drag coefficient[29] of 0.26 for a frontal area of 913 mm^2 (Supplementary Fig. 1).

Calculation of glide path. To evaluate the correct part of the total polar, we used a body weight of 43 g (ref. 30) to calculate glide paths. By solving for the unknown parameters glide angle, γ, bank angle, μ, and turn radius, R (Fig. 1), we obtained equations of motion that contain only measurable quantities—body mass, m, body weight, W, flight speed, V, lift, L, and drag, D (Supplementary equations (1)):

$$\begin{pmatrix} \gamma \\ \mu \\ R \end{pmatrix} = \begin{pmatrix} \arcsin\left(\frac{D}{W}\right) \\ \arccos\left(\frac{\sqrt{W^2-D^2}}{L}\right) \\ \frac{mV^2}{W^2}\left(\frac{W^2-D^2}{\sqrt{L^2+D^2-W^2}}\right) \end{pmatrix} \quad (1)$$

We then determined all possible helicoidal glide paths[21] (assuming constant glide speed and no side slip).

Calculation of performance maxima. We formulated six figures of merit[21] (Fig. 3): (1) maximum glide distance[21], max(1/tanγ) (=maximum glide ratio), (2) maximum glide duration[21], max(1/(Vsinγ)), (3) maximum turning angle for a given height loss, max (1/(Rtanγ)), (4) maximum angular velocity in a turn[21], max(Vcosγ/R), (5) maximum curvature of the turn path[21], max(1/R), and (6) maximum horizontal component of the flight speed, max(Vcosγ) (Supplementary equations (1)).

We limited the performance analysis by two criteria. First, calculations were only valid for non-zero turning radii (no pure roll around body axis). Second, values for dives (glide angle $\gamma > 45°$) were calculated, but ignored in the search for performance maxima. We linearly interpolated the force coefficients of the two adjacent polars to calculate the 45° dive-angle cut-off speed, straight flight performance maxima, and minimum and maximum flight speeds. Finally, we averaged performance indices per sweep angle to construct Fig. 3 (Supplementary Fig. 2).

Sensitivity analysis. The performance maxima occur at the same wing configuration when we change body drag coefficient (-100%, $+200\%$), body weight ($\pm23\%$) and add the tail's contribution to lift ($\pm20\%$ of wing lift).

Accuracy of roost speed prediction. Air density at the average roosting height[11,18] (1,700 m) was not reported. We measured air density at sea level (1.201 ± 0.005 kg m^{-3}), underestimating optimal speed by maximally 9%. However, roosting flight speed does not correlate strongly with altitude (A.H., unpublished observation based on refs 11, 18).

1. Rosén, M. & Hedenström, A. Gliding flight in a jackdaw. *J. Exp. Biol.* **204**, 1153–1166 (2001).
2. Tucker, V. A. Gliding birds: the effect of variable wing span. *J. Exp. Biol.* **133**, 33–58 (1987).
3. Pennycuick, C. J. Gliding flight of the fulmar petrel. *J. Exp. Biol.* **37**, 330–338 (1960).
4. Newman, B. G. Soaring and gliding flight of the black vulture. *J. Exp. Biol.* **35**, 280–285 (1958).
5. Pennycuick, C. J. Wind-tunnel study of gliding flight in the pigeon *Columba livia*. *J. Exp. Biol.* **49**, 509–526 (1968).
6. Müller, U. K. & Lentink, D. Turning on a dime. *Science* **306**, 1899–1900 (2004).
7. Weiss, P. Wings of change: shape-shifting aircraft ply future skyways. *Sci. News* **164**, 359 (2003).
8. Rayner, J. M. V. in *Current Ornithology* Vol. 5 (ed. Johnston, R. F.) 1–66 (Plenum, New York, 1988).
9. Hoerner, S. F. & Borst, H. V. *Fluid-dynamic Lift* (Hoerner, Bakersfield, California, 1985).
10. Azuma, A. *The Biokinetics of Flying and Swimming* 2nd edn (AIAA Education Series, Reston, Virginia, 2006).
11. Bäckman, J. & Alerstam, T. Confronting the winds: orientation and flight behaviour of roosting swifts, *Apus apus. Proc. R. Soc. Lond. B* **268**, 1081–1087 (2001).
12. Bruderer, B. & Weitnauer, E. Radarbeobachtungen über Zug und Nachtflüge des Mauerseglers (*Apus apus*). *Rev. Suisse Zool.* **79**, 1190–1200 (1972).
13. Parrott, G. C. Aerodynamics of gliding flight of a black vulture *Coragyps atratus. J. Exp. Biol.* **53**, 363–374 (1970).
14. Nachtigall, W. Der Taubenflügel in Gleitflugstellung: geometrische Kenngrössen der Flügelprofile und Luftkrafterzeugung. *J. Ornithol.* **120**, 30–40 (1979).
15. Withers, P. C. An aerodynamic analysis of bird wings as fixed aerofoils. *J. Exp. Biol.* **90**, 143–162 (1981).
16. Thomas, A. L. R. The flight of birds that have wings and tails: variable geometry expands the envelope of flight performance. *J. Theor. Biol.* **183**, 237–245 (1996).
17. Tucker, V. A. & Parrott, G. C. Aerodynamics of gliding flight in a falcon and other birds. *J. Exp. Biol.* **52**, 345–367 (1970).
18. Bäckman, J. & Alerstam, T. Harmonic oscillatory orientation relative to the wind in nocturnal roosting flights of the swift *Apus apus. J. Exp. Biol.* **205**, 905–910 (2002).
19. Lack, D. *Swifts in a Tower* (Methuen, London, 1956).
20. Pennycuick, C. J. Flight of auks (Alcidae) and other Northern sea birds compared with Southern Procellariiformes: ornithodolite observations. *J. Exp. Biol.* **128**, 335–347 (1987).
21. Ruijgrok, G. J. J. *Elements of Airplane Performance* (Delft Univ. Press, Delft, 1994).
22. Vogel, S. *Life in Moving Fluids* 2nd edn (Princeton Univ. Press, Princeton, 1994).
23. Schmitz, F. W. *Aerodynamik des Flugmodells* (C.J.E. Volckmann, Berlin, 1942).
24. Schlichting, H. *Boundary Layer Theory* 7th edn (McGraw-Hill, New York, 1979).
25. Videler, J. J., Stamhuis, E. J. & Povel, G. D. E. Leading-edge vortex lifts swifts. *Science* **306**, 1960–1962 (2004).
26. Hedenström, A. & Rosén, M. Predator versus prey: on aerial hunting and escape strategies in birds. *Behav. Ecol.* **12**, 150–156 (2001).
27. Veldhuis, L. L. M. *Configuration and Propulsion Aerodynamics Research in the Low Speed Aerodynamics Laboratory* [in Dutch] (Internal Report LSW 93–1, Faculty of Aerospace Engineering, Delft University of Technology, Delft, 1993).
28. Bird, J. D. *Tuft-Grid Surveys at Low Speeds for Delta Wings* (Technical Note D-5045, NASA, Hampton, Virginia, 1969).
29. Pennycuick, C. J., Alerstam, T. & Hedenström, A. A new low-turbulence windtunnel for bird flight experiments at Lund University, Sweden. *J. Exp. Biol.* **200**, 1441–1449 (1997).
30. Glutz von Blozheim, U. N. & Bauer, K. M. *Handbuch der Vögel Mitteleuropas* (Akademischer, Wiesbaden, 1980).

LETTERS

Opioids block long-term potentiation of inhibitory synapses

Fereshteh S. Nugent[1]*, Esther C. Penick[1]*† & Julie A. Kauer[1]

Excitatory brain synapses are strengthened or weakened in response to specific patterns of synaptic activation, and these changes in synaptic strength are thought to underlie persistent pathologies such as drug addiction, as well as learning[1]. In contrast, there are few examples of synaptic plasticity of inhibitory GABA (γ-amino-butyric acid)-releasing synapses. Here we report long-term potentiation of GABA$_A$-mediated synaptic transmission (LTP$_{GABA}$) onto dopamine neurons of the rat brain ventral tegmental area, a region required for the development of drug addiction. This novel form of LTP is heterosynaptic, requiring postsynaptic NMDA (N-methyl-D-aspartate) receptor activation at glutamate synapses, but resulting from increased GABA release at neighbouring inhibitory nerve terminals. NMDA receptor activation produces nitric oxide, a retrograde signal released from the postsynaptic dopamine neuron. Nitric oxide initiates LTP$_{GABA}$ by activating guanylate cyclase in GABA-releasing nerve terminals. Exposure to morphine both *in vitro* and *in vivo* prevents LTP$_{GABA}$. Whereas brief treatment with morphine *in vitro* blocks LTP$_{GABA}$ by inhibiting presynaptic glutamate release, *in vivo* exposure to morphine persistently interrupts signalling from nitric oxide to guanylate cyclase. These neuroadaptations to opioid drugs might contribute to early stages of addiction, and may potentially be exploited therapeutically using drugs targeting GABA$_A$ receptors.

High-frequency electrical stimulation (HFS) induced LTP of the inhibitory postsynaptic current (IPSC) onto dopamine-containing but not GABA-containing (dopaminergic and GABAergic, respectively) neurons in ventral tegmental area (VTA) slices (Fig. 1a, b; see also Supplementary Fig. 1). HFS also potentiated GABA$_A$ IPSCs in gramicidin-perforated patch recordings, suggesting that LTP$_{GABA}$ can occur under physiological conditions and is not simply an artefact of chloride-loaded cells (Fig. 1c, d). To our knowledge this is the first demonstration of LTP of GABAergic synapses in the VTA. Blocking GABA$_A$ receptors both *in vivo* and *in vitro* strongly increases dopamine cell firing[2,3], and LTP$_{GABA}$ is therefore a mechanism normally available to increase inhibitory drive and dampen dopamine neuron excitability.

LTP$_{GABA}$ was associated with a decrease in the paired-pulse ratio and an increase in 1/CV², where CV is the coefficient of variation (Fig. 1e, f). Together, these data strongly suggest that LTP$_{GABA}$ is maintained by persistently increased GABA release. We proposed that as in other forms of presynaptically maintained LTP, high-frequency firing of the presynaptic GABAergic afferents would be sufficient both to trigger and maintain LTP$_{GABA}$[4]. If so, postsynaptic manipulations should not influence induction of LTP. Consistent with this prediction, loading the postsynaptic neuron with the G-protein inhibitor guanosine 5′ [β-thio]diphosphate (GDPβS; 200 μM) did not affect LTP$_{GABA}$ (IPSC amplitudes, 178 ± 7% of control values before HFS; n = 5). However, loading the postsynaptic

dopamine cell with a calcium chelator (BAPTA) entirely prevented the induction of LTP$_{GABA}$, suggesting that an increase in postsynaptic Ca²⁺ concentration is required to trigger LTP$_{GABA}$ (Fig. 2a, b). The NMDA receptor (NMDAR) at glutamatergic synapses might be activated during HFS, providing the essential postsynaptic Ca²⁺ for
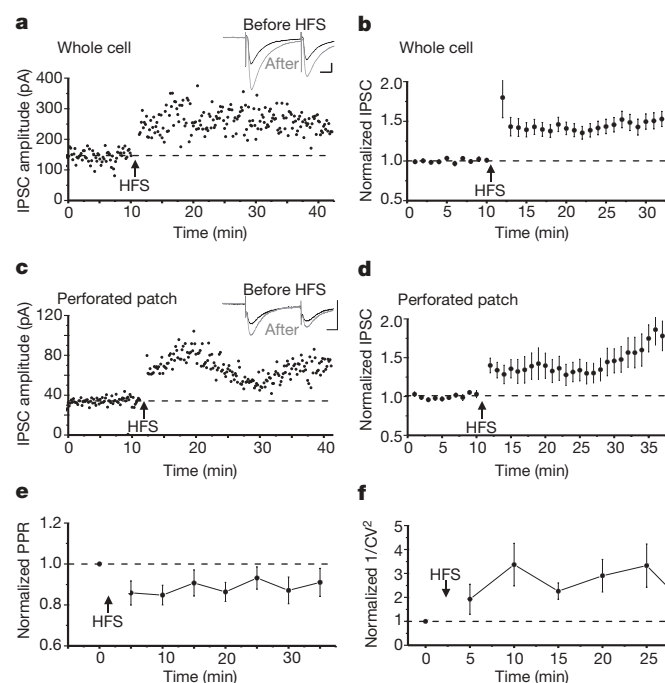

Figure 1 | GABAergic synapses on dopamine neurons are potentiated after HFS. a, LTP$_{GABA}$ in a dopamine neuron using whole-cell recording methods. HFS was delivered at the arrow. Inset: averaged IPSCs before (black) and 25 min after HFS (grey). In this and all figures, ten consecutive IPSCs from each condition were averaged for illustration. Calibration for insets: 10 ms, 50 pA. The dotted line in this and other figures is an approximation of the mean response before HFS. **b**, Average of 71 experiments from dopamine cells. LTP$_{GABA}$ was not triggered in all cells, but data from all cells are included in this and subsequent graphs. The dotted line represents the mean normalized IPSC value before HFS. **c**, LTP$_{GABA}$ in a dopamine neuron using gramicidin-perforated patch recording. Inset: averaged IPSCs before (black) and 25 min after HFS (grey). **d**, Average of eight gramicidin-perforated patch experiments from dopamine cells (LTP$_{GABA}$, 146 ± 4% of pre-HFS values; n = 8). All cells but one exhibited LTP. **e**, LTP$_{GABA}$ in dopamine cells was accompanied by a decrease in the paired-pulse ratio (PPR, IPSC2/IPSC1; n = 35). **f**, LTP$_{GABA}$ was accompanied by an increase in 1/CV² (squared mean IPSC amplitude divided by IPSC variance; n = 35). Error bars in all figures indicate means ± s.e.m.

[1]Brown University, Department of Molecular Pharmacology, Physiology and Biotechnology, Providence, Rhode Island 02912, USA. †Present address: Knox College, 2 East South Street, Galesburg, Illinois 61401, USA.
*These authors contributed equally to this work.

LTP_{GABA}. Consistent with this idea, the NMDAR antagonist D(-)-2-amino-5-phosphonovaleric acid (AP5) prevented LTP_{GABA} (Fig. 2c, d). This form of LTP is therefore heterosynaptic, triggered when glutamate activates NMDARs but influencing neighbouring GABAergic synapses.

To determine whether activation of $GABA_A$ receptors during HFS is also necessary for LTP_{GABA} induction, $GABA_A$ receptors were blocked using bicuculline, and NMDA was briefly bath-applied to activate NMDARs. After washout of both drugs, IPSCs recovered to potentiated levels (Fig. 2e, f). These results clearly demonstrate that $GABA_A$ receptor activation is unnecessary whereas NMDAR activation is sufficient to induce heterosynaptic LTP_{GABA}.

To explain the postsynaptic induction and presynaptic maintenance of LTP_{GABA} requires a retrograde signal generated by an NMDAR-mediated rise in postsynaptic Ca^{2+} yet acting at a presynaptic site. Several lines of evidence supported nitric oxide (NO) as a possible retrograde signal mediating LTP_{GABA}. NO increases GABA release at several central nervous system (CNS) synapses[5–7]. Moreover, the Ca^{2+}/calmodulin-dependent enzyme nitric oxide synthase (NOS) is found in VTA dopamine neurons[8] and is closely associated with NMDA receptors at postsynaptic densities[9]. We therefore tested the effects of a NOS inhibitor, L-NAME (Nω-nitro-L-arginine), and two NO scavengers, haemoglobin and PTIO (2-phenyl-4,4,5,5-tetra-methylimidazoline-1-oxyl 3-oxide), and found that they all potently blocked LTP_{GABA} (Fig. 3a–c). Haemoglobin is too large to cross cell membranes, so its blockade of LTP_{GABA} is consistent with the requirement for NO to pass through the extracellular space, as expected of a retrograde messenger. Furthermore, application of the NO donor, SNAP (S-nitroso-N-acetylpenicillamine), rapidly enhanced GABAergic IPSCs (Fig. 3d) and decreased the paired-pulse ratio (Fig. 3e). Bath application of haemoglobin prevented this effect of SNAP, as expected if the IPSC potentiation by SNAP requires the release of NO (Fig. 3f). Together, these results provide strong support for the idea that NO is produced in dopamine neurons and released extracellularly to enhance GABA release from GABAergic terminals.

At other CNS sites NO facilitates GABA release by activating pre-synaptic guanylate cyclase[5,9]. The guanylate cyclase inhibitor, ODQ (1H-[1,2,4]oxadiazolo[4,3-a]quinoxalin-1-one), prevented LTP induction, implicating the NO–cGMP signalling cascade in LTP_{GABA} (Fig. 3g). A cGMP analogue (pCPT-cGMP) mimicked LTP_{GABA} by robustly increasing GABAergic synaptic currents; furthermore, once the potentiation had plateaued, HFS produced no further synaptic potentiation (Fig. 3h, i). This experiment suggests that cGMP potentiates GABAergic synaptic transmission at this inhibitory synapse using a mechanism shared by LTP_{GABA}. In some systems, endocannabinoids appear to interact with NO[10]; however, in our system pre-treatment with the CB1 receptor antagonist SR141716A (1 μM) did not block LTP_{GABA} (IPSC amplitude, $143 \pm 7\%$ of pre-HFS values; $n = 3$), suggesting that potentiation by NO occurs independently of the CB1 receptor.

Under our experimental conditions, AMPA receptors (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid receptors; AMPARs) were blocked to isolate the IPSC. Under physiological conditions, AMPARs are the most likely source of the depolarization necessary for NMDA receptor activation. We propose that LTP_{GABA} in the VTA results after NMDA channels open, permitting Ca^{2+} to enter the dopamine cell and activate NO synthase. NO released from the dopamine neuron then passes into the extracellular space where it is accessible to haemoglobin, which prevents LTP_{GABA}. NO in turn activates guanylate cyclase, producing cGMP. How cGMP enhances GABA release is not yet known.

This is the first demonstration of heterosynaptic LTP at GABAergic synapses requiring NO as a retrograde messenger. LTP has been reported previously at relatively few inhibitory synapses in the CNS[11–14], and is perhaps best characterized at synapses in the deep cerebellar nuclei where the LTP depends on postsynaptic mechanisms and is not thought to require NO[11,14]. In our experiments, HFS might, in theory, release NO from many dopamine neurons, promoting widespread potentiation of GABAergic release. Experimentally, however, intracellular BAPTA in the recorded neuron prevented LTP_{GABA}; moreover, HFS selectively potentiated synapses on dopamine but not GABAergic neurons. These observations indicate that even NO released during HFS acts quite locally, increasing GABA release primarily onto the neuron in which NO is generated.

The VTA is necessary for the processing of naturally rewarding stimuli such as food and sex, as well as for self-administration of drugs of abuse. Addictive drugs rapidly potentiate excitatory synapses on dopamine neurons[15–19], and GABAergic synapses in the VTA are also altered as a result of exposure in vivo to drugs of abuse[19–22]. As previously reported, morphine and the μ-opioid agonist DAMGO ([D-Ala², N-Me-Phe⁴, Gly⁵-ol]enkephalin) depressed inhibitory synaptic transmission onto VTA dopamine neurons in vitro[3] (Supplementary Fig. 2).

We found that bathing the VTA slice in either morphine or DAMGO also prevented LTP_{GABA} (Fig. 4a). The μ-opioids may block LTP_{GABA} either through presynaptic interactions between opioid signalling pathways and the targets of NO, or by influencing LTP_{GABA} induction upstream of NO production. Application of
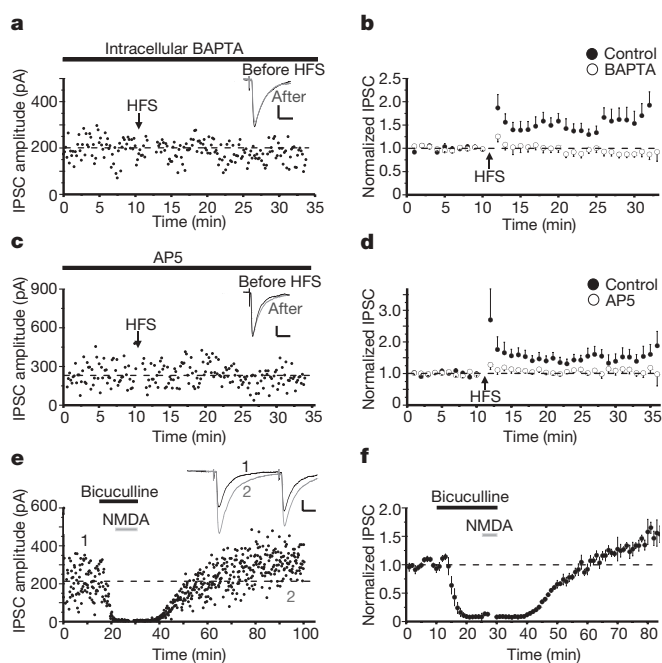
Figure 2 | Chelating postsynaptic Ca^{2+} or blocking NMDARs prevents LTP_{GABA}, whereas $GABA_A$ receptor activation is not required.
a, Intracellular BAPTA (30 mM) blocked LTP_{GABA}. BAPTA was allowed to perfuse the neuron for at least 20 min before HFS. Inset: averaged IPSCs before (black) or 25 min after HFS (grey). Calibration for insets: 10 ms, 50 pA. **b**, Averaged experiments with and without 30 mM BAPTA in the pipette solution (control LTP, filled symbols: $166 \pm 12\%$ of pre-HFS values, $n = 13$; BAPTA cells, open symbols: $95 \pm 5\%$ of pre-HFS values, $n = 14$). **c**, The NMDAR antagonist AP5 (50 μM) blocked LTP_{GABA}. Inset: averaged IPSCs before (black) or 25 min after HFS (grey). **d**, Averaged experiments with and without AP5 present (control LTP, filled symbols: $148 \pm 10\%$ of pre-HFS values, $n = 19$; AP5 cells, open symbols: $108 \pm 7\%$ of pre-HFS values, $n = 13$). **e**, $GABA_A$ receptor activation is not required for LTP_{GABA}. Bicuculline (30 μM) was used to block the IPSC and then 10 μM NMDA was bath-applied; the recording mode was switched to current clamp. Two minutes after NMDA elicited depolarization, both drugs were washed from the bath and the cell was voltage clamped. DNQX (10 μM) and strychnine (1 μM) were present throughout. Inset: averaged IPSCs before (1, black) and 60 min after washout of NMDA and bicuculline (2, grey). **f**, Averaged data from four experiments using the protocol outlined in **e** (IPSC amplitudes after washout of bicuculline, $146 \pm 4\%$ of pre-HFS values; $n = 4$). Bicuculline treatment alone did not produce potentiation.

the NO donor potentiated IPSCs, even with DAMGO present (Fig. 4b), suggesting that opioids do not prevent NO efficacy at the level of the presynaptic terminals. Instead, μ-opioids inhibit glutamate release upstream of NO production. NMDAR synaptic currents were depressed by approximately 40% by the μ-opioid receptor agonist. When we depressed NMDAR currents experimentally by approximately 40% using an NMDAR antagonist, LTP_{GABA} was blocked (Supplementary Fig. 3). This mechanism may contribute to the early effects of μ-opioids on VTA circuit function *in vivo*[23].

In vivo treatment with morphine potentiates VTA glutamatergic synapses 24 h later[17]. We therefore asked whether *in vivo* exposure to morphine also alters plasticity of GABAergic synapses. Sprague–Dawley rats were treated either with morphine or with saline. After HFS in slices cut 24 h after morphine injections, there was no LTP_{GABA} (Fig. 4c–e), whereas slices from saline-treated animals exhibited normal LTP_{GABA}. These experiments provide direct evidence that inhibitory synapses in the VTA can no longer be potentiated 24 h after a single *in vivo* exposure to morphine.

Instead of a persistent block of LTP_{GABA}, saturation of LTP_{GABA} might preclude further potentiation, as reported 24 h after ethanol exposure[22]. In contrast, in our experiments basal IPSCs from morphine- and saline-treated rats had indistinguishable paired-pulse ratio values (Fig. 4f). These data indicate that the release probability of GABAergic synapses was not increased after *in vivo* morphine

exposure, as expected if LTP_{GABA} were already maximally activated. Instead, morphine produced a long-lasting inability to generate LTP_{GABA}.

We next explored the cellular mechanisms underlying this phenomenon using the NO donor and cGMP analogue. Whereas synapses from saline-treated rats were potentiated after exposure to SNAP, synapses from morphine-treated rats were unaffected (Fig. 4g). Thus, 24 h after morphine exposure, NO cannot potentiate GABAergic synapses. In contrast, when we applied the cGMP analogue, synapses from morphine-treated rats were potentiated (Fig. 4h). These data are consistent with a model in which *in vivo* morphine exposure interrupts the signalling between NO and cGMP generation, perhaps at the level of guanylate cyclase. This mechanism is entirely different from that mediating the acute blockade of LTP_{GABA} *in vitro*, and suggests that neuroadaptations in GABAergic terminals persist for hours after exposure to morphine.

The inability of GABAergic synapses to potentiate 24 h after *in vivo* morphine exposure may promote LTP of glutamatergic synapses, observed as an increased AMPAR/NMDAR ratio 24 h after morphine[17], cocaine, amphetamine, nicotine or ethanol exposure[16–19]. Repeated cocaine exposure over a period of days also gradually diminishes $GABA_A$ receptor-mediated inhibition in the VTA[19]. Here we show that even a single morphine exposure produces a
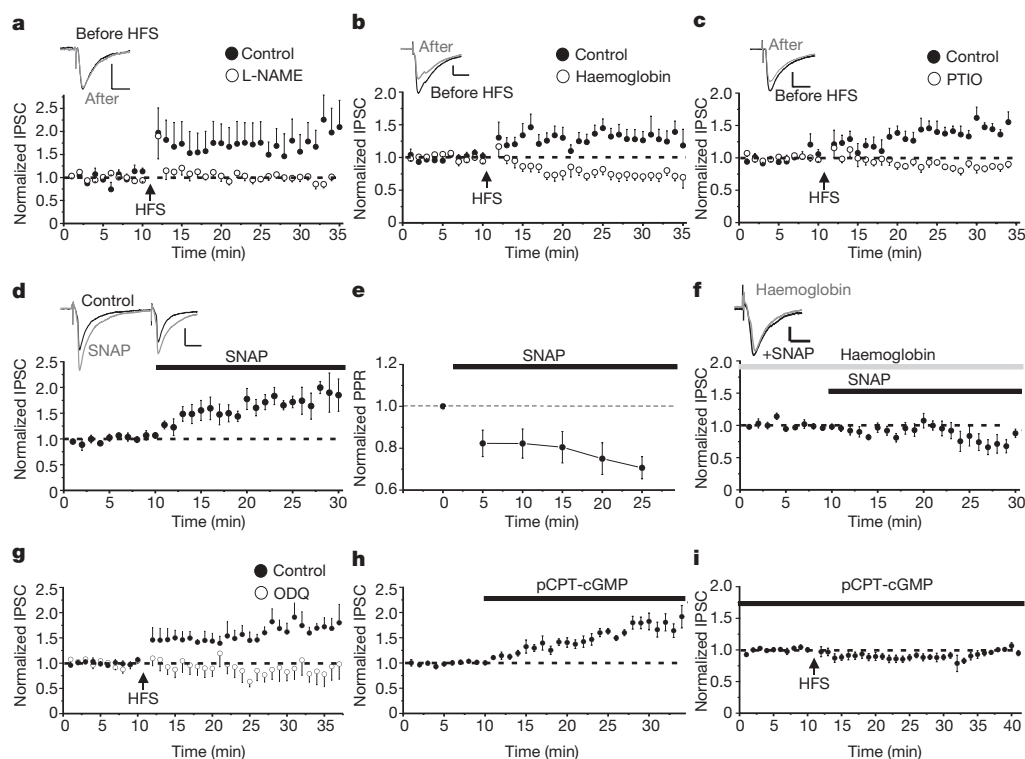
Figure 3 | LTP_{GABA} requires NO–cGMP signalling. a, The NOS inhibitor L-NAME (200 μM) blocked LTP_{GABA} (control LTP, filled symbols: 162 ± 14% of pre-HFS values, $n = 7$; L-NAME cells, open symbols: 98 ± 7% of pre-HFS values, $n = 14$). Insets: averaged IPSCs in L-NAME before (black) and 25 min after HFS (grey). Calibration for insets: 10 ms, 50 pA. **b**, Bath-applied haemoglobin (100 μM), an NO scavenger, prevented LTP_{GABA} (control LTP, filled symbols: 136 ± 6% of pre-HFS values, $n = 7$; haemoglobin cells, open symbols: 74 ± 4% of pre-HFS values, $n = 9$). Inset as for **a** but with haemoglobin instead of L-NAME. **c**, Another NO scavenger, PTIO (300 μM), also blocked LTP_{GABA} (control LTP, filled symbols: 142 ± 11% of pre-HFS values, $n = 7$; PTIO cells, open symbols: 86 ± 4% of pre-HFS values, $n = 7$). Inset as for **a** but with PTIO instead of L-NAME. **d**, The NO donor SNAP (200 μM) potentiated GABAergic IPSCs without HFS (171 ± 8% of pre-drug values; $n = 6$). Inset: averaged IPSCs before (black) and after 10 min in SNAP (grey). **e**, The SNAP-induced potentiation

was accompanied by a decrease in the paired-pulse ratio. Five-minute blocks of data are shown (paired-pulse ratio before SNAP, 1.06 ± 0.07; in SNAP, 0.81 ± 0.08). **f**, Haemoglobin (100 μM) blocked the enhancement of IPSCs by 200 μM SNAP (IPSC amplitudes, 91 ± 9% of pre-drug values; $n = 6$). Haemoglobin was introduced at least 20 min before the addition of SNAP. Inset: averaged IPSCs in haemoglobin (black) and 10 min after SNAP (grey). **g**, The guanylate cyclase inhibitor ODQ (10 μM) blocks LTP_{GABA}, implicating the NO–cGMP signalling cascade (control LTP, filled symbols: 160 ± 5.7% of pre-HFS values, $n = 6$; ODQ-treated cells, open symbols: 81 ± 4% of pre-HFS values, $n = 8$). **h**, pCPT-cGMP (100 μM; 8-(*p*-chlorophenylthio)-cGMP), a cGMP analogue, potentiated IPSCs (169 ± 5% of pre-drug values; $n = 11$). **i**, After the IPSC in 100 μM pCPT-cGMP reached a stable potentiated level, HFS was delivered. pCPT-cGMP prevented further potentiation of IPSCs by HFS (89 ± 0.5 of pre-HFS values; $n = 6$).
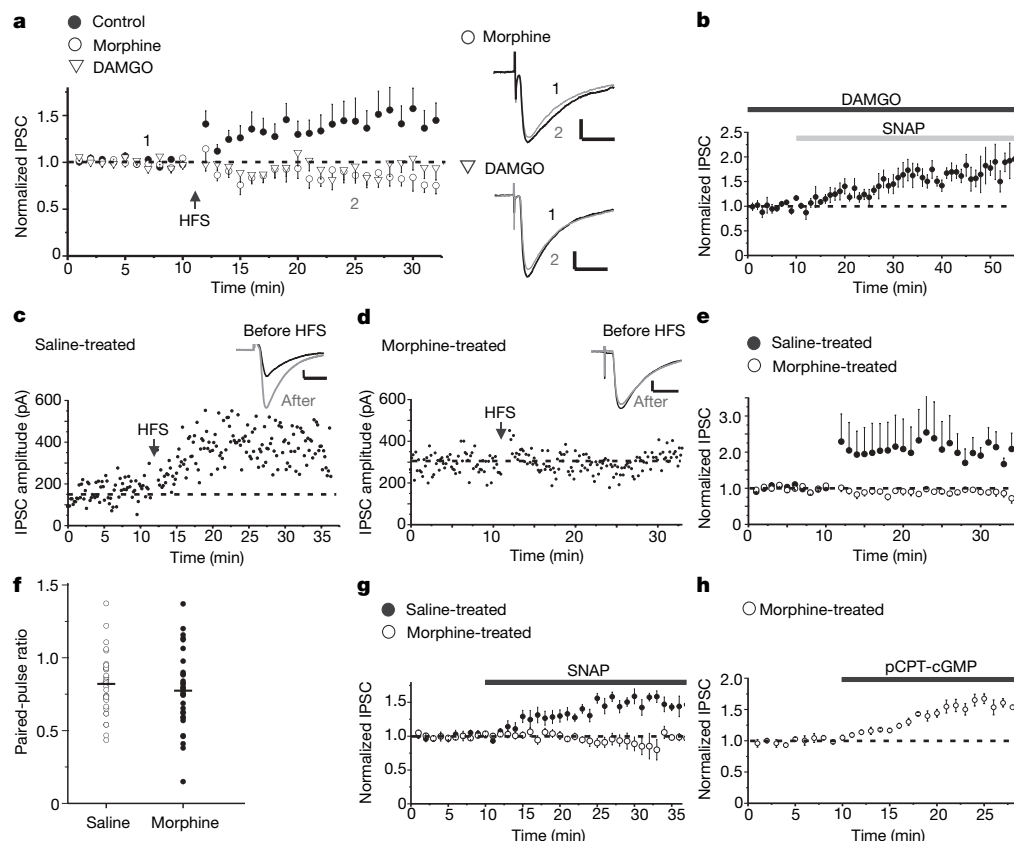
Figure 4 | Opioids block LTP$_{GABA}$ both *in vitro* and *in vivo* by distinct mechanisms. a, Bath-applied μ-opioids block LTP$_{GABA}$. Morphine (1 μM) or DAMGO (1 μM) was applied and HFS was delivered at the arrow (control cells, filled circles: 148 ± 8% of pre-HFS values, n = 18; morphine cells, open circles: 85 ± 2% of pre-HFS values, n = 10; DAMGO cells, open triangles: 93 ± 10% of pre-HFS values, n = 15). Insets: averaged IPSCs before (1, black) and 15 min after HFS (2, grey) from a single morphine (top) or DAMGO (bottom) experiment. Calibration for insets: 10 ms, 50 pA. **b**, Acutely applied opioids do not prevent NO-induced potentiation of GABA$_A$ IPSCs. DAMGO (1 μM) was bath-applied at least 10 min before application of SNAP (200 μM). Once the IPSC had reached a new stable level in DAMGO-treated cells, SNAP was bath-applied (IPSC amplitude in SNAP, 159 ± 3.5% of pre-drug values; n = 4). **c, d**, *In vivo* exposure to morphine prevents LTP$_{GABA}$ in slices prepared 24 h later. Example experiments and IPSCs (insets) from slices from a saline-injected animal (**c**) and a morphine-injected animal (**d**). **e**, Averaged

experiments from slices prepared 24 h after either saline or morphine injection (saline, filled symbols: 196 ± 20% of pre-HFS values, n = 10; morphine, open symbols: 91 ± 4% of pre-HFS values, n = 11). **f**, Paired-pulse ratios are similar in slices 24 h after *in vivo* injection with either saline (open circles, n = 31) or morphine (filled circles, n = 32). Each symbol represents the paired-pulse ratio value for one animal (mean basal paired-pulse ratio values (black horizontal bars) from saline-treated rats: 0.82 ± 0.04, n = 31; morphine-treated rats: 0.77 ± 0.05, n = 32). **g**, In slices from morphine-treated rats, NO no longer potentiates GABA$_A$-mediated IPSCs. A two-hundred micromolar concentration of SNAP potentiated IPSCs in slices cut 24 h after saline injection (filled symbols, 151 ± 2.7% of pre-drug values; n = 8), but had no effect in slices from morphine-treated rats (open symbols, 92 ± 1% of pre-drug values; n = 10). **h**, In contrast, the cGMP analogue (100 μM pCPT-cGMP) potentiated GABA$_A$-mediated IPSCs in slices from morphine-treated rats (156 ± 3.5% of pre-drug values; n = 4).

similar persistent loss of normal GABAergic function. Morphine's reinforcing and stimulant effects increase after a single dose of morphine *in vivo*[24,25], and also after overexpression of the AMPAR subunit GluR1 in the VTA[26]. The loss of normal inhibitory control we observe coupled with potentiation of excitatory synapses may represent neuroadaptations that increase the incentive properties of morphine[27]. These early adaptations may synergize to increase vulnerability to addiction during subsequent drug exposures. Moreover, GABA$_A$ receptor modulators modify the abuse-related effects of addictive drugs[28–30], and targeting these receptors may prove to be an effective therapeutic strategy.

METHODS

See Supplementary Information for detailed methods.

Electrophysiological recordings. Horizontal midbrain slices (250 μm thick) were prepared from Sprague–Dawley rats (16–21 days old) anaesthetized using halothane or isoflurane. Midbrain slices were continuously perfused with artificial cerebrospinal fluid (ACSF). In all experiments recording IPSCs, 6,7-dinitroquinoxaline-2,3-dione (DNQX, 10 μM) and strychnine (1 μM) were added to block AMPA- and glycine-mediated synaptic currents, respectively. VTA neurons were voltage clamped at −70 mV except during HFS, when neurons were held in current-clamp mode. IPSCs were stimulated with a bipolar stainless steel

electrode 200–500 μm rostral to the recording site in the VTA. LTP$_{GABA}$ was induced by stimulating afferents at 100 Hz for 1 s, the train repeated twice 20 s apart (HFS). For perforated patch recordings, patch pipettes were filled with gramicidin-KCl-based internal solution (gramicidin 10 μg ml^{-1}). Drugs were added directly to the ACSF perfusing the slice chamber at known concentrations for at least 15 min before HFS. Control drug-free experiments were interleaved with experiments using drug application, and individual graphs illustrate the appropriate set of controls.

Analysis. Results are expressed as means ± s.e.m. Significance was determined using a Student's *t*-test with significance level of $P < 0.05$. Levels of LTP are reported as averaged IPSC amplitudes for 5 min just before LTP induction compared with averaged IPSC amplitudes during the 5-min period from 15–20 min after HFS. Paired-pulse ratios (50-ms interstimulus interval) and coefficient of variation were measured over 5-min epochs of 30 IPSCs each.

Treatment with morphine. Rats (16–21 days old) were maintained on a 12 h light/dark cycle and provided food and water ad libitum. They were injected intraperitoneally with either 10 mg kg^{-1} morphine or a comparable volume of saline, placed in a new cage for 2 h, and then returned to the home cage. They were killed for brain slice preparation 24 h after injection.

1. Malenka, R. C. & Bear, M. F. LTP and LTD: an embarrassment of riches. *Neuron* **44**, 5–21 (2004).

2. Yim, C. Y. & Mogenson, G. Electrophysiological studies of neurons in the ventral tegmental area of Tsai. *Brain Res.* **181**, 301–313 (1980).
3. Johnson, S. W. & North, R. A. Opioids excite dopamine neurons by hyperpolarization of local interneurons. *J. Neurosci.* **12**, 483–488 (1992).
4. Zalutsky, R. A. & Nicoll, R. A. Comparison of two forms of long-term potentiation in single hippocampal neurons. *Science* **248**, 1619–1624 (1990).
5. Stern, J. E. & Ludwig, M. NO inhibits supraoptic oxytocin and vasopressin neurons via activation of GABAergic synaptic inputs. *Am. J. Physiol.* **28**, R1815–R1822 (2001).
6. Li, D. P., Chen, S. R. & Pan, H. L. Nitric oxide inhibits spinally projecting paraventricular neurons through potentiation of presynaptic GABA release. *J. Neurophysiol.* **88**, 2664–2674 (2002).
7. Yu, D. & Eldred, W. D. Nitric oxide stimulates γ-aminobutyric acid release and inhibits glycine release in retina. *J. Comp. Neurol.* **483**, 278–291 (2005).
8. Klejbor, I., Domaradzka-Pytel, B., Ludkiewicz, B., Wojcik, S. & Morys, J. The relationships between neurons containing dopamine and nitric oxide synthase in the ventral tegmental area. *Folia Histochem. Cytobiol.* **42**, 83–87 (2004).
9. Brenman, J. E. & Bredt, D. S. Synaptic signaling by nitric oxide. *Curr. Opin. Neurobiol.* **7**, 374–378 (1997).
10. Safo, P. K., Cravatt, B. F. & Regehr, W. G. Retrograde endocannabinoid signaling in the cerebellar cortex. *Cerebellum* **5**, 134–145 (2006).
11. Aizenman, C. D., Manis, P. B. & Linden, D. J. Polarity of long-term synaptic gain change is related to postsynaptic spike firing at a cerebellar inhibitory synapse. *Neuron* **21**, 827–835 (1998).
12. Shew, T., Yip, S. & Sastry, B. R. Mechanisms involved in tetanus-induced potentiation of fast IPSCs in rat hippocampal CA1 neurons. *J. Neurophysiol.* **83**, 3388–3401 (2001).
13. Komatsu, Y. GABA$_B$ receptors, monoamine receptors, and postsynaptic inositol trisphosphate-induced Ca^{2+} release are involved in the induction of long-term potentiation at visual cortical inhibitory synapses. *J. Neurosci.* **16**, 6342–6352 (1996).
14. Ouardouz, M. & Sastry, B. R. Mechanisms underlying LTP of inhibitory synaptic transmission in the deep cerebellar nuclei. *J. Neurophysiol.* **84**, 1414–1421 (2000).
15. Mansvelder, H. D. & McGehee, D. S. Long-term potentiation of excitatory inputs to brain reward areas by nicotine. *Neuron* **27**, 349–357 (2000).
16. Ungless, M. A., Whistler, J. L., Malenka, R. C. & Bonci, A. Single cocaine exposure *in vivo* induces long-term potentiation in dopamine neurons. *Nature* **411**, 583–587 (2001).
17. Saal, D., Dong, Y., Bonci, A. & Malenka, R. C. Drugs of abuse and stress trigger a common synaptic adaptation in dopamine neurons. *Neuron* **37**, 577–582 (2003).
18. Faleiro, L. J., Jones, S. & Kauer, J. A. Rapid synaptic plasticity of glutamatergic synapses on dopamine neurons in the ventral tegmental area in response to acute amphetamine injection. *Neuropsychopharmacology* **29**, 2115–2125 (2004).
19. Liu, Q. S., Pu, L. & Poo, M. M. Repeated cocaine exposure *in vivo* facilitates LTP induction in midbrain dopamine neurons. *Nature* **437**, 1027–1031 (2005).
20. Bonci, A. & Williams, J. T. A common mechanism mediates long-term changes in synaptic transmission after chronic cocaine and morphine. *Neuron* **16**, 631–639 (1996).
21. Shoji, Y., Delfs, J. & Williams, J. T. Presynaptic inhibition of GABA$_B$-mediated synaptic potentials in the ventral tegmental area during morphine withdrawal. *J. Neurosci.* **19**, 2347–2355 (1999).
22. Melis, M., Camarini, R., Ungless, M. A. & Bonci, A. Long-lasting potentiation of GABAergic synapses in dopamine neurons after a single *in vivo* ethanol exposure. *J. Neurosci.* **22**, 2074–2082 (2002).
23. Margolis, E. B. *et al.* κ opioids selectively control dopaminergic neurons projecting to the prefrontal cortex. *Proc. Natl Acad. Sci. USA* **103**, 2938–2942 (2006).
24. Schulteis, G., Heyser, C. J. & Koob, G. F. Opiate withdrawal signs precipitated by naloxone following a single exposure to morphine: potentiation with a second morphine exposure. *Psychopharmacology* **129**, 56–65 (1997).
25. Vanderschuren, L. J., De Vries, T. J., Wardeh, G., Hogemboom, F. A. & Schoffelmeer, A. N. A single exposure to morphine induces long-lasting behavioural and neurochemical sensitization in rats. *Eur. J. Neurosci.* **14**, 1533–1538 (2001).
26. Carlezon, W. A. Jr *et al.* Sensitization to morphine induced by viral-mediated gene transfer. *Science* **277**, 812–814 (1997).
27. Williams, J. T., Christie, M. J. & Manzoni, O. Cellular and synaptic adaptations mediating opioid dependence. *Physiol. Rev.* **81**, 299–343 (2001).
28. Stromberg, M. F., Mackler, S. A., Volpicelli, J. R., O'Brien, C. P. & Dewey, S. L. The effect of γ-vinyl-GABA on the consumption of concurrently available oral cocaine and ethanol in the rat. *Pharmacol. Biochem. Behav.* **68**, 291–299 (2001).
29. Brodie, J. D., Figueroa, E. & Dewey, S. L. Treating cocaine addiction: from preclinical to clinical trial experience with γ-vinyl GABA. *Synapse* **50**, 261–265 (2003).
30. Barrett, A. C., Negus, S. S., Mello, N. K. & Caine, S. B. Effect of GABA agonists and GABA-A receptor modulators on cocaine- and food-maintained responding and cocaine discrimination in rats. *J. Pharmacol. Exp. Ther.* **315**, 858–871 (2005).

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

UDP acting at P2Y$_6$ receptors is a mediator of microglial phagocytosis

Schuichi Koizumi[1,2]*, Yukari Shigemoto-Mogami[1]*, Kaoru Nasu-Tada[1], Yoichi Shinozaki[1,3], Keiko Ohsawa[4], Makoto Tsuda[3], Bhalchandra V. Joshi[5], Kenneth A. Jacobson[5], Shinichi Kohsaka[4] & Kazuhide Inoue[3]

Microglia, brain immune cells, engage in the clearance of dead cells or dangerous debris, which is crucial to the maintenance of brain functions. When a neighbouring cell is injured, microglia move rapidly towards it or extend a process to engulf the injured cell. Because cells release or leak ATP when they are stimulated[1,2] or injured[3,4], extracellular nucleotides are thought to be involved in these events. In fact, ATP triggers a dynamic change in the motility of microglia *in vitro*[5,6] and *in vivo*[3,4], a previously unrecognized mechanism underlying microglial chemotaxis[5,6]; in contrast, microglial phagocytosis has received only limited attention. Here we show that microglia express the metabotropic P2Y$_6$ receptor whose activation by endogenous agonist UDP triggers microglial phagocytosis. UDP facilitated the uptake of microspheres in a P2Y$_6$-receptor-dependent manner, which was mimicked by the leakage of endogenous UDP when hippocampal neurons were damaged by kainic acid *in vivo* and *in vitro*. In addition, systemic administration of kainic acid in rats resulted in neuronal cell death in the hippocampal CA1 and CA3 regions, where increases in messenger RNA encoding P2Y$_6$ receptors that colocalized with activated microglia were observed. Thus, the P2Y$_6$ receptor is upregulated when neurons are damaged, and could function as a sensor for phagocytosis by sensing diffusible UDP signals, which is a previously unknown pathophysiological function of P2 receptors in microglia.

Microglia express several functional P2 receptors, and their P2X$_4$, P2X$_7$ and P2Y$_{12}$ receptors have already been described in relation to their physiological and pathophysiological consequences[5–9]. To investigate the expression of mRNAs for P2 receptors that are at a higher concentration in cultured rat microglia, we conducted reverse-transcriptase-mediated polymerase chain reaction (RT–PCR) analysis with complementary DNA coding for P2Y and P2X receptors (Fig. 1a). In accordance with previous reports[5–9], microglia expressed mRNAs encoding P2X$_4$, P2X$_7$ and P2Y$_{12}$ receptors. However, we found unexpectedly that cultured rat microglia expressed a large amount of mRNA coding for P2Y$_6$ receptors, which was also confirmed by western blotting for the expression of P2Y$_6$ receptor protein (Fig. 1b). The P2Y$_6$ receptor is coupled to the activation of phospholipase C (PLC), leading to the production of inositol 1,4,5-trisphosphate (InsP$_3$) and the release of Ca^{2+} from InsP$_3$-receptor-sensitive stores[10,11]. We therefore examined changes in the intracellular Ca^{2+} concentration ([Ca^{2+}]$_i$) in microglia and found that the P2Y$_6$ receptor agonist UDP evoked increases in [Ca^{2+}]$_i$ in a concentration-dependent manner, and it also increased the fraction of the UDP-responsive cells (Supplementary Fig. 1a). The elevations in [Ca^{2+}]$_i$ induced by 100 μM

UDP were significantly inhibited by the PLC inhibitor U73122, the Ca^{2+}-ATPase inhibitor in sarcoplasmic/endoplasmic reticulum thapsigargin, and the membrane-permeable InsP$_3$ receptor inhibitor xestospongin C, but were little affected by pertussis toxin (Supplementary Fig. 1b). The UDP-evoked [Ca^{2+}]$_i$ increases in microglia were significantly inhibited by reactive blue 2 (RB2), known as a potent P2Y$_6$ antagonist[11], suramin, which inhibits P2Y$_6$ receptor at higher concentrations, the diisothiocyanate derivative MRS2578, which is a selective antagonist of the P2Y$_6$ receptor[12], and an antisense oligonucleotide (AS) for P2Y$_6$ receptors, but not by a random-sequence oligonucleotide (R-oligo) (Fig. 1c). All these data show that rat microglia express functional P2Y$_6$ receptors by which UDP mobilizes Ca^{2+}.
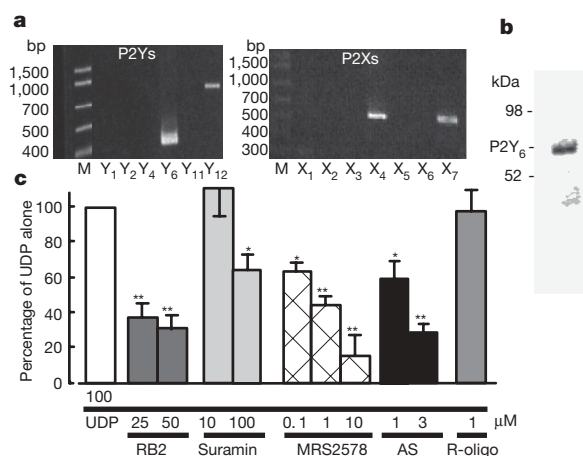
Figure 1 | Expression of P2Y$_6$ receptor and UDP-evoked increase in [Ca^{2+}]$_i$ in cultured microglia. a, RT–PCR analysis of the expression of mRNAs for P2Y$_6$, P2Y$_{12}$, P2X$_4$ and P2X$_7$ receptors in microglial cells. **b**, Expression of P2Y$_6$ receptor protein confirmed by western blotting analysis. **c**, Effects of various chemicals on the increase in [Ca^{2+}]$_i$ (measured as the change in ratio of fluorescence at 340 nm to that at 380 nm) evoked by 100 μM UDP in microglia. The maximum increase in Fura-2 fluorescence evoked by 100 μM UDP was considered as the control response, and values are expressed as a percentage of control. Data show means and s.e.m. for 24–36 cells obtained from at least three independent experiments. Significant differences from the response to UDP alone: asterisk, $P < 0.05$; two asterisks, $P < 0.01$ (Student's *t*-test).

[1]Division of Pharmacology, National Institute of Health Sciences, 1-18-1 Kamiyoga, Setagaya, Tokyo 158-8501, Japan. [2]Department of Pharmacology, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, 1110 Shimokato, Chuo, Yamanashi 409-3893, Japan. [3]Department of Molecular and System Pharmacology, Graduate School of Pharmaceutical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi, Fukuoka 812-8582, Japan. [4]Department of Neurochemistry, National Institute of Neuroscience, 4-1-1 Ogawahigashi, Kodaira, Tokyo 187-8502, Japan. [5]Molecular Recognition Section, Laboratory of Bioorganic Chemistry, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0810, USA.
*These authors contributed equally to this work.

Morphogenesis, cell movement and phagocytosis are driven by dynamic reorganization of the actin cytoskeleton[13,14]. We showed previously that activation of P2Y$_{12/13}$ receptors, another microglial G-protein-coupled receptor, resulted in membrane ruffling and chemotaxis in microglia[5,6], and therefore we sought first to determine whether the P2Y$_6$-receptor-mediated signals affect the cell movement of microglia. Membrane ruffles are structures that are found primarily at the front edges of migrating cells[15]. To determine whether P2Y$_6$ activation stimulates microglial chemotaxis, cells were stimulated with either UDP or ATP. Neither lamellipodia-like membrane ruffles (Fig. 2a left) nor chemotaxis (Fig. 2b left) were observed when stimulated with UDP, whereas ATP produced both responses (Fig. 2a right and Fig. 2b right). However, UDP caused actin reorganization and formed aggregates of F-actin in the interior of the cells (Fig. 2a left, arrows). On stimulation with UDP (100 μM), microglia rapidly changed their morphology (Supplementary Fig. 2a); namely, to microglial processes with filopodia-like protrusions (arrows) and phagosome-like vacuoles (arrowheads). A crown-like circular structure rich in F-actins, termed the 'phagocytotic cup'[16], was also observed around the zymosan particles (Supplementary Fig. 2b, red). We speculated that UDP somehow regulates the morphogenesis of microglia, which may be involved in microglial endocytotic activities such as pinocytosis, macropinocytosis and phagocytosis. Phagocytosis is one of the most important physiological functions of microglia[17] and is the process activating the uptake of larger particles (more than 0.5 μm) by actin-based mechanisms. We investigated the UDP-evoked phagocytosis process by time-lapse videomicroscopy and flow cytometry (fluorescence-activated cell sorting; FACS)-based assay. When stimulated with 100 μM UDP, microglia rapidly phagocytosed fluorescent zymosan particles (green) (Fig. 2c, see also Supplementary Video). A quantitative phagocytosis assay by FACS shows that UDP induced the phagocytosis of latex beads in a concentration-dependent fashion (5–1,000 μM) in a 20-min incubation period (Fig. 2d). GDP (100–1,000 μM), a weak agonist of the P2Y$_6$ receptor, caused a slight uptake of microspheres (at 100 μM this was 49.7 ± 8.6% of UDP alone; $n = 4$) but ADP, also known as a weak partial agonist of the mouse P2Y$_6$ receptor, failed to stimulate the uptake (at 100 μM it was 0.3 ± 2.3% of UDP alone; $n = 4$). This is in good agreement with the previous finding that ADP does not activate rat P2Y$_6$ receptors[18]. The phagocytosis induced by 100 μM UDP was significantly inhibited by 30–100 μM RB2, a higher concentration of suramin (300 μM) and MRS2578 (0.01–3 μM), and was nearly abolished by P2Y$_6$ AS (Fig. 2e; see also Supplementary Fig. 2c, d). Recent reports indicate the existence of functional cross-talk between the nucleotides and cysteinyl leukotrienes (CysLTs, for example LTD4) in orchestrating inflammatory responses[19], indicating that some nucleotides may reveal their functions by means of a CysLT receptor (CysLTR). Microglia express a functional CysLT1R, whose activation by LTD4 resulted in an increase in $[Ca^{2+}]_i$ in microglia (Supplementary Fig. 3a). Thus, UDP acting on CysLT1R may reveal various microglial responses. However, MRS2578, a selective P2Y$_6$ receptor antagonist, did not block the LTD4-evoked Ca^{2+} responses in CysLT1R-transfected Chinese hamster ovary cells (Supplementary Fig. 3b) at a dose that inhibited the UDP-evoked increase in $[Ca^{2+}]_i$ and phagocytosis in microglia (Figs 1c and 2e). In addition, 1 μM LTD4 did not induce phagocytosis in microglia (Supplementary Fig. 3c, 4.8 ± 4.2% of that with 100 μM UDP alone; $n = 3$). All these findings suggest that the contribution of the CysLT1R to the UDP-evoked phagocytosis in microglia is negligible. Taken together, these data strongly suggest that rat microglial P2Y$_6$ receptors are coupled with phagocytic functions. The UDP-evoked phagocytosis was inhibited by 1 μM thapsigargin, the protein kinase C inhibitor staurosporin at 5 μM, and 10 μM U73122 (see Supplementary Fig. 4), indicating that activation of the P2Y$_6$ receptor seems to trigger phagocytosis through the pathway(s) mediated by PLC-linked Ca^{2+} and protein kinase C.

Because phagocytes remove dead or damaged cells, debris and invading pathogens, recognition is the first step in phagocytosis. It is initiated by activation of the phagocytosis-promoting receptors such as Fc receptors and complement receptors[20]. In the central nervous system, microglia possess these receptors and remove amyloid-β, a key molecule in Alzheimer's disease, and attenuate Alzheimer's disease-like pathology[21]. With regard to apoptotic cells, microglia may also remove such cells by recognizing so-called 'eat-me' signals[20]. However, in the present study we used non-opsonized zymosan (Fig. 2c) and latex beads (Fig. 2d, e), which were not recognized by opsonin-dependent receptors such as Fc receptors, complement receptors or vitronectin receptors. Phagocytosis-promoting receptors also include opsonin-independent ones such as β$_1$-integrins, mannose receptors, scavenger receptors and phosphatidylserine receptors[13]; in fact, microglia expressed all these receptors (Supplementary Fig. 5a–e, cell lysates). Among these receptors, β$_1$-integrin was detected as a bead-associated protein that was slightly increased on stimulation with UDP (Supplementary Fig. 5a, bead-associated) and localized at membrane ruffle-like or phagocytic cup-like structures (see also Supplementary Fig. 2b), to which fluorescent microspheres were attached (Supplementary Fig. 5f). However, we do not know whether β$_1$-integrin itself binds or recognizes the microspheres. β$_1$-Integrin might be involved in some way in the machinery of phagocytosis or in the uptake processes of the microspheres in response to UDP, but the precise target molecule or molecules that bind or recognize microspheres to be phagocytosed remains to be identified. The microglial phagocytosis seen in the present study is a new type that is promoted by the diffusible
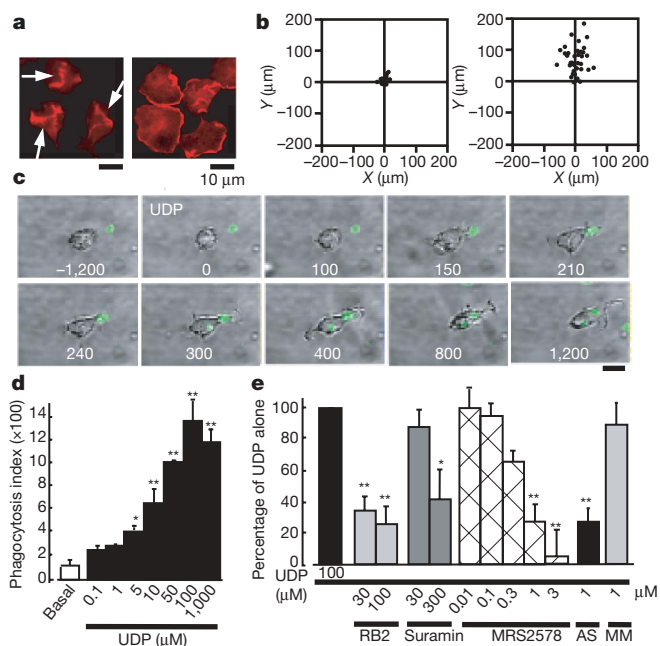

Figure 2 | Changes in cell motilities of microglia. a, UDP- and ATP-evoked membrane ruffles. Cultured microglia were stimulated for 5 min with 100 μM UDP (left) and 10 μM ATP (right), fixed, permeabilized, and then stained with anti-phalloidin. Scale bar, 10 μm. **b,** Typical chemotactic responses of microglia towards 100 μM UDP (left) and 100 μM ATP (right) assessed by the Dunn chemotaxis chamber (see Methods). **c,** Time-lapse images showing the effect of UDP on the microglial morphogenic changes and the uptake of fluorescent zymosan particles (green). The time after addition of UDP is shown in seconds in each picture. **d,** The UDP-evoked uptake of microspheres was assessed quantitatively as a phagocytosis index by using FACS. Data are mean and s.e.m. for three experiments (asterisk, $P < 0.05$; two asterisks, $P < 0.01$ compared with basal). **e,** Effects of the P2 receptor antagonists reactive blue 2 and suramin, the P2Y$_6$ receptor antagonist MRS2578, and P2Y$_6$ AS or MM on the UDP-evoked phagocytosis. Data are means and s.e.m. for three or four experiments (asterisk, $P < 0.05$; two asterisks, $P < 0.01$ compared with UDP alone).

extracellular molecule UDP. However, we cannot deny the possibility that the UDP may simply facilitate the machinery of phagocytosis and that UDP-evoked phagocytosis observed in this study may even include macropinocytosis.

To determine the expression and function of microglial P2Y$_6$ receptors *in vivo*, the excitotoxicity of brain injury was induced by kainic acid (KA) (Fig. 3). KA is an excitatory amino acid that is often used to cause limbic motor epilepsy or excitatory neuronal cell death *in vivo* and *in vitro*. KA acts on non-NMDA glutamate receptors to facilitate excess excitability, thereby leading to necrosis and even apoptosis of neurons. The hippocampal CA1 and CA3 regions are susceptible to neuronal death in response to KA[22]. When KA was injected intraperitoneally into rats (10 mg kg^{-1}), it produced typical limbic seizure within 60 min. At 72 h after the administration of KA, the brains were removed and were used for western blotting, immunohistochemical assays and *in situ* hybridization (ISH). Western blotting analysis showed that KA increased the expression of P2Y$_6$ receptors in comparison with the saline-injected control groups (Fig. 3A, B). Double staining of microglia and neurons by anti-Iba1 (green) and anti-neuronal nuclei (NeuN, red) antibodies, respectively, showed that KA induced severe neuronal loss in the hippocampal CA1 and CA3 regions, where intense Iba1-positive signals—indicative of microglia—were observed. KA increased the number of microglia appearing in the activated form with poorly ramified, short and thick processes (Fig. 3C, f–h). Small NeuN signals seemed to be incorporated in some microglia (see g and h in Fig. 3C), suggesting that microglia phagocytose damaged or dead neurons.

These findings suggest that microglia might migrate or proliferate, probably as a result of KA-induced neuronal damage.

We further examined the cell types that produced increases in P2Y$_6$ receptor protein in response to the administration of KA, and found that P2Y$_6$ immunoreactivities (green in Fig. 3D) were associated with the microglia (OX-42, red in Fig. 3D, c) but not with astrocytes (glial fibrillary acidic protein (GFAP), red in Fig. 3D, d) or neurons (NeuN, red in Fig. 3D, e). Furthermore, we performed ISH to characterize the expression of mRNA coding for P2Y$_6$ receptors with the use of digoxigenin-labelled antisense RNA probe. Signals for P2Y$_6$ receptor mRNA were very low in the naive animals but were upregulated three days after treatment with KA (Fig. 3E, b; blue dots indicated by arrowheads). At this time, the number of microglia increased markedly, especially at the hippocampal CA3 and CA1 regions (Fig. 3C). After ISH, the sections were stained with anti-Iba1 antibody to characterize P2Y$_6$ receptor mRNA signals. In the hippocampal CA3 regions of naive rats, there were very few anti-Iba1-positive microglia that did not show P2Y$_6$ receptor mRNA. In contrast, in the hippocampal CA3 of KA-injected rats, there was an increased number of anti-Iba1-positive microglia, in which P2Y$_6$ receptor signals were colocalized with microglia (Fig. 3E, c; KA, black arrows, see also inset at higher magnification).

There is a growing literature about 'eat-me' signals that are expressed on the cell surface of apoptotic or dying cells. However, diffusible signals that trigger phagocytosis have received only limited attention. When neurons or cells are exposed to traumatic injury such as ischaemia, they swell and subsequently shrink as a result of
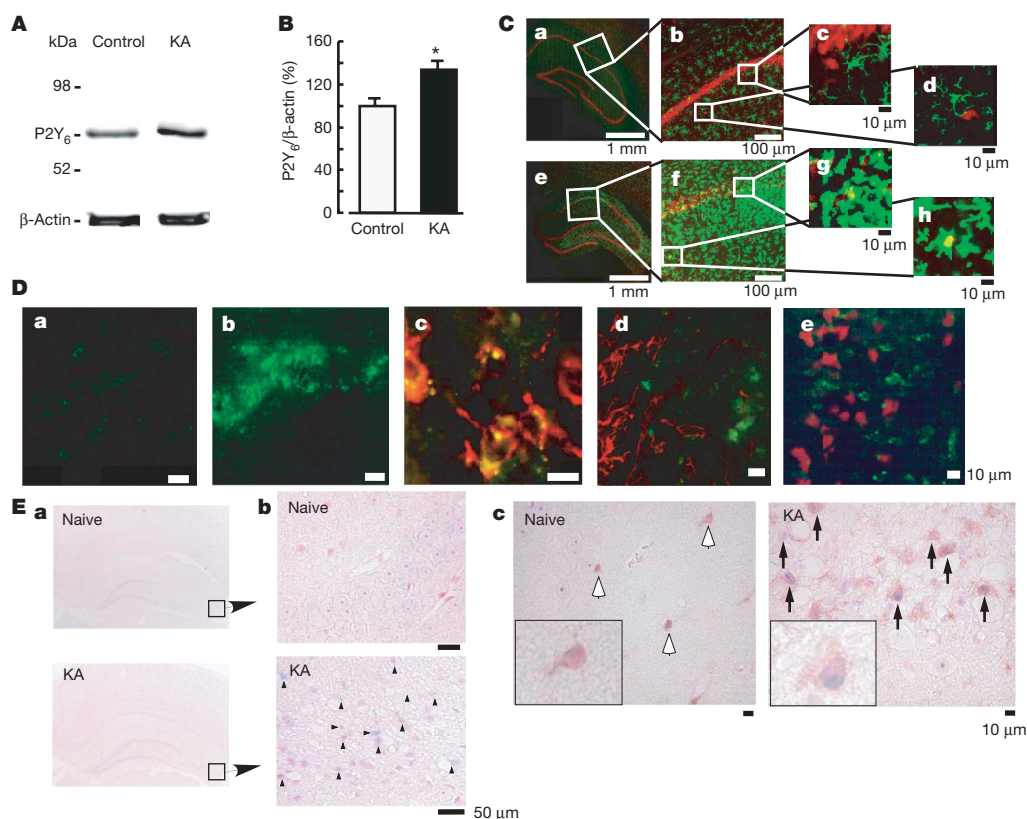

Figure 3 | Increase in P2Y$_6$ receptors in the hippocampus after kainic acid (KA)-treatment. **A,** Western blot analysis, showing increase in P2Y$_6$ receptor protein in rats treated intraperitoneally with 10 mg kg^{-1} KA, 72 h after treatment. **B,** Summary of quantitative data; KA was applied at 10 mg kg^{-1}. Results are means and s.e.m. for 8 (control) and 7 (KA) experiments (asterisk, $P < 0.05$ compared with control). **C,** Immunohisto-chemical analysis in naive control (**a–d**) and KA-treated (**e–h**) rats; red, anti-NeuN antibody; green, anti-Iba1 antibody. Rectangles in **a** and **e** are expanded in **b** and **f**, respectively. Rectangles in **b** and **f** also correspond to **c, d** and **g, h**, respectively. **D,** Anti-P2Y$_6$ antibody signals (green) were increased by KA (**a**, control; **b**, KA), which was colocalized with microglia (red in **c**, anti-OX42) but not with astrocytes (red in **d**, anti-GFAP) or neurons (red in **e**, anti-NeuN). **E,** ISH analysis. **a, b,** mRNA coding for P2Y$_6$ receptor in naive rats was very low but was increased at the hippocampal CA3 region by KA (3 days later) (blue dots and arrowheads, KA). **c,** Double staining with P2Y$_6$ antisense RNA probe (blue dots) and anti-Iba-1 antibody (brown signals, white (naive) or black (KA) arrows). In KA-treated rats there was an increased number of microglia, which was associated with P2Y$_6$ receptor mRNA (blue signals, inset at higher magnification in KA).

increased permeability. This is followed by leakage of cytoplasmic molecules, leading to necrotic cell death. Thus, cytoplasmic nucleotides could be diffusible messengers that signal the crisis state to adjacent cells including microglia. In fact, the diffusible messenger ATP promotes microglial chemotaxis and/or migration[3–6]. Diffusible molecules might be insufficiently precise to cause phagocytes to recognize and eat cells. However, released or leaked nucleotides are immediately degraded by the extracellular nucleotide-degrading enzymes. In this respect, UDP might be a localized and transient marker of traumatized or necrotic cells.

Cell injury results in a leakage of ATP that affects the motility of adjacent cells, including microglia[3,4]. In addition, cells release or leak uridine nucleotides[23] and nucleotide sugars[24] in response to various stimuli or ischaemic injury[25]. We therefore next investigated whether KA increases the release of extracellular UDP from neurons to induce microglial phagocytosis. Cultured hippocampal neurons were stimulated with and without $100\,\mu M$ KA for 1 h; the supernatant was then collected for nucleotide assay by high-performance liquid chromatography (HPLC) or for phagocytosis assay by FACS (Fig. 4). Because released or leaked UTP is rapidly degraded into UDP, UMP and uridine by ARL67156-sensitive ectonucleotidases, we monitored the amount of UTP rather than UDP, and collected the supernatant and the microdialysates in the presence of $20\,\mu M$ ARL67156 throughout experiments. There was a close relationship between the HPLC peak corresponding to UTP and the concentration of the standard UTP ($R^2 = 0.9947$). The amount of UTP in the KA-treated supernatant was significantly larger than that in the KA-untreated control supernatant (Fig. 4b; control, $2.3 \pm 1.1\,\mu M$; KA treated, $10.5 \pm 3.9\,\mu M$, $P < 0.05$). We also tested whether the KA-treated supernatant obtained from cultured hippocampal neurons facilitated microglial phagocytosis. Hippocampal neurons were treated with and without $100\,\mu M$ KA for 1 h; each supernatant was collected and added to microglia; this was followed by a phagocytosis assay. As shown in Fig. 4c, when microglia were incubated with the KA-treated supernatant for 20 min, there was a significant increase in phagocytosis, which was blocked by the P2Y6 receptor antagonist MRS2578 ($1\,\mu M$). KA alone did not stimulate phagocytosis.

Finally, we tested whether KA induces the release of UDP and P2Y6-receptor-mediated phagocytosis in vivo. An increase in extracellular UTP concentration ($[UTP]_o$) was observed soon after injection of KA (from 1 to 4 h after injection), which reached 2–3-fold higher than the KA-untreated control (data not shown). At 1 day after KA injection, $[UTP]_o$ was about 1.5–2.0-fold higher than the KA-untreated control (Fig. 4e). Then, at day 3, $[UTP]_o$ reached almost 10-fold higher levels (9.4 ± 1.2-fold; Fig. 4e and inset), which decreased slightly at day 5. A higher (5–10-fold) $[UTP]_o$ was observed 2–3 days after the injection of KA, which lasted at least another couple of days. It should be noted that loss of neurons (removal of neurons) also became obvious 2–3 days after the KA injection. We further injected fluorescent microspheres into the hippocampal CA3 regions of KA-treated rats, and then counted the numbers of the microspheres phagocytosed or attached by microglia. The P2Y6 receptor antagonist MRS2578 was injected into the hippocampal CA3 region, and P2Y6 AS or MM (mismatch oligonucleotide) was injected into the third ventricle. The number of microspheres taken or attached by microglia was markedly increased by KA treatment, which is significantly inhibited by MRS2578 or P2Y6 AS but not by MM (Fig. 4g; see also Supplementary Fig. 6). These findings all suggest that UDP/ P2Y6-receptor-mediated signals are important for microglial phagocytosis even in vivo.

A recent review described that dying cells use both 'find-me' and 'eat-me' signals for phagocyte attraction and recognition[26]. Nucleotides could be both 'find-me' and 'eat-me' signals. The intracellular ATP concentration is estimated to be high (more than 5 mM) and the UTP concentration is reported to be one-third that of ATP[23]. Cells release ATP, and here we showed that KA caused an increase in extracellular UTP or UDP. Microglia might therefore be attracted by

ATP or ADP[5,6] and subsequently recognize UDP, leading to the removal of the dying cells or their debris. It is interesting that ATP/ ADP is not able to efficiently activate P2Y6 receptors; neither can UDP act on P2Y12/13 receptors. Thus, even if these nucleotides were leaked or released simultaneously, adenine and uridine nucleotides would regulate microglial motilities, namely chemotaxis and phagocytosis, in a mutually exclusive but coordinated fashion.

So far we have not shown quantitative data indicating that individual microglia upregulate the expression of P2Y6 receptors. A significant, but not drastic, increase in P2Y6 receptor protein in the hippocampus was observed after injection of KA (Fig. 3A, B). ISH data show that expression of mRNA coding for P2Y6 receptors in microglia was very low in naive animals but became obvious in an increased number of microglia after KA injection (Fig. 3E), suggesting that the increase in P2Y6 receptor protein is not due simply to an increased number of microglia but is upregulated in individual
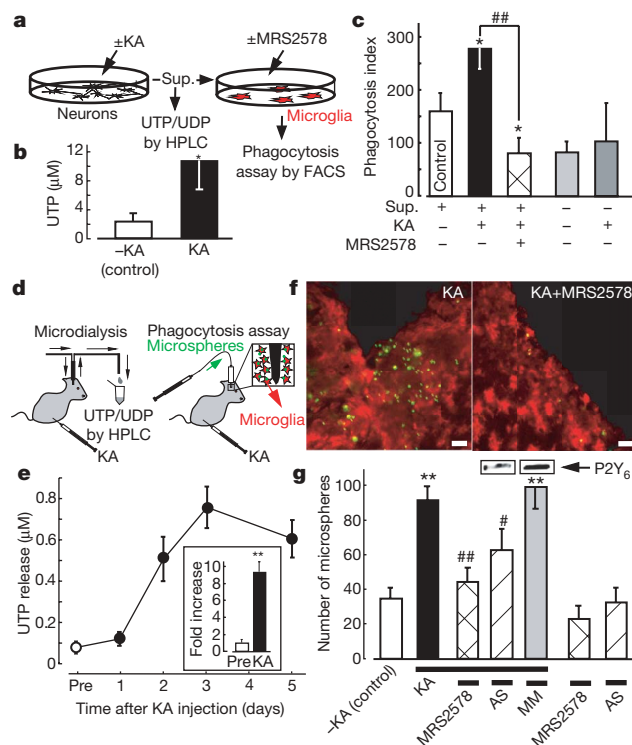

Figure 4 | KA-evoked increases in extracellular uridine nucleotides and P2Y6-receptor-mediated microglial phagocytosis in vitro and in vivo. a, Schematic diagram of the experiments in vitro. Sup., supernatant. b, Summary of the UTP concentration in the KA-treated and control supernatants. Data show means and s.e.m. for at least five independent experiments (asterisk, $P < 0.05$ compared with control). c, Effects of the KA-treated and control supernatant on microglial uptake of fluorescent latex beads. Data show means and s.e.m. for at least four independent experiments (asterisk, $P < 0.05$ compared with control; hash sign, $P < 0.05$ compared with KA-treated supernatant). d, Schematic diagram of the experiments in vivo. KA was applied intraperitoneally at $10\,mg\,kg^{-1}$. e, Time course of changes in $[UTP]_o$ in baseline dialysates (before treatment with KA (Pre), and 1, 2, 3 and 5 days afterwards). Inset, fold increase at day 3 (compared with before treatment). f, Typical pictures of fluorescent microspheres (green) attached or taken up by microglia (red, anti-Iba1) in the KA-treated (left) and KA + MRS2578-treated (right) hippocampal CA3 regions. Scale bar, $20\,\mu m$. g, Quantitative analysis of phagocytosis in vivo (details are provided in Supplementary Methods). Changes in P2Y6 receptor protein by P2Y6 AS or MM are shown at the top of corresponding columns. Values are means and s.e.m. (asterisk, $P < 0.01$ compared with control (−KA); hash sign, $P < 0.05$; two hash signs, $P < 0.01$ compared with KA-treated group). Statistical analyses were performed by ANOVA with Scheffe's multiple comparison. At least three sections containing the injection sites were analysed per animal, and at least three animals were used in each group for analysis.

microglia. We also emphasize that even if the extent of $P2Y_6$ receptor upregulation is not drastic, an increase in extracellular UDP, a ligand for $P2Y_6$ receptor, is markedly increased after treatment with KA (detected as UTP, almost 10-fold; Fig. 4e) and therefore that the UDP/$P2Y_6$ receptor system would be sufficiently activated to cause microglial phagocytosis after treatment with KA. In comparison with the extensive knowledge of the molecular events involved in the regulation of apoptosis or necrosis, relatively little is known about the processes responsible for the clearance of dead cells and the degradation of waste materials. Considering the present findings that injured neurons leak diffusible UTP/UDP and cause the upregulation of $P2Y_6$ receptors in microglia, the UDP/$P2Y_6$ receptor system might function as a critical device covering the phagocytosis of both apoptotic and necrotic cells if they release or leak UDP by sensing diffusible UDP signals.

Thus we have shown that microglia express $P2Y_6$ receptors that function as a sensor of phagocytosis. The $P2Y_6$ receptor agonist UDP is released (as UTP) when neurons are damaged by KA. Thus, the activation of $P2Y_6$ receptors by UDP would be a key event in initiating the clearance of dying cells or debris in the central nervous system.

METHODS

Detailed methods are provided in Supplementary Information.

Microglia culture. The protocol was reviewed and approved by the Committee for Institutional Laboratory Animal Care of the National Institute of Health Sciences. Rat primary cultured microglia were prepared in accordance with the method described previously[27].

Phagocytosis assay *in vitro* and *in vivo*. *In vitro* microglial phagocytosis was assessed by either FACScan analysis or imaging analysis with fluorescently labelled microspheres. For the *in vivo* phagocytosis assay, fluorescently labelled microspheres were injected into the hippocampal CA3 region after injection of KA, and then the number of fluorescent microspheres associated with microglia was analysed by confocal microscopy (LSM 5 Pascal; Carl Zeiss).

Microdialysis. A microdialysis probe (A-I type probe; Eicom) was inserted into the hippocampal CA3 region by means of a guide cannula, and was perfused continuously at a flow rate of $3.0\ \mu l\ min^{-1}$ (collected for 60 min) supplemented with the ectonuclease inhibitor ARL67156 ($20\ \mu M$) (Sigma).

Quantification of UTP by HPLC. The concentration of nucleotides in the supernatant of the hippocampal cultures was analysed with an HPLC system (Jasco) combined with a C_{18} column (4.6×250 mm, Shodex) as described[28], with minor modifications.

Data analysis and statistics. All results are expressed as means ± s.e.m. A statistical analysis was performed with Student's *t*-test or analysis of variance, followed by Scheffe's multiple comparison test. Differences were considered to be significant at $P < 0.05$.

1. Guthrie, P. B. *et al.* ATP released from astrocytes mediates glial calcium waves. *J. Neurosci.* **19**, 520–528 (1999).
2. Koizumi, S., Fujishita, K., Tsuda, M., Shigemoto-Mogami, Y. & Inoue, K. Dynamic inhibition of excitatory synaptic transmission by astrocyte-derived ATP in hippocampal cultures. *Proc. Natl Acad. Sci. USA* **100**, 11023–11028 (2003).
3. Nimmerjahn, A., Kirchhoff, F. & Helmchen, F. Resting microglial cells are highly dynamic surveillants of brain parenchyma *in vivo. Science* **308**, 1314–1318 (2005).
4. Davalos, D. *et al.* ATP mediates rapid microglial response to local brain injury *in vivo. Nature Neurosci.* **8**, 752–758 (2005).
5. Honda, S. *et al.* Extracellular ATP or ADP induce chemotaxis of cultured microglia through $G_{i/o}$-coupled P2Y receptors. *J. Neurosci.* **21**, 1975–1982 (2001).
6. Nasu-Tada, K., Koizumi, S. & Inoue, K. Involvement of β1 integrin in microglial chemotaxis and proliferation on fibronectin: different regulations by ADP through PKA. *Glia* **52**, 98–107 (2005).
7. Ferrari, D. *et al.* P2Z purinoreceptor ligation induces activation of caspases with distinct roles in apoptotic and necrotic alterations of cell death. *FEBS Lett.* **447**, 71–75 (1999).
8. Suzuki, T. *et al.* Production and release of neuroprotective tumor necrosis factor by P2X7 receptor-activated microglia. *J. Neurosci.* **24**, 1–7 (2004).
9. Tsuda, M. *et al.* P2X4 receptors induced in spinal microglia gate tactile allodynia after nerve injury. *Nature* **424**, 778–783 (2003).
10. Chang, K., Hanaoka, K., Kumada, M. & Takuwa, Y. Molecular cloning and functional analysis of a novel P2 nucleotide receptor. *J. Biol. Chem.* **270**, 26152–26158 (1995).
11. Communi, D., Parmentier, M. & Boeynaems, J. M. Cloning, functional expression and tissue distribution of the human P2Y6 receptor. *Biochem. Biophys. Res. Commun.* **222**, 303–308 (1996).
12. Mamedova, L. K., Joshi, B. V., Gao, Z. G., von Kugelgen, I. & Jacobson, K. A. Diisothiocyanate derivatives as potent, insurmountable antagonists of P2Y6 nucleotide receptors. *Biochem. Pharmacol.* **67**, 1763–1770 (2004).
13. Greenberg, S. Signal transduction of phagocytosis. *Trends Cell Biol.* **5**, 93–99 (1995).
14. Mitchison, T. J. & Cramer, L. P. Actin-based cell motility and cell locomotion. *Cell* **84**, 371–379 (1996).
15. Lauffenburger, D. A. & Horwitz, A. F. Cell migration: a physically integrated molecular process. *Cell* **84**, 359–369 (1996).
16. Ohsawa, K., Imai, Y., Kanazawa, H., Sasaki, Y. & Kohsaka, S. Involvement of Iba1 in membrane ruffling and phagocytosis of macrophages/microglia. *J. Cell Sci.* **113**, 3073–3084 (2000).
17. Kreutzberg, G. W. Microglia: a sensor for pathological events in the CNS. *Trends Neurosci.* **19**, 312–318 (1996).
18. Nicholas, R. A. *et al.* Pharmacological and second messenger signalling selectivities of cloned P2Y receptors. *J. Auton. Pharmacol.* **16**, 319–323 (1996).
19. Mellor, E. A., Maekawa, A., Austen, K. F. & Boyce, J. A. Cysteinyl leukotriene receptor 1 is also a pyrimidinergic receptor and is expressed by human mast cells. *Proc. Natl Acad. Sci. USA* **98**, 7964–7969 (2001).
20. Lauber, K. *et al.* Apoptotic cells induce migration of phagocytes via caspase-3-mediated release of a lipid attraction signal. *Cell* **113**, 717–730 (2003).
21. Bard, F. *et al.* Peripherally administered antibodies against amyloid β-peptide enter the central nervous system and reduce pathology in a mouse model of Alzheimer disease. *Nature Med.* **6**, 916–919 (2000).
22. Sperk, G. *et al.* Kainic acid induced seizures: neurochemical and histopathological changes. *Neuroscience* **10**, 1301–1315 (1983).
23. Lazarowski, E. R., Homolya, L., Boucher, R. C. & Harden, T. K. Direct demonstration of mechanically induced release of cellular UTP and its implication for uridine nucleotide receptor activation. *J. Biol. Chem.* **272**, 24348–24354 (1997).
24. Lazarowski, E. R., Shea, D. A., Boucher, R. C. & Harden, T. K. Release of cellular UDP-glucose as a potential extracellular signaling molecule. *Mol. Pharmacol.* **63**, 1190–1197 (2003).
25. Erlinge, D. *et al.* Uridine triphosphate (UTP) is released during cardiac ischemia. *Int. J. Cardiol.* **100**, 427–433 (2005).
26. Ravichandran, K. S. 'Recruitment signals' from apoptotic cells: invitation to a quiet meal. *Cell* **113**, 817–820 (2003).
27. Nakajima, K. *et al.* Identification of elastase as a secretory protease from cultured rat microglia. *J. Neurochem.* **58**, 1401–1408 (1992).
28. Lazarowski, E. R., Boucher, R. C. & Harden, T. K. Constitutive release of ATP and evidence for major contribution of ecto-nucleotide pyrophosphatase and nucleoside diphosphokinase to extracellular nucleotide concentrations. *J. Biol. Chem.* **275**, 31061–31068 (2000).

Author Contributions S.K. designed most experiments, performed Ca^{2+} imaging and *in vivo* experiments and wrote the paper. Y.S.M. conducted major parts of the experiments. K.N.T. and Y.S. carried out the FACS assay and the HPLC analysis, respectively. K.O. and S.K. performed the chemotaxis analysis. B.V.J. and K.A.J. made the $P2Y_6$ receptor antagonist MRS2578. M.T. analysed the data. K.I. analysed the data and coordinated the project. K.I. also designed the project. 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. Correspondence and requests for materials should be addressed to K.I. (inoue@phar.kyushu-u.ac.jp).

LETTERS

Phosphorylation of Erp1 by p90rsk is required for cytostatic factor arrest in *Xenopus laevis* eggs

Tomoko Nishiyama[1], Keita Ohsumi[1] & Takeo Kishimoto[1]

Until fertilization, the meiotic cell cycle of vertebrate eggs is arrested at metaphase of meiosis II by a cytoplasmic activity termed cytostatic factor (CSF)[1], which causes inhibition of the anaphase-promoting complex/cyclosome (APC/C), a ubiquitin ligase that targets mitotic cyclins—regulatory proteins of meiosis and mitosis—for degradation[2,3]. Recent studies indicate that Erp1/Emi2, an inhibitor protein for the APC/C, has an essential role in establishing and maintaining CSF arrest[4–6], but its relationship to Mos, a mitogen-activated protein kinase (MAPK) kinase kinase that also has an essential role in establishing CSF arrest[7] through activation of p90 ribosomal S6 kinase (p90rsk)[8,9], is unclear. Here we report that in *Xenopus* eggs Erp1 is a substrate of p90rsk, and that Mos-dependent phosphorylation of Erp1 by p90rsk at Thr 336, Ser 342 and Ser 344 is crucial for both stabilizing Erp1 and establishing CSF arrest in meiosis II oocytes. Semi-quantitative analysis with CSF-arrested egg extracts reveals that the Mos-dependent phosphorylation of Erp1 enhances, but does not generate, the activity of Erp1 that maintains metaphase arrest. Our results also suggest that Erp1 inhibits cyclin B degradation by binding the APC/C at its carboxy-terminal destruction box[10], and this binding is also enhanced by the Mos-dependent phosphorylation. Thus, Mos and Erp1 collaboratively establish and maintain metaphase II arrest in *Xenopus* eggs. The link between Mos and Erp1 provides a molecular explanation for the integral mechanism of CSF arrest in unfertilized vertebrate eggs.

Since the discovery of CSF (the transferable cytoplasmic activity in amphibian oocytes that is responsible for metaphase arrest in unfertilized eggs[1]), the mechanism for CSF arrest in vertebrate eggs has been extensively investigated[11–13]. Recent studies[4,14–16] have shown that release from CSF arrest occurs when the calcium ion flux at fertilization activates calmodulin-dependent protein kinase II (CaMKII)[17], which phosphorylates Erp1/Emi2, thereby acting as a priming kinase for the recruitment of polo-like kinase (Plx1). Plx1, in turn, triggers the degradation of Erp1 by phosphorylation of a DSGX$_3$S degron, which results in cyclin-dependent kinase 1 (Cdk1) inactivation through APC/C-mediated degradation of B-type cyclin. In contrast, the mechanism for the establishment and maintenance of CSF arrest remains elusive. Although it has recently been demonstrated that Erp1/Emi2 is an essential component of CSF[4,6], its relationship to regulatory kinases required for CSF arrest, especially Mos, is unclear[18]. The Mos–MAPK–p90rsk pathway has been shown to have an essential role in establishing CSF arrest[7–9,19,20]; however, its downstream target(s) involved in the regulation of cyclin B degradation remain(s) undetermined.

We first examined the changes of Erp1 levels in *Xenopus* oocytes and embryos during meiotic maturation and cleavage periods. As shown in Fig. 1a, Erp1 was undetectable in oocytes until meiotic interphase, and rapidly accumulated in meiosis II (Fig. 1a), consistent with the

occurrence of metaphase arrest in meiosis II, but not meiosis I (see Supplementary Fig. 1). In fertilized eggs, Erp1 disappeared following fertilization, and accumulated during the subsequent interphase to a
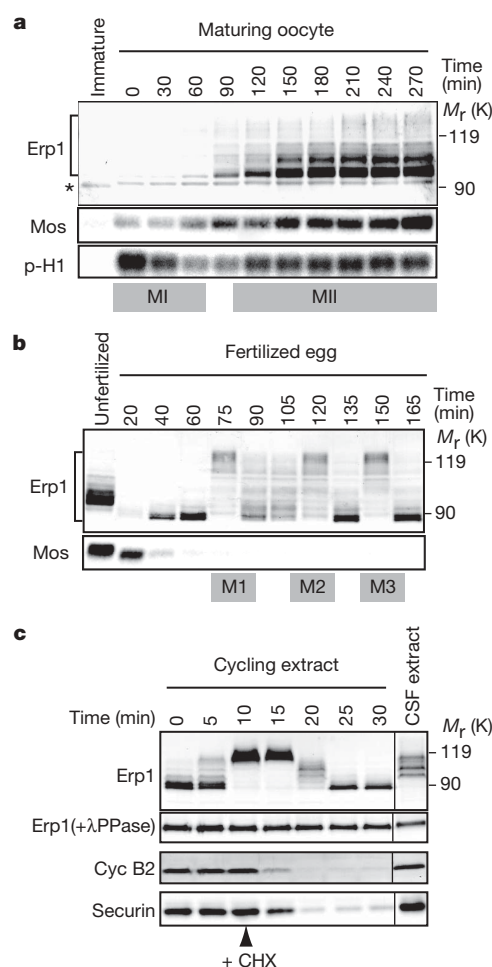

Figure 1 | Erp1 is present throughout the early embryonic cell cycle.
a, **b**, Erp1 and Mos levels in oocytes (**a**) and fertilized eggs (**b**) were examined by immunoblotting. Cdk1 activity in maturing oocytes was also examined (p-H1). The timing of meioses I (MI) and II (MII) and early embryonic mitoses (M1, 2 and 3) are indicated. The asterisk indicates non-specific signals. **c**, Erp1, cyclin B2 (cyc B2) and securin protein levels in mitotic cycling extracts, with protein synthesis inhibited at the entry to M-phase, were examined by immunoblotting. To precisely compare differences in the amount of Erp1 in various cell cycle phases, extracts were treated with λ-protein phosphatase (λPPase) before immunoblotting. The timing of the cycloheximide (CHX) addition is indicated.

[1]Laboratory of Cell and Developmental Biology, Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8501, Japan.

level comparable to that in unfertilized eggs, which was estimated at 13 nM (Fig. 1b; and Supplementary Fig. 2). Erp1 persisted in embryos until the late-blastula stage and disappeared by the late-gastrula stage (data not shown).

We then monitored the Erp1 level in egg extracts that can undergo embryonic cell cycles[21], and found that it did not change through mitotic phase (M-phase), whereas the cyclin B and securin levels dropped at M-phase exit (Fig. 1c). Thus, Erp1 survives through M-phases in early embryonic cells, indicating that Erp1 expression is not sufficient for inducing metaphase arrest. Because Mos–MAPK–p90rsk activity is also essential for establishing metaphase II arrest[7–9,19,20], we predicted that Erp1 is modified by the Mos–MAPK–p90rsk pathway, and that the modification is required for Erp1 to function as a metaphase-arrest inducer.

To test this hypothesis, we investigated whether Erp1 is post-translationally modified, depending on Mos–MAPK activity. Because the electrophoretic mobility of Erp1 changed during the cell cycle (Fig. 1b), we compared Erp1 in CSF-arrested extracts[22,23] (CSF extracts), M-phase extracts (M extracts) and interphase extracts (I extracts). As shown in Fig. 2a, Erp1 in I extracts migrated faster than Erp1 in either CSF or M extracts (Erp1-CSF or Erp1-M, respectively), and Erp1-CSF migrated faster than Erp1-M. The difference in the mobility between Erp1-CSF and Erp1-M was attributed to the presence or absence of Mos–MAPK activity, respectively, because after addition of glutathione S-transferase (GST)–Mos to M extracts to activate endogenous MAPK[24], Erp1 exhibited the same mobility as seen for Erp1-CSF (Fig. 2a). Erp1 shifted upward to an even slower migrating form, 3 min after Ca^{2+} addition to CSF extracts, as a result of phosphorylation by CaMKII[15,16].

To identify the Erp1 site(s) that is modified by Mos–MAPK activity, we prepared in vitro translated, truncated versions of haemagglutinin (HA)-tagged Erp1. After incubation in either CSF or M-phase-like (Mit) extracts, which have Cdk1 activity but not Mos–MAPK activity, the electrophoretic mobility of each truncated HA–Erp1 was examined; in this experiment, Erp1 mutated at the amino-terminal DSGX$_3$S degron (N-Erp1)[4,5] was used to increase its stability in extracts (Supplementary Figs. 3 and 4). The results revealed that truncated HA–Erp1, including the amino acids 329–356, exhibited the differential mobility shifts indicative of phosphorylation (Supplementary Fig. 3); the domain contained four serine and threonine residues, Ser 335, Thr 336, Ser 342 and Ser 344, the latter three of which are shared by other mammalian Erp1 (Fig. 2b).

We next examined the mobility of in vitro translated full-length HA–N-Erp1 mutated at one of the four residues. The results showed that after incubation in CSF extracts, HA–N-Erp1(S335A) migrated to the Erp1-CSF position as did the wild-type, whereas HA–N-Erp1(T336A), HA–N-Erp1(S342A) or HA–N-Erp1(S344A) migrated close to the Erp1-M position (Fig. 2b); these three residues are referred to as TSS hereafter. Consistently, HA–N-Erp1 mutated at all residues of TSS (3A) exhibited a mobility shift to the Erp1-M position after treatment with CSF extracts (Fig. 2b). Thus, TSS phosphorylation is crucial for the CSF-specific mobility shift of Erp1.

To examine whether Erp1 phosphorylation at TSS depends on Mos–MAPK activity, we incubated a GST fusion of the Erp1(325–354) peptide (GST-325–354) in either CSF or Mit extracts in the presence of ^{32}P-labelled ATP to measure its label incorporation. The ^{32}P incorporation into the peptide was five times higher in CSF extracts than in Mit extracts, whereas histone H1 phosphorylation was higher in Mit extracts (Fig. 2c). Furthermore, when I extracts were supplemented with ΔN-Ste11—a constitutively active form of a MAPK kinase kinase[25]—to activate endogenous MAPK, the phosphorylation of GST-325–354 was increased by more than fourfold. Thus, phosphorylation of Erp1(325–354) is dependent on Mos–MAPK activity.

We also examined the extent of GST–325-354 phosphorylation in CSF extracts deprived of p90rsk2, the major p90rsk in Xenopus eggs[8]. These results showed that in p90rsk2-deprived CSF extracts,

GST–325-354 phosphorylation was reduced to a low level, even though MAPK remained activated (Fig. 2c). The phosphorylation of GST–325-354 in the p90rsk2-deprived extracts was restored by the addition of wild-type p90rsk1 or constitutively active p90rsk2, but not by the addition of kinase-negative p90rsk2. In addition, full-length, wild-type maltose binding protein (MBP)–Erp1, but not its 3A mutant, was phosphorylated in egg extracts, depending on MAPK activity, and also was phosphorylated with p90rsk2 immunoprecipitated from CSF extracts (Fig. 2d). Thus, p90rsk2 is a kinase responsible for Erp1 phosphorylation at TSS in Xenopus eggs.

We next examined whether TSS phosphorylation is involved in the regulation of CSF activity. We first confirmed the requirement for Erp1 in establishing CSF arrest in meiosis II oocytes[6]. When immature oocytes were injected with antisense oligonucleotides for Erp1 and subsequently treated with progesterone, they resumed meiosis—as indicated by breakdown of the germinal vesicle (GVBD) and an
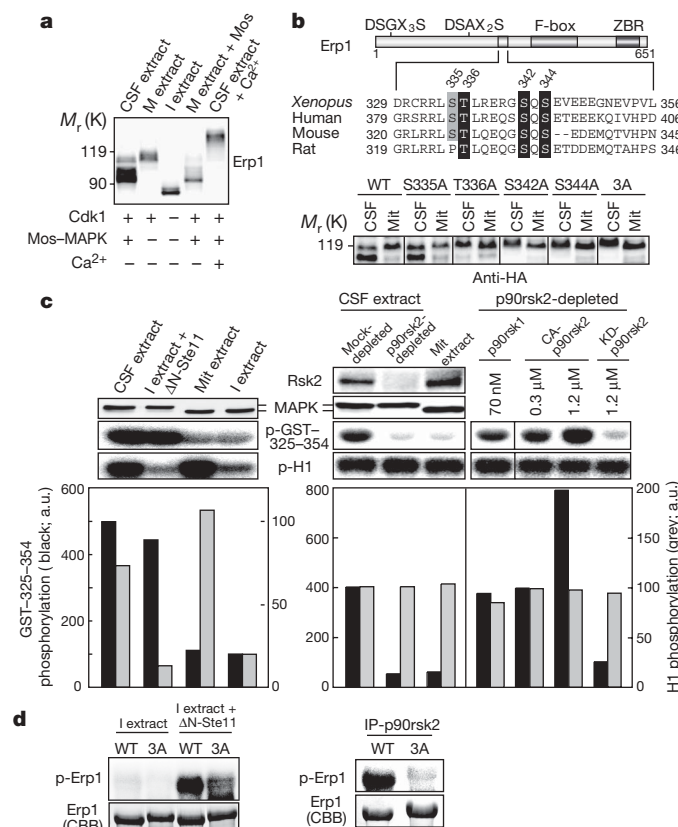

Figure 2 | Erp1 is phosphorylated by p90rsk, depending on Mos–MAPK activity. a, CSF extract, cycling extracts in M-phase and interphase, M extract to which Mos was added, and CSF extract at 3 min after Ca^{2+}-addition were immunoblotted with Erp1 antibodies. **b**, Sequence comparison of the 329–356 domain of Xenopus Erp1 with mammalian Erp1 is shown (top). Conserved Thr and Ser residues are shaded in black. The HA-tagged, full-length Erp1 proteins with the indicated mutations were immunoblotted, after incubation in either CSF or Mit extract. **c**, GST–Erp1(325–354) was incubated in the indicated extracts containing [γ-^{32}P]ATP, and the label incorporation was examined by autoradiography. The extracts were simultaneously examined for Cdk1 activity using histone H1 as substrate (p-H1), and for MAPK and p90rsk2 by immunoblotting; upper and lower bands of MAPK indicate the active and inactive forms, respectively. p90rsk2-deprived extracts were supplemented with either recombinant, full-length p90rsk1, constitutively active (CA) or kinase-negative (KD) p90rsk2. The extent of GST–Erp1(325–354) and histone H1 phosphorylation was quantified (bottom). a. u., arbitrary units. **d**, Phosphorylation of the full-length, MBP-tagged N-Erp1(WT) or N-Erp1(3A) in the indicated extracts (left) or with p90rsk2 immunoprecipitated from CSF extract (right) was examined by autoradiography. The gels were also stained with Coomassie brilliant blue (CBB).

increase in Cdk1 activity—but did not accumulate Erp1 (Fig. 3a). In Erp1-deprived oocytes, cyclin B levels decreased and Cdk1 was not fully activated after exit from meiosis I, and oocytes degenerated within 4 h after GVBD, indicating a failure to establish metaphase II arrest.

When Erp1-deprived oocytes were injected with 30 nM MBP–N-Erp1(WT) 60 min after GVBD, cyclin B levels increased, Cdk1 activity was elevated and oocytes maintained a healthy appearance for at least 4 h (Fig. 3b). In contrast, those injected with the same amount of MBP–N-Erp1(3A) did not restore Cdk1 activation and degenerated after meiosis I exit. Immunoblot analysis revealed, however, that MBP–N-Erp1(3A) was unstable in maturing oocytes, whereas the wild-type construct was not (Fig. 3b), indicating that the failure of MBP–N-Erp1(3A) to induce metaphase II arrest could be due to the instability. To examine this possibility, we prepared MBP–N-Erp1(3A), which was also mutated at the phospho degron at the central domain (NA-Erp1(3A); Supplementary Fig. 4). Even though NA-Erp1(3A) was found to be as stable as the wild-type construct in Erp1-deprived oocytes, it failed to restore meiosis II arrest, as did N-Erp1(3A) (Fig. 3c). These results confirm that Erp1 phosphorylation at TSS is required for establishing CSF arrest and, furthermore, indicate that TSS phosphorylation is also important to protect Erp1 against degradation that is mediated by the central degron, which is much slower than the N-terminal degron-mediated degradation.

Interestingly, in post-meiosis I oocytes, Mos was unstable in the absence of Erp1 phosphorylated at TSS (Fig. 3c). Similarly, Erp1 was less stable in maturing oocytes that had been deprived of Mos[26],

probably because of the lack of TSS phosphorylation, which was dependent on Mos (Supplementary Figs. 4 and 5). Thus, Mos and Erp1 stabilize each other in meiosis II oocytes.

To evaluate the requirement of Erp1 phosphorylation at TSS for CSF activity, we performed quantitative analysis of the M-phase-maintaining activity of MBP–N-Erp1, using CSF extracts. As shown in a previous study[4], immunodepletion of Erp1 from CSF extracts abolished their potential for metaphase arrest, inducing Cdk1 inactivation through cyclin B degradation (Fig. 4a, b). When Erp1-depleted CSF extracts were supplemented with 25 nM MBP–N-Erp1(WT) immediately after Erp1 depletion, M-phase state was maintained, whereas it was not with the same concentration of MBP–N-Erp1 (3A) (Fig. 4b). Importantly, the 3A mutant was preserved as well as
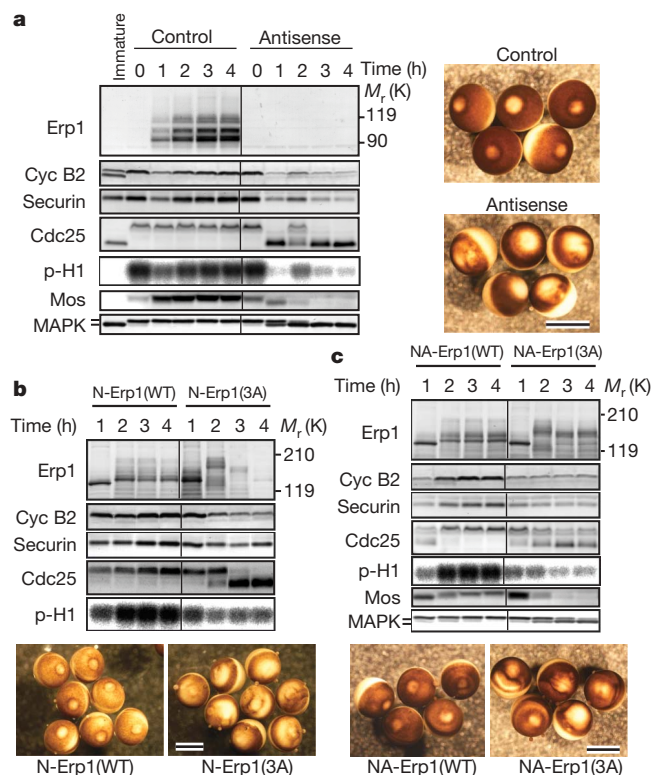

Figure 3 | Phosphorylation of Erp1 at TSS increases its stability and is required for the establishment of metaphase arrest in meiosis II oocytes. a, Immature oocytes that had been injected with antisense oligonucleotides for Erp1 were treated with progesterone and examined for changes of Erp1, cyclin B2, securin, Cdc25, Cdk1 activity (p-H1), Mos, MAPK and their appearance during the post-GVBD period. **b, c**, Some antisense-injected oocytes received an injection of either 30 nM MBP–N-Erp1(WT) or MBP–N-Erp1(3A) (**b**), or 20 nM MBP–NA-Erp1(WT) or MBP–NA-Erp1(3A) (**c**), 60 min after GVBD. The slower and faster migrating forms of Cdc25 reflect high and low Cdk1 activity, respectively. The time of GVBD is taken as 0 h, and the photographs were taken 4 h after GVBD. Scale bars, 1 mm.
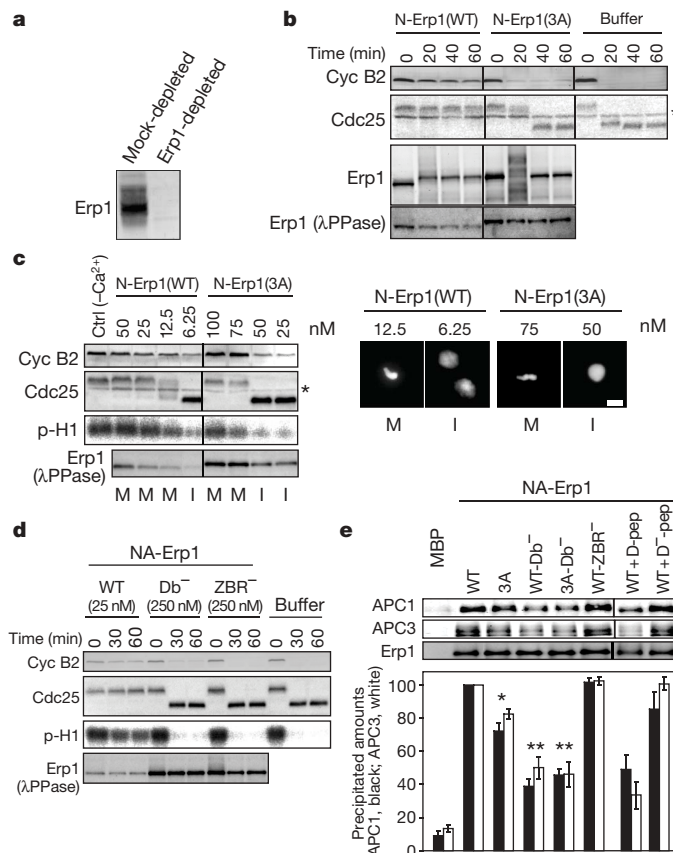
Figure 4 | Phosphorylation of Erp1 at TSS enhances its CSF activity by increasing its binding affinity to the APC/C. a, b, CSF extracts deprived of Erp1 by immunodepletion (**a**) were combined with MBP–N-Erp1(WT) or MBP–N-Erp1(3A) and examined for cyclin B2, Cdc25 and MBP–N-Erp1 (**b**). To compare the amount of Erp1, extracts were treated with λ protein phosphatase (λPPase) before immunoblotting. **c**, CSF extracts were supplemented with various amounts of MBP–N-Erp1(WT) or Erp1(3A), followed by Ca^{2+} addition to degrade endogenous Erp1, and examined 45 min after Ca^{2+} addition for cyclin B2, Cdc25 and Cdk1 activity (p-H1), and the cell cycle phase determined by sperm chromatin morphology. M and I indicate the formation by sperm chromatin of condensed chromosomes and nuclei, respectively. Scale bar, 10 μm. The asterisks in **b** and **c** indicate non-specific signals. ctrl, control. **d**, Erp1-depleted CSF extracts were supplemented with the indicated MBP–NA-Erp1 proteins, and examined for cyclin B2, Cdc25C, Cdk1 activity (p-H1) and MBP–NA-Erp1. **e**, MBP or each of the indicated MBP–NA-Erp1 proteins (200 nM) was incubated in Erp1-depleted CSF extracts for 10 min, and immunoprecipitated to examine co-precipitated APC1 and APC3 by immunoblotting. Some extracts were supplemented with 400 μM D-box peptides (WT + D-pep) or dysfunctional D-box peptides (WT + D⁻-pep), before Erp1(WT) addition. The signal intensity was quantified, then normalized by the NA-Erp1 amount of each precipitate, and expressed as relative values, taking the value of NA-Erp1(WT) as 100. Error bars, s.e.m. ($n = 3–5$; $*P < 0.01$ and $**P < 0.001$).

the wild type during the incubation period, indicating that the failure of MBP–N-Erp1(3A) to maintain CSF arrest was not due to its instability. We note, however, that when MBP–N-Erp1(3A) was added to CSF extracts at higher concentrations ($>75 \, nM$), and followed by addition of Ca^{2+} to induce degradation of endogenous Erp1, M-phase arrest of the extracts was maintained (Fig. 4c). Similar results were obtained with Erp1-depleted CSF extracts (Supplementary Fig. 6). These results strongly suggest that Erp1 phosphorylation at TSS enhances (by approximately fivefold), but does not generate, CSF activity. We also confirmed the requirement of Erp1 phosphorylation at TSS for inducing metaphase arrest in the embryonic cell cycle, using extracts from activated eggs (Supplementary Fig. 7).

A recent study has demonstrated that Emi1 carries a destruction box (D-box) at the C terminus, and inhibits the APC/C by binding[10]. Because the D-box is conserved in Erp1, Erp1 might similarly inhibit the APC/C. To examine this, we tested whether NA-Erp1 mutated at the D-box (R529A, L532A; Db⁻) sustained M-phase arrest in Erp1-depleted CSF extracts. As shown in Fig. 4d, Erp1-Db⁻ did not sustain M-phase arrest in Erp1-depleted extracts, even at 250 nM. CSF activity of NA-Erp1 was also abolished by a mutation at the ZBR (C583A; ZBR⁻), consistent with results that show the APC/C-inhibiting activity of Emi1 is abolished by a similar mutation[10]. To examine whether D-box-dependent CSF activity of Erp1 is regulated by its APC/C binding, we isolated NA-Erp1 proteins that had been added to Erp1-deprived extracts, and quantified the amounts of co-precipitated APC/C subunits, APC1 and APC3. As shown in Fig. 4e, the amounts of the APC/C subunits associated with Erp1-Db⁻ were no more than half of those associated with Erp1(WT). The association of Erp1(WT) with the APC/C was inhibited by peptides carrying a D-box, but was not inhibited by those carrying a mutated D-box (Fig. 4e). Thus, like Emi1, Erp1 stably binds the APC/C using its D-box. Erp1 binding to the APC/C was significantly compromised by the Erp1 mutation at TSS. It is likely, therefore, that Erp1 inhibits APC/C-mediated cyclin B degradation through D-box binding to the APC/C—as does Emi1—and the inhibitory binding is enhanced by Mos-dependent phosphorylation of Erp1 at TSS.

This study demonstrates that phosphorylation of Erp1 by p90rsk at TSS is required for the establishment and maintenance of metaphase II arrest in *Xenopus* eggs (see Supplementary Fig. 1). Thus, in the original cytoplasmic transfer experiment leading to the proposal of CSF[1], Mos was the essential component that was transferred, whereas Erp1 was also the requisite component that resided in the recipient blastomere. Accordingly, CSF activity of unfertilized eggs should be ascribed to a regulatory system in which both Mos and Erp1 are indispensable components, with the stability and activity of each dependent on the other, but not to a system of either Mos or Erp1 alone. In the mouse oocyte, p90rsk is dispensable for establishing CSF arrest[27], though Mos and Erp1 are essential. Although this may indicate that Erp1 is regulated differently in *Xenopus* than in mouse, it is also possible that another kinase(s) substitutes for p90rsk in the mouse oocyte, because there exists a downstream pathway of Mos–MAPK other than the one leading to p90rsk activation[28]. The mechanisms by which Erp1 phosphorylation at TSS enhances its D-box binding to the APC/C, and by which the ZBR inhibits the APC/C remain to be elucidated.

METHODS

Oocytes, eggs and embryos of *Xenopus laevis* were obtained and meiotic resumption of oocytes and activation of eggs were induced as described[29]. Other methods and reagents are described in detail in Supplementary Methods.

1. Masui, Y. & Markert, C. L. Cytoplasmic control of nuclear behavior during meiotic maturation of frog oocytes. *J. Exp. Zool.* **177**, 129–145 (1971).
2. Harper, J. W., Burton, J. L. & Solomon, M. J. The anaphase-promoting complex: it's not just for mitosis any more. *Genes Dev.* **16**, 2179–2206 (2002).
3. Peters, J. M. The anaphase-promoting complex: proteolysis in mitosis and beyond. *Mol. Cell* **9**, 931–943 (2002).
4. Schmidt, A. et al. *Xenopus* polo-like kinase Plx1 regulates XErp1, a novel inhibitor of APC/C activity. *Genes Dev.* **19**, 502–513 (2005).
5. Tung, J. J. et al. A role for the anaphase-promoting complex inhibitor Emi2/XErp1, a homolog of early mitotic inhibitor 1, in cytostatic factor arrest of *Xenopus* eggs. *Proc. Natl Acad. Sci. USA* **102**, 4318–4323 (2005).
6. Shoji, S. et al. Mammalian Emi2 mediates cytostatic arrest and transduces the signal for meiotic exit via Cdc20. *EMBO J.* **25**, 834–845 (2006).
7. Sagata, N., Watanabe, N., Vande Woude, G. F. & Ikawa, Y. The c-*mos* proto-oncogene product is a cytostatic factor responsible for meiotic arrest in vertebrate eggs. *Nature* **342**, 512–518 (1989).
8. Bhatt, R. R. & Ferrell, J. E. Jr. The protein kinase p90 rsk as an essential mediator of cytostatic factor activity. *Science* **286**, 1362–1365 (1999).
9. Gross, S. D., Schwab, M. S., Lewellyn, A. L. & Maller, J. L. Induction of metaphase arrest in cleaving *Xenopus* embryos by the protein kinase p90Rsk. *Science* **286**, 1365–1367 (1999).
10. Miller, J. J. et al. Emi1 stably binds and inhibits the anaphase-promoting complex/cyclosome as a pseudosubstrate inhibitor. *Genes Dev.* **20**, 2410–2420 (2006).
11. Masui, Y. The elusive cytostatic factor in the animal egg. *Nature Rev. Mol. Cell Biol.* **1**, 228–232 (2000).
12. Kishimoto, T. Cell-cycle control during meiotic maturation. *Curr. Opin. Cell Biol.* **15**, 654–663 (2003).
13. Tunquist, B. J. & Maller, J. L. Under arrest: cytostatic factor (CSF)-mediated metaphase arrest in vertebrate eggs. *Genes Dev.* **17**, 683–710 (2003).
14. Liu, J. & Maller, J. L. Calcium elevation at fertilization coordinates phosphorylation of XErp1/Emi2 by Plx1 and CaMK II to release metaphase arrest by cytostatic factor. *Curr. Biol.* **15**, 1458–1468 (2005).
15. Rauh, N. R., Schmidt, A., Bormann, J., Nigg, E. A. & Mayer, T. U. Calcium triggers exit from meiosis II by targeting the APC/C inhibitor XErp1 for degradation. *Nature* **437**, 1048–1052 (2005).
16. Hansen, D. V., Tung, J. J. & Jackson, P. K. CaMKII and polo-like kinase 1 sequentially phosphorylate the cytostatic factor Emi2/XErp1 to trigger its destruction and meiotic exit. *Proc. Natl Acad. Sci. USA* **103**, 608–613 (2006).
17. Lorca, T. et al. Calmodulin-dependent protein kinase II mediates inactivation of MPF and CSF upon fertilization of *Xenopus* eggs. *Nature* **366**, 270–273 (1993).
18. Schmidt, A., Rauh, N. R., Nigg, E. A. & Mayer, T. U. Cytostatic factor: an activity that puts the cell cycle on hold. *J. Cell Sci.* **119**, 1213–1218 (2006).
19. Haccard, O. et al. Induction of metaphase arrest in cleaving *Xenopus* embryos by MAP kinase. *Science* **262**, 1262–1265 (1993).
20. Yamamoto, T. M., Iwabuchi, M., Ohsumi, K. & Kishimoto, T. APC/C–Cdc20-mediated degradation of cyclin B participates in CSF arrest in unfertilized *Xenopus* eggs. *Dev. Biol.* **279**, 345–355 (2005).
21. Murray, A. W. & Kirschner, M. W. Cyclin synthesis drives the early embryonic cell cycle. *Nature* **339**, 275–280 (1989).
22. Lohka, M. J. & Maller, J. L. Induction of nuclear envelope breakdown, chromosome condensation, and spindle formation in cell-free extracts. *J. Cell Biol.* **101**, 518–523 (1985).
23. Murray, A. W., Solomon, M. J. & Kirschner, M. W. The role of cyclin synthesis and degradation in the control of maturation promoting factor activity. *Nature* **339**, 280–286 (1989).
24. Ohsumi, K., Koyanagi, A., Yamamoto, T. M., Gotoh, T. & Kishimoto, T. Emi1-mediated M-phase arrest in *Xenopus* eggs is distinct from cytostatic factor arrest. *Proc. Natl Acad. Sci. USA* **101**, 12531–12536 (2004).
25. Gotoh, Y. et al. Characterization of recombinant *Xenopus* MAP kinase kinases mutated at potential phosphorylation sites. *Oncogene* **9**, 1891–1898 (1994).
26. Dupre, A., Jessus, C., Ozon, R. & Haccard, O. Mos is not required for the initiation of meiotic maturation in *Xenopus* oocytes. *EMBO J.* **21**, 4026–4036 (2002).
27. Dumont, J., Umbhauer, M., Rassinier, P., Hanauer, A. & Verlhac, M. H. p90Rsk is not involved in cytostatic factor arrest in mouse oocytes. *J. Cell Biol.* **169**, 227–231 (2005).
28. Lefebvre, C. et al. Meiotic spindle stability depends on MAPK-interacting and spindle-stabilizing protein (MISS), a new MAPK substrate. *J. Cell Biol.* **157**, 603–613 (2002).
29. Iwabuchi, M., Ohsumi, K., Yamamoto, T. M., Sawada, W. & Kishimoto, T. Residual Cdc2 activity remaining at meiosis I exit is essential for meiotic M–M transition in *Xenopus* oocyte extracts. *EMBO J.* **19**, 4513–4523 (2000).

nature

LETTERS

A direct link of the Mos–MAPK pathway to Erp1/Emi2 in meiotic arrest of *Xenopus laevis* eggs

Daigo Inoue[1,2]*, Munemichi Ohe[1]*, Yoshinori Kanemori[1], Toshiya Nobui[1] & Noriyuki Sagata[1,2]

In vertebrates, unfertilized eggs (or mature oocytes) are arrested at metaphase of meiosis II by a cytoplasmic activity called cytostatic factor (CSF)[1]. The classical Mos–MAPK pathway has long been implicated in CSF arrest of vertebrate eggs, but exactly how it exerts CSF activity remains unclear[2–4]. Recently, Erp1 (also called Emi2), an inhibitor of the anaphase-promoting complex/cyclosome (APC/C) required for degradation of the mitotic regulator cyclin B (ref. 5), has also been shown to be a component of CSF in both *Xenopus* and mice[6–8]. Erp1 is destroyed on fertilization or egg activation[9,10], like Mos[11]. However, despite these similarities the Mos–MAPK (mitogen-activated protein kinase) pathway and Erp1 are thought to act rather independently in CSF arrest[3,6,8]. Here, we show that p90rsk, the kinase immediately downstream from Mos–MAPK, directly targets Erp1 for CSF arrest in *Xenopus* oocytes. Erp1 is synthesized immediately after meiosis I, and the Mos–MAPK pathway or p90rsk is essential for CSF arrest by Erp1. p90rsk can directly phosphorylate Erp1 on Ser 335/Thr 336 both *in vivo* and *in vitro*, and upregulates both Erp1 stability and activity. Erp1 is also present in early embryos, but has little CSF activity owing, at least in part, to the absence of p90rsk activity. These results clarify the direct link of the classical Mos–MAPK pathway to Erp1 in meiotic arrest of vertebrate oocytes.

Recent studies in *Xenopus* oocytes show that Erp1 synthesis begins after germinal vesicle breakdown (GVBD; a hallmark for entry into meiosis I) and is essential for cyclin B reaccumulation (or cyclin-dependent Cdc2 re-activation) after meiosis I, or for the meiosis I (MI)/meiosis II (MII) transition and CSF establishment[12,13]. However, the classical Mos–MAPK pathway (Mos–MEK1–MAPK–p90rsk) is also implicated in the MI/MII transition and CSF establishment in *Xenopus* oocytes[14,15]. Thus, although the role of the Mos–MAPK pathway is questioned in CSF maintenance[6], it might have some link to Erp1 in the MI/MII transition or CSF establishment. If so, the Mos–MAPK pathway might be involved in Erp1 synthesis after MI, thereby enabling the MI/MII transition. Inhibition of MAPK activation in oocytes by the MEK1 inhibitor U0126 prevented cyclin B2 reaccumulation after MI, apparently by allowing the APC/C-mediated degradation[15] (Supplementary Fig. 1), but had little effect on Erp1 expression after GVBD or MI (Fig. 1a; and Supplementary Fig. 2). Thus, unexpectedly, Erp1 was expressed even without the Mos–MAPK pathway, but could not function for the MI/MII transition. (In this and some other experiments, Erp1 was treated with λ-protein phosphatase because of its heterogeneous hyperphosphorylation[12].) Interestingly, inhibition of MAPK activation by U0126 and that of Erp1 synthesis by antisense morpholino oligos prevented post-MI cyclin B2 reaccumulation, Cdc2 reactivation or Cdc27 phosphorylation (a marker of APC/C inhibition[15]) to a similar extent, and simultaneous inhibition of them had no additive inhibitory effects on the events tested (Fig. 1b). Furthermore, although premature, ectopic

expression of Erp1 at endogenous MII levels in MI could prevent cyclin B2 degradation after GVBD in control oocytes (as shown recently[12]), it could not do so in MAPK-inactivated oocytes (Fig. 1c). Thus, clearly, Erp1's activity during the MI/MII transition or CSF establishment depended largely on the Mos–MAPK pathway.

On fertilization, Mos is degraded but, when re-introduced into two-cell embryos, it can cause rapid cleavage arrest at metaphase, similar to metaphase II arrest in unfertilized eggs[16]. Erp1 is also degraded on fertilization or CSF release in egg extracts[6,9,12]. Interestingly, after degradation on fertilization, Erp1 was rapidly re-synthesized and accumulated during the first embryonic cell cycle, and remained constant at least up to the early blastula stage (Fig. 1d). Moreover, the embryonic Erp1 was heterogeneously hyperphosphorylated, as in unfertilized eggs, and was present at levels comparable to Erp1 in unfertilized eggs (Fig. 1d and Supplementary Fig. 3). However, morpholino oligo inhibition of Erp1 synthesis after fertilization had little if any effect on cyclin B2 levels or cell division in early cleavage embryos (Fig. 1e), indicating a very low CSF activity of embryonic Erp1. Notably, however, whereas ectopically expressed Mos could induce accumulation of a larger, mitotic form of cyclin B2 and rapid 2/4-cell arrest in control embryos (consistent with metaphase arrest[16]), it could not do so in Erp1-depleted embryos, which instead showed a Cdc2 Tyr 15-phosphorylated, interphasic arrest at much later 16/32-cell stages (perhaps owing to Wee1 activation by Mos[17]) (Fig. 1e). Thus, obviously, the long known CSF activity of (ectopic) Mos in early embryos[16] absolutely required the presence of endogenous Erp1. Together, these results strongly suggest that Erp1 activity during oocyte maturation strictly depends on the Mos–MAPK pathway.

Given these results, the Mos–MAPK pathway might be directly linked to Erp1. If so, the MAPK substrate p90rsk might directly phosphorylate and upregulate Erp1, because it has CSF activity in *Xenopus* oocytes[18,19]. Interestingly, an amino-terminal region (residues 200–400) of *Xenopus* Erp1 could be efficiently phosphorylated by *Xenopus* p90rsk *in vitro* (see Fig. 2b). This region (termed here NT) contains two neighbouring residues, Ser 335 and Thr 336, which are both evolutionarily well conserved and lie in the consensus p90rsk phosphorylation motif (Arg-X-X-Ser/Thr) (Fig. 2a). *In vitro* p90rsk kinase assays using glutathione *S*-transferase (GST)-fused NT (with Ala mutations) and [γ-^{32}P]ATP suggested that p90rsk can phosphorylate Ser 335 and Thr 336 equally well (Fig. 2b). p90rsk could also phosphorylate Erp1 NT on other sites but only inefficiently (Supplementary Fig. 4). On immunoblotting, p90rsk-phosphorylated GST–NT(WT) (wild-type), but not GST–NT(2A) (with double S335A/T336A mutations), could be specifically recognized by either anti-phospho-Ser 335 or -Thr 336 antibodies (Fig. 2c; and Supplementary Fig. 5). Moreover, even p90rsk-phosphorylated full-length Erp1(WT) but not Erp1(2A) could be specifically recognized by either
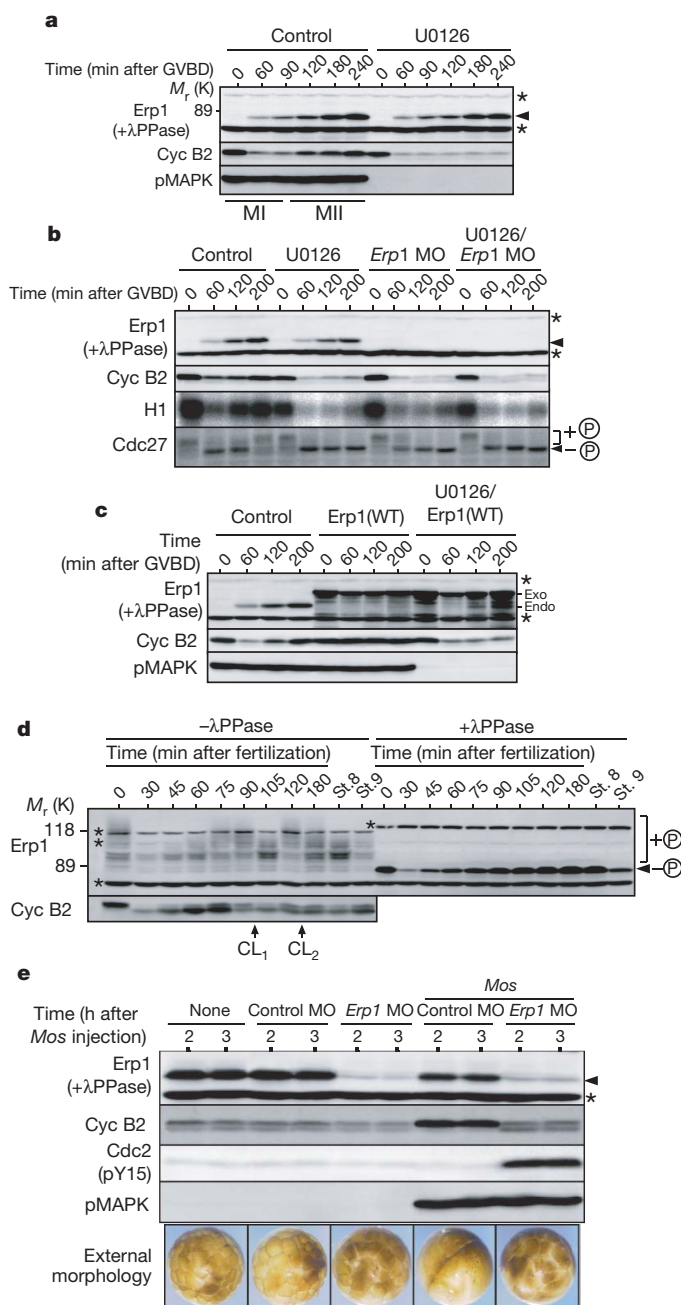
Figure 1 | Requirement of the Mos–MAPK pathway for Erp1 activity.
a, Immature oocytes pretreated or not with U0126 were treated with progesterone and subjected to immunoblotting (**b**–**e** also show immunoblots). For Erp1 protein, oocyte extracts were treated with λ-protein phosphatase[12]. The periods of MI and MII are shown for control oocytes. pMAPK, activated MAPK; arrowhead, (dephosphorylated) Erp1; asterisk, background protein. **b**, Immature oocytes treated with U0126 and/or injected with *Erp1* antisense morpholino oligos (MOs) were processed as in **a**. H1, Cdc2 kinase activity; + circled P, phosphorylated; − circled P, nonphosphorylated. **c**, Immature oocytes injected with 1.2 ng of mRNA encoding Myc₃–Erp1(WT) protein were cultured overnight and then processed as in **a**. For the expression patterns of Myc₃–Erp1(WT) (exo) and endogenous Erp1 (endo), see ref. 12. **d**, Extracts from unfertilized eggs and embryos were treated or not with λ-protein phosphatase (λPPase). + circled P, phosphorylated; − circled P, dephosphorylated; asterisk, background protein; CL₁, first cleavage; CL₂, second cleavage; st. 8, early blastula; st. 9, late blastula. **e**, Fertilized eggs were injected with either control or *Erp1* morpholino oligos at the one-cell stage, re-injected or not with 3 ng of *Mos* mRNA in both blastomeres at the two-cell stage, and cultured. Embryos were photographed 2.5 h after the *Mos* mRNA injection.

phospho-specific antibody (Fig. 2d). Thus, p90rsk can phosphorylate Erp1 at least on Ser 335 and Thr 336 *in vitro*.

We then asked whether Erp1 would be phosphorylated on Ser 335/Thr 336 by p90rsk during oocyte maturation. Endogenous Erp1 could be specifically recognized by both anti-pSer 335 and -pThr 336 antibodies after immunoprecipitation from unfertilized egg or post-GVBD oocyte extracts (Fig. 2e; and Supplementary Fig. 6). To confirm these results, we also expressed either inactive Cys583→Ala Erp1 (Erp1(C583A); refs 6, 12) or Erp1(C583A/2A; C583A with double S335A/T336A mutations) ~15-fold over endogenous levels in oocytes (which could mature normally; see cyclin B2 in Fig. 2f). By direct immunoblotting, both anti-pSer 335 and -pThr 336 antibodies could recognize—above background proteins and after GVBD—Erp1 (C583A) but not Erp1(C583A/2A)(Fig. 2f; and Supplementary Figs 7 and 8), indicating Ser 335/Thr 336 phosphorylation in normally maturing oocytes. Notably, in MAPK-inactivated oocytes, however, recognition of Erp1(C583A) by either antibody was abolished but could be restored significantly by ectopic expression of constitutively active p90rsk (which could also restore entry into MII) (Fig. 2g). Furthermore, when expressed in early embryos, neither Erp1(C583A) nor Erp1(C583A/2A) could be significantly recognized by either antibody; however, when co-expressed with active p90rsk, Erp1(C583A) but not Erp1(C583A/2A) could be specifically recognized (Fig. 2h; and see Supplementary Fig. 7). Thus, these results strongly suggest that p90rsk phosphorylates Erp1 on both Ser 335 and Thr 336 during oocyte maturation.

We then addressed whether Ser 335/Thr 336 phosphorylation could upregulate Erp1 during maturation. We first asked whether this phosphorylation could affect Erp1 stability. For this, we injected immature oocytes with a small amount (100 pg per oocyte) of full-length 3′-UTR-containing Erp1 messenger RNA (encoding either WT or S335A/T336A mutants) so that the encoded proteins would be expressed in an endogenous-like pattern and produced at low, endogenous levels without affecting the meiotic cell cycle[12]. Interestingly, Erp1(S335A) was produced approximately twofold less abundantly than Erp1(WT), particularly after metaphase II arrest (or after 3 h of GVBD), and Erp1(T336A) and Erp1(2A) were produced even (about three to fourfold) less abundantly than Erp1(WT) (Fig. 3a). These results were highly reproducible (not shown, but compare the levels of Erp1(C583A) and Erp1(C583A/2A) in Fig. 2f) and, as expected, the respective Erp1 mutants had significantly shorter half-lives than Erp1(WT) (Supplementary Fig. 9). Thus, Ser 335/Thr 336 phosphorylation was required for normal Erp1 stability during maturation and metaphase II arrest.

We next compared activities of the wild-type and the Erp1 mutants during maturation. To do this, we expressed the various Erp1 constructs in oocytes depleted of endogenous Erp1 by morpholino oligos, and then compared their abilities to rescue the MI/MII transition or CSF establishment[12]. The oocytes treated with morpholino oligos alone failed to reaccumulate cyclin B2 after 1 h of GVBD (or after MI) and showed severely depigmented morphology (indicative of parthenogenetic activation[12]); expression of Erp1(WT) in such oocytes could rescue both cyclin B2 reaccumulation and external morphology after MI (Fig. 3b). Notably, however, expression of either Erp1(S335A) or Erp1(T336A) could rescue those events only partially, and expression of Erp1(2A) could hardly do so (suggesting a cumulative effect of Ser 335 and Thr 336 phosphorylations on Erp1 activity) (Fig. 3b). In these experiments, despite injection of the same 100 pg per oocyte of mRNAs as in Fig. 3a, the Erp1 mutants were all produced approximately at the same (endogenous) levels as Erp1 (WT) after 2 h of GVBD, probably owing to either their increased stability or synthesis in the compromised, mitotic-like cell cycle after MI[12]. A similar explanation could be applied to the similar expression patterns of endogenous Erp1 in control and MAPK-inactivated oocytes (see Fig. 1a). In any event, Erp1(2A) was expressed at the same (endogenous) levels as Erp1(WT) even at the end of MI (or ~1 h after GVBD), but could not function for the MI/MII transition

(Fig. 3b). These results indicate that Ser 335/Thr 336 phosphorylation is also required for normal Erp1 activity during the MI/MII transition or CSF establishment.

We also examined the requirement for the Mos–MAPK pathway in CSF maintenance, which has been controversial[6,20–22]. First we tested whether mErp1(WT) or mErp1(2A), mutants that would resist SCF β-TrCP-dependent destruction upon CSF release[9,10] (see Methods), could maintain CSF arrest on calcium ionophore treatment of mature oocytes. Injection of a very small amount (30 pg per oocyte) of their individual mRNAs into immature oocytes produced mErp1(WT) and mErp1(2A) proteins at comparable, endogenous levels and at significantly higher levels than Erp1(WT) in mature oocytes (Fig. 3c, time 0). This implies that even during maturation and metaphase II arrest, parental Erp1(WT) and Erp1(2A) proteins

were both being significantly degraded, albeit at different rates (Figure 3a; and Supplementary Fig. 9), in an SCF β-TrCP-dependent manner. On calcium ionophore treatment, both mErp1(WT) and mErp1(2A) proteins were stable as expected, whereas endogenous Erp1 as well as Erp1(WT) was readily degraded. Despite its persistence, however, mErp1(2A) could not prevent cyclin B2 degradation, whereas mErp1(WT) could do so, if not perfectly (suggesting a partial non-proteolytic inactivation of Erp1 by calcium signalling; Fig. 3c; and Supplementary Fig. 10). Furthermore, mErp(2A), but not mErp(WT), failed to prevent dephosphorylation of the mitotic phosphatase Cdc25 and later inactivation of MAPK (Fig. 3c), indicating the requirement for Ser 335/Thr 336 phosphorylation of Erp1 in CSF maintenance. We then tested whether p90rsk was indeed required for CSF maintenance. Treating mature oocytes with the
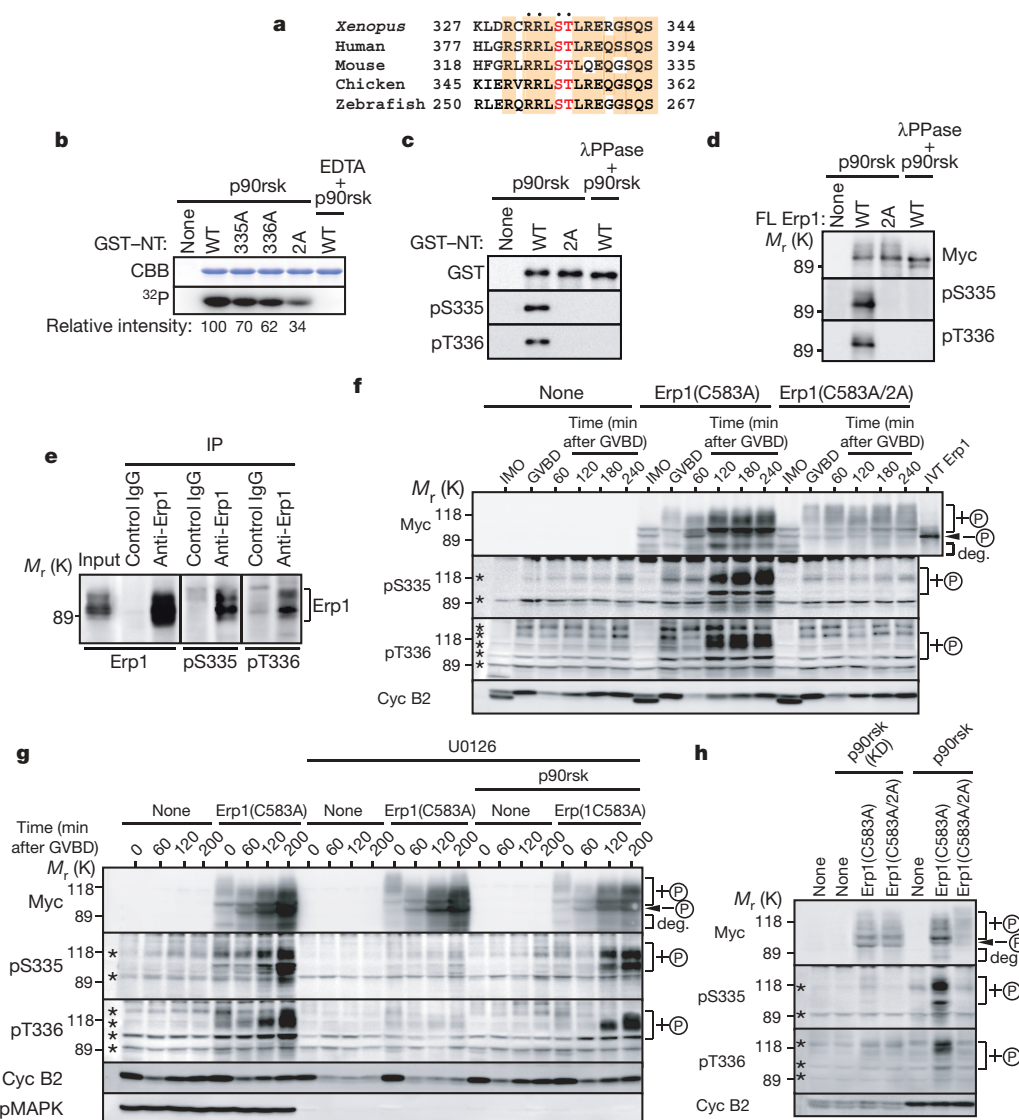

Figure 2 | Phosphorylation of Erp1 by p90rsk. a, Conservation of the consensus p90rsk phosphorylation motif (R-X-X-S/T) in Erp1 proteins. **b,** GST–NT proteins were incubated with *Xenopus* p90rsk2 (p90rsk) and [γ-^{32}P]ATP, subjected to SDS–polyacrylamide gel electrophoresis (PAGE), and stained with Coomassie brilliant blue (CBB) or autoradiographed (^{32}P). **c,** GST–NT proteins were phosphorylated by p90rsk and immunoblotted. **d,** Myc$_3$-tagged full-length (FL) Erp1 proteins were synthesized *in vitro*, immunoprecipitated, phosphorylated by p90rsk, and immunoblotted. **e,** Unfertilized egg extracts were subjected to immunoprecipitation (IP) with anti-Erp1 antibody and immunoblotted. **f,** Immature oocytes (IMO) were injected with 2 ng of mRNA encoding the indicated Myc$_3$–Erp1 proteins, cultured overnight, treated with progesterone and extracts immunoblotted.

Myc$_3$–Erp1 proteins before GVBD are those that were synthesized prematurely[12]. + circled P, phosphorylated; − circled P, nonphosphorylated; deg., degradation products; asterisk, background protein; IVT, *in vitro*-translated. **g,** Immature oocytes were injected with 2 ng of mRNA encoding Myc$_3$–Erp1(C583A), cultured overnight, re-injected with or without 50 ng of *p90rsk* mRNA, subsequently treated or not with U0126, and treated with progesterone. Extracts were immunoblotted. **h,** Fertilized eggs were injected with 4 ng of mRNA encoding the indicated Myc$_3$–Erp1 proteins at the one-cell stage, re-injected in both blastomeres at the two-cell stage with 20 ng of mRNA encoding either p90rsk or kinase-dead (KD) p90rsk, and cultured for 2.5 h.
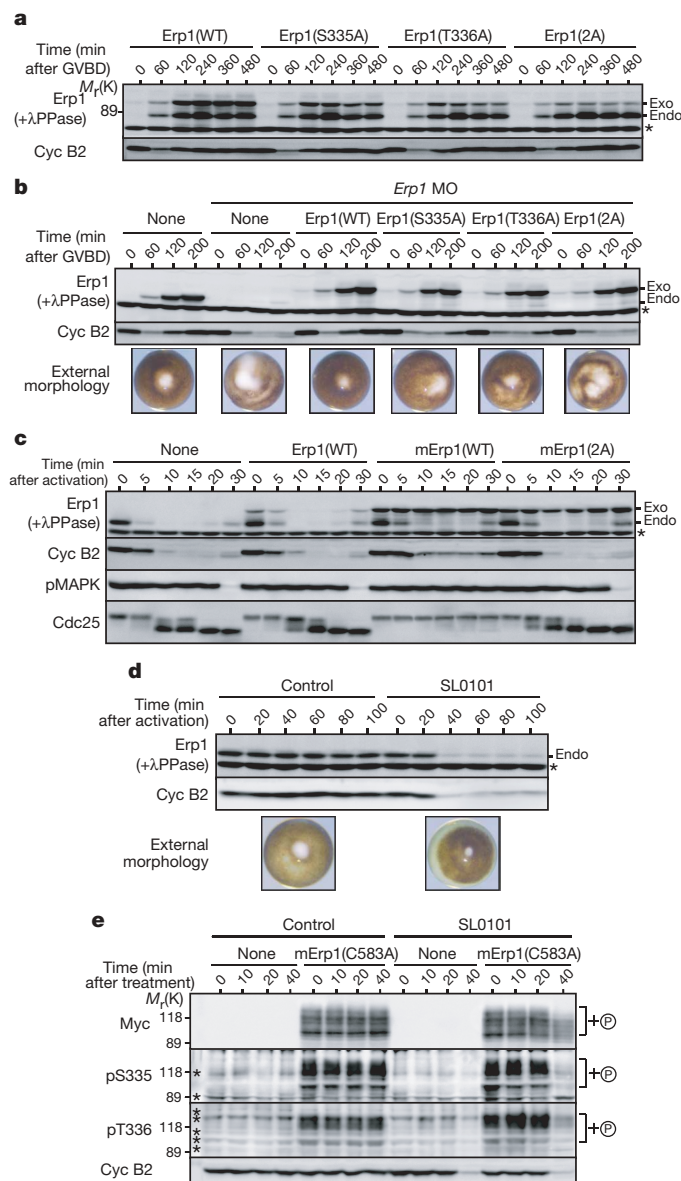
Figure 3 | Upregulation of Erp1 stability and activity by Ser 335/Thr 336 phosphorylation. a, Immature oocytes were injected with 100 pg of mRNA encoding the indicated Myc3–Erp1 proteins, cultured overnight, treated with progesterone, and subjected to immunoblotting (**b**–**e** also show immunoblots). Asterisk, background protein; exo, exogenous; endo, endogenous. **b**, Immature oocytes were injected with Erp1 morpholino oligos together with or without 100 pg of morpholino-oligo-resistant mRNA encoding the indicated Myc3–Erp1 proteins, processed as in **a**, and also photographed 4 h after GVBD. **c**, Immature oocytes were injected with 30 pg of mRNA encoding the indicated Myc3–Erp1 proteins, cultured overnight, treated with progesterone, and, 4 h after GVBD, treated with calcium ionophore to induce egg activation. **d**, Mature oocytes were treated or not with 4 mM SL0101 and also photographed 30 min after the treatment. **e**, Immature oocytes were injected with 600 pg of mRNA encoding Myc3–mErp1(C583A/2A), cultured overnight, treated with protesterone, and, 4 h after GVBD, treated or not with SL0101. + circled P, phosphorylated; asterisk, background protein.

recently isolated p90rsk inhibitor SL0101 (ref. 23) could inhibit endogenous p90rsk activity (Supplementary Fig. 11) and induce degradation of both endogenous Erp1 and cyclin B2, as well as surface contraction of the oocytes (Fig. 3d). (In maturing oocytes, SL0101 could significantly inhibit post-MI, cyclin B2 reaccumulation; Supplementary Fig. 12.) SL0101 could also induce dephosphorylation of Ser 335/Thr 336 of a proteolysis-resistant, inactive Erp1 mutant
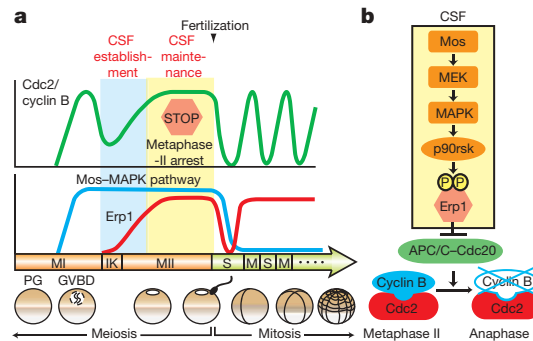

Figure 4 | Model of the relationship between the Mos–MAPK pathway and Erp1 in metaphase II arrest of *Xenopus* oocytes. a, Changes in levels of the Mos–MAPK pathway (activity), Cdc2/cyclin B and Erp1 during oocyte maturation and after fertilization. Erp1 has strong CSF activity only when the Mos–MAPK pathway is operating. PG, progesterone; IK, interkinesis (MI/MII transition); S, DNA synthesis phase; M, mitotic phase. **b**, The main pathway leading to CSF arrest. CSF consists of Mos, the three transducer kinases and the effector protein Erp1, which binds and inhibits the APC/C.

(mErp1(C583A)), as well as inducing its partial degradation, which is not enough to explain the substantial dephosphorylation (Fig. 3e). Taken together, these results strongly suggest that the Mos–MAPK pathway is required for the maintenance of CSF arrest in mature oocytes, with p90rsk upregulating both Erp1 stability and activity. Erp1 has been shown to inhibit the APC/C (ref. 6). Our additional data suggest that Ser 335/Thr 336 phosphorylation promotes inhibitory binding of Erp1 to the APC/C (data not shown).

Although the Mos–MAPK pathway operates during both MI and MII and somehow functions in CSF arrest in *Xenopus* oocytes[2,3,14], Erp1 is expressed only after MI and acts as a key component of CSF[6,7,12] (Fig. 4a). Our study using intact oocytes demonstrates that the Mos–MAPK pathway is directly linked to and upregulates Erp1 for both CSF establishment and maintenance, thus answering the long-standing question of how Mos exerts CSF activity[2,24]. By definition, (amphibian) CSF appears during oocyte maturation (to cause metaphase II arrest), disappears after fertilization (never to reappear), and can cause metaphase arrest when re-introduced into early embryos[24]. Among the known CSF components, only Mos seems to satisfy all of the criteria for CSF[16,24]. However, for Mos to function as CSF, the three downstream kinases—MEK1, MAPK and p90rsk—and Erp1 are all essential (refs 2, 3, 6 and this work), and, as shown here, p90rsk directly targets Erp1 for both its stability and activity. Thus, it seems that (amphibian) CSF is composed of Mos, its transducer kinases and the effector protein Erp1 (Fig. 4b).

In mice as well, both the Mos–MAPK pathway and Erp1 (or Emi2) are required for CSF arrest of mature oocytes[4,8,25]. In this species, however, p90rsk itself is not likely to be essential for CSF arrest[26]. Therefore, if the Mos–MAPK pathway is still linked to Erp1 in mouse oocytes, then Erp1 might be regulated by yet another kinase, for example, MAPK itself[27] or the immediately downstream p90rsk-related kinase MSK (ref. 28).

METHODS

Oocytes, eggs, embryos and CSF extracts. Oocytes, eggs and embryos were prepared, cultured and microinjected as described[29]. Oocyte maturation and egg activation were induced by progesterone and calcium ionophore (A23187), respectively, as described[29]. In some experiments, oocytes were treated with 200 µM U0126 (Promega) before progesterone treatment[15], and mature oocytes were treated with 4 mM SL0101 (Toronto Research Chemicals). CSF extracts from unfertilized eggs were prepared as described[6].

Complementary DNAs and *in vitro* transcription and translation. cDNAs encoding *Xenopus* Erp1 (tagged with three Myc epitopes), *Xenopus* Mos or a constitutively active form of *Xenopus* p90rsk2 were prepared and *in vitro* transcribed[12,29]. To generate proteolysis-resistant mutants of Erp1 (mErp1), DSGYSDS and DSAFHS, the known and potential β-TrCP-binding motifs in *Xenopus* Erp1 (ref. 6), were mutated to DAGYSDA and DAAFHA, respectively.

In vitro translation of mRNAs was performed using wheat germ extracts (Promega).

Morpholino oligonucleotides, full-length *Erp1* mRNA, morpholino-oligo-resistant mRNA and recombinant proteins. These are described in Supplementary Information.

Antibodies, immunoblotting and immunoprecipitation. Anti-*Xenopus* Erp1 phospho-Ser 335 and -Thr 336 antibodies were raised in rabbits against phosphorylated peptides RRLpSTLRERG (S being phosphorylated) and RRLSpTLRERGSQ (T being phosphorylated), respectively. Other antibodies and immunoblotting were described[12]. For immunoprecipitation of endogenous Erp1, 100 μl of CSF extracts supplemented with 10 mM NaF and 2 mM ATP was first incubated with 4 μg of anti-Erp1 antibody for 30 min at 4 °C, and then incubated with 20 μl of slurry of Protein A beads (PIERCE) for 30 min at 4 °C. The beads were washed three times with a wash buffer (50 mM Tris-HCl at pH7.5, 10 mM MgCl$_2$, 1 mM DTT).

In vitro **kinase assays.** These are described in Supplementary Information.

1. Masui, Y. & Markert, C. L. Cytoplasmic control of nuclear behavior during meiotic maturation of frog oocytes. *J. Exp. Zool.* **177**, 129–145 (1971).
2. Sagata, N. Meiotic metaphase arrest in animal oocytes: its mechanisms and biological significance. *Trends Cell Biol.* **6**, 22–28 (1996).
3. Tunquist, B. J. & Maller, J. L. Under arrest: cytostatic factor (CSF)-mediated metaphase arrest in vertebrate eggs. *Genes Dev.* **17**, 683–710 (2003).
4. Brunet, S. & Maro, B. Cytoskeleton and cell cycle control during meiotic maturation of the mouse oocyte: integrating time and space. *Reproduction* **130**, 801–811 (2005).
5. Peters, J. M. The anaphase promoting complex/cyclosome: a machine designed to destroy. *Nature Rev. Mol. Cell Biol.* **7**, 644–656 (2006).
6. Schmidt, A. *et al. Xenopus* polo-like kinase Plx1 regulates XErp1, a novel inhibitor of APC/C activity. *Genes Dev.* **19**, 502–513 (2005).
7. Tung, J. J. *et al.* A role for the anaphase-promoting complex inhibitor Emi2/XErp1, a homolog of early mitotic inhibitor 1, in cytostatic factor arrest of *Xenopus* eggs. *Proc. Natl Acad. Sci. USA* **102**, 4318–4323 (2005).
8. Shoji, S. *et al.* Mammalian Emi2 mediates cytostatic arrest and transduces the signal for meiotic exit via Cdc20. *EMBO J.* **25**, 834–845 (2006).
9. Rauh, N. R., Schmidt, A., Bormann, J., Nigg, E. A. & Mayer, T. U. Calcium triggers exit from meiosis II by targeting the APC/C inhibitor XErp1 for degradation. *Nature* **437**, 1048–1052 (2005).
10. Hansen, D. V., Tung, J. J. & Jackson, P. K. CaMKII and polo-like kinase 1 sequentially phosphorylate the cytostatic factor Emi2/XErp1 to trigger its destruction and meiotic exit. *Proc. Natl Acad. Sci. USA* **103**, 608–613 (2006).
11. Watanabe, N., Hunt, T., Ikawa, Y. & Sagata, N. Independent inactivation of MPF and cytostatic factor (Mos) upon fertilization of *Xenopus* eggs. *Nature* **352**, 247–248 (1991).
12. Ohe, M., Inoue, D., Kanemori, Y. & Sagata, N. Erp1/Emi2 is essential for the meiosis I to meiosis II transition in *Xenopus* oocytes. *Dev. Biol.* **303**, 157–164 (2007).
13. Liu, J., Grimison, B., Lewellyn, A. L. & Maller, J. L. The anaphase-promoting complex/cyclosome inhibitor Emi2 is essential for meiotic but not mitotic cell cycles. *J. Biol. Chem.* **281**, 34736–34741 (2006).
14. Furuno, N. *et al.* Suppression of DNA replication via Mos function during meiotic divisions in *Xenopus* oocytes. *EMBO J.* **13**, 2399–2410 (1994).
15. Gross, S. D. *et al.* The critical role of the MAP kinase pathway in meiosis II in *Xenopus* oocytes is mediated by p90^{Rsk}. *Curr. Biol.* **10**, 430–438 (2000).
16. Sagata, N., Watanabe, N., Vande Woude, G. F. & Ikawa, Y. The c-*mos* proto-oncogene product is a cytostatic factor responsible for meiotic arrest in vertebrate eggs. *Nature* **342**, 512–518 (1989).
17. Murakami, M. S., Copeland, T. D. & Vande Woude, G. F. Mos positively regulates Xe-Wee1 to lengthen the first mitotic cell cycle of *Xenopus*. *Genes Dev.* **13**, 620–631 (1999).
18. Bhatt, R. R. & Ferrell, J. E. Jr. The protein kinase p90 rsk as an essential mediator of cytostatic factor activity. *Science* **286**, 1362–1365 (1999).
19. Gross, S. D., Schwab, M. S., Lewellyn, A. L. & Maller, J. L. Induction of metaphase arrest in cleaving *Xenopus* embryos by the protein kinase p90Rsk. *Science* **286**, 1365–1367 (1999).
20. Reimann, J. D. & Jackson, P. K. Emi1 is required for cytostatic factor arrest in vertebrate eggs. *Nature* **416**, 850–854 (2002).
21. Furuno, N., Ogawa, Y., Iwashita, J., Nakajo, N. & Sagata, N. Meiotic cell cycle in *Xenopus* oocytes is independent of cdk2 kinase. *EMBO J.* **16**, 3860–3865 (1997).
22. Yamamoto, T. M., Iwabuchi, M., Ohsumi, K. & Kishimoto, T. APC/C-Cdc20-mediated degradation of cyclin B participates in CSF arrest in unfertilized *Xenopus* eggs. *Dev. Biol.* **279**, 345–355 (2005).
23. Smith, J. A. *et al.* Identification of the first specific inhibitor of p90 ribosomal S6 kinase (RSK) reveals an unexpected role for RSK in cancer cell proliferation. *Cancer Res.* **65**, 1027–1034 (2005).
24. Masui, Y. The elusive cytostatic factor in the animal egg. *Nature Rev. Mol. Cell Biol.* **1**, 228–232 (2000).
25. Madgwick, S., Hansen, D. V., Levasseur, M., Jackson, P. K. & Jones, K. T. Mouse Emi2 is required to enter meiosis II by reestablishing cyclin B1 during interkinesis. *J. Cell Biol.* **174**, 791–801 (2006).
26. Dumont, J., Umbhauer, M., Rassinier, P., Hanauer, A. & Verlhac, M. H. p90Rsk is not involved in cytostatic factor arrest in mouse oocytes. *J. Cell Biol.* **169**, 227–231 (2005).
27. Lefebvre, C. *et al.* Meiotic spindle stability depends on MAPK-interacting and spindle-stabilizing protein (MISS), a new MAPK substrate. *J. Cell Biol.* **157**, 603–613 (2002).
28. Hauge, C. & Frodin, M. RSK and MSK in MAP kinase signalling. *J. Cell Sci.* **119**, 3021–3023 (2006).
29. Inoue, D. & Sagata, N. The Polo-like kinase Plx1 interacts with and inhibits Myt1 after fertilization of *Xenopus* eggs. *EMBO J.* **24**, 1057–1067 (2005).

LETTERS

Expanding the diversity of chemical protein modification allows post-translational mimicry

Sander I. van Kasteren[1], Holger B. Kramer[1], Henrik H. Jensen[1], Sandra J. Campbell[2], Joanna Kirkpatrick[1], Neil J. Oldham[1], Daniel C. Anthony[2] & Benjamin G. Davis[1]

One of the most important current scientific paradoxes is the economy with which nature uses genes. In all higher animals studied, we have found many fewer genes than we would have previously expected. The functional outputs of the eventual products of genes seem to be far more complex than the more restricted blueprint. In higher organisms, the functions of many proteins are modulated by post-translational modifications (PTMs)[1]. These alterations of amino-acid side chains lead to higher structural and functional protein diversity and are, therefore, a leading contender for an explanation for this seeming incongruity. Natural protein production methods typically produce PTM mixtures within which function is difficult to dissect or control. Until now it has not been possible to access pure mimics of complex PTMs. Here we report a chemical tagging approach that enables the attachment of multiple modifications to bacterially expressed (bare) protein scaffolds: this approach allows reconstitution of functionally effective mimics of higher organism PTMs. By attaching appropriate modifications at suitable distances in the widely-used LacZ reporter enzyme scaffold, we created protein probes that included sensitive systems for detection of mammalian brain inflammation and disease. Through target synthesis of the desired modification, chemistry provides a structural precision and an ability to retool with a chosen PTM in a manner not available to other approaches. In this way, combining chemical control of PTM with readily available protein scaffolds provides a systematic platform for creating probes of protein–PTM interactions. We therefore anticipate that this ability to build model systems[2] will allow some of this gene product complexity to be dissected, with the aim of eventually being able to completely duplicate the patterns of a particular protein's PTMs from an *in vivo* assay into an *in vitro* system.

Glycosylation is the most diverse PTM process, and numerous examples have illustrated its key role in the modulation of protein properties[3]. For systematic study of the effects of protein glycosylation, efficient methods for the synthesis of pure, complex, differentially glycosylated proteins or mimics are needed[4]. Although we have described examples of site-selective chemical protein glycosylation[4,5], this method was limited to incorporation of only one type of PTM or sugar. In principle, the development of a second chemical protein modification reaction that is orthogonal (mutually compatible) with this first reaction would allow dual and differential modification. Here we describe the development of two such compatible reactions and their use in a convergent strategy that for the first time allows the synthesis of differentially modified proteins. Our approach (Supplementary Fig. S1) combined the incorporation of natural (cysteine (Cys)) and unnatural amino acids (azidohomoalanine (Aha)[6] and homopropargylglycine (Hpg)[7]) as chemical tags at sites for selective attachment of modifications using these two reactions. In this way, dual combination of selective synthetic methodology with site-directed gene mutagenesis allowed modification at multiple, predetermined sites in a peptide sequence.

We chose triplet codons for Cys and methionine (Met) to code for our chemical tags (Supplementary Fig. S1b) because of their low incidence in the gene sequences of proteins[8]. We combined site-directed gene mutagenesis with residue-specific replacement of methionine by analogues[6,7] to access differentially tagged precursor proteins. Representative sequences and structures were chosen: the protein fold present in ~10% of proteins (in $(\alpha/\beta)_8$-TIM-barrel protein, SSβG; ref. 9) and the most abundant mammalian glycoprotein, Tamm-Horsfall protein (THp)[10]. Importantly, the enzymatic galactosidase activity of SSβG is the same as that of LacZ-type reporter activity[11], and allowed us to monitor the functional effect of chemical protein modification and to verify a benign response.

For a first modification reaction, we chose glycoconjugation of Cys with glycomethanethiosulphonates (glyco-MTS)[5]. As a second key step in our expansion of protein modification methodology, we required a chemoselective reaction for modification of Met analogues such as Hpg or Aha. A promising candidate reaction, which needed to be compatible with not only our first reaction but also protein environments, was the Cu(I)-catalysed modification[12,13] of the Huisgen cycloaddition (CCHC). This reaction has found numerous applications (see Supplementary Information) owing to its compatibility with aqueous conditions. However, protein glycosylation and complete, quantitative, protein modifications have not been demonstrated.

A prerequisite for incorporation of representative natural glycans with free hydroxyl groups into proteins in such a putative glyco-cycloaddition (glyco-CCHC) would be the use of unprotected glycosylating reagents. Therefore, as potential reaction partners to alkyne(Hpg)- and azide(Aha)-containing proteins, representative azido-sugars **1–4** (GlcNAcβ–N$_3$ **1**, disaccharide GlcNAcβ(1,4)-GlcNAcβ–N$_3$ **2**, spacer-linked mannoside Manα-O-CH$_2$CH$_2$–N$_3$ **3**, and glucoside Glcβ-O-CH$_2$CH$_2$–N$_3$ **4**) and the alkynyl-C-glycosides **5** and **6** (Galβ–CCH **5**, and spacer-linked Glcβ-O-CH$_2$CH$_2$–CCH **6**), respectively, were synthesized from their parent carbohydrates (Supplementary Information). To survey reaction conditions and to demonstrate first reactions with unprotected glycans, a range of model amino acid glyco-CCHC reactions were successfully evaluated (Supplementary Information). Having demonstrated compatibility with aqueous buffer and the absence of sugar protecting groups, we then tested the glyco-CCHC on proteins.

To create model protein substrate for optimization studies, we engineered TIM barrel protein SSβG to replace unneeded Met and Cys (ref. 8) with near-isosteric amino acids isoleucine (Ile) and serine

[1]Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford OX1 3TA, UK. [2]Department of Pharmacology, University of Oxford, Mansfield Road, Oxford OX1 3QT, UK.

(Ser), respectively. Expression of a ten-point $(Met)_{10}(Cys)_1 \rightarrow (Met43)_1(Ile)_9(Ser)_1$ mutant plasmid in the Met-auxotrophic expression host *Escherichia coli* B834(DE3) in the presence of Met analogue Aha allowed formation of protein SSβG-Aha43 incorporating a single azide chemical tag in the side chain of azidohomoalanine (Aha) at single Met-encoded site 43, which despite functional group alteration at ten residues remained an active, competent enzyme. Optimization of glyco-CCHC reaction on monoazido protein SSβG-Aha43 revealed the critical use of a pre-formed solution of highly pure (99.999%) Cu(I) source stabilized by coordinating ligand. These generally applicable conditions allowed the first complete conversion (>95%) modifications of proteins in a site-selective manner using CCHC. Alkyne-containing protein SSβG-Hpg1 was also produced, and used as substrate for inverse mode (alkyne on protein, azide on sugar) glyco-CCHC reactions with anomeric glycosyl azides **1** and **2**; these reacted completely to yield glycoproteins that contain the carbohydrate 'root' of the conserved glycan core found in *N*-linked glycoproteins, the motifs GlcNAcβ-*N*– and GlcNAcβ(1,4)GlcNAcβ-*N*–. Precise multi-site glycosylation was also demonstrated; mutant proteins with multiple tags allowed multi-site modification (see Supplementary Table S1 for product range).

Next, multi-site modification with different glycans was demonstrated with proteins containing two different chemical tags. In tagged TIM-barrel protein SSβG-Aha43-Cys439 containing a thiol tag at site 439 and an azide tag at site 43, the Cys 439 thiol group was selectively α-glucosylated using the glyco-MTS[5] reaction and then, in a second step, the Aha 43 azide group was β-galactosylated using the glyco-CCHC reaction (Supplementary Fig. S2). The ability to carry out Cu(I)-catalysed glyco-CCHC in the presence of attached, disulphide-linked saccharide shows the advantages of optimized conditions; triazole formation was promoted without excess reducing agents that might disrupt disulphides. Similarly, sulphur-containing functionality did not impair rate or efficiency of CCHC reactions. Full orthogonality was demonstrated by successfully reversing the order of reactions: glyco-CCHC then glyco-MTS. To demonstrate substrate generality, the same method was also applied to fragments of THp. THp$_{295-306}$ contains *N*-glycosylation (Asn 298) and *O*-glycosylation (Ser 303) sites[10]. Our two-step differential glycoconjugation strategy gave complete, clean conversion using Cys as a tag for *O*-glycosylation and either Aha or Hpg as tags for *N*-glycosylation (THp$_{295-306}$-Aha298-Cys303 or THp$_{295-306}$-Hpg298-Cys303) (Fig. 1).

The benign nature of our chemical approach was confirmed by assessment of the native function of proteins after modification; all SSβG enzymes were active. Furthermore, these glycosylated proteins, as well as retaining inherent function, showed additional protein–protein binding capabilities with lectins (Supplementary Fig. S3) mediated by the newly-introduced carbohydrate side chains. This was highlighted by successful detection of GlcNAc-modification of SSβG-GlcNAc1 (Supplementary Information) using the same GlcNAc-binding lectin, WGA, that has been used to detect the natural PTM of GlcNAc-modification[14].

Next, we used our approach of differential chemical post-translational modification to reconstitute function in a mimic of a human protein known to possess differential PTMs important to its function, P-selectin-glycoprotein-ligand-1 (PSGL-1). PSGL-1 is a transmembrane protein expressed as a homodimer on leukocyte microvilli[15,16]. It is the best characterized of the ligands of P-selectin and plays a critical role in the initial rolling-adhesion phases of the inflammatory response[16]. Studies[15] of the naturally-occurring protein have identified two key PTMs: first, an *O*-linked oligosaccharide modification containing trisaccharide sialylated-*N*-acetyllactosamine (siaLacNAc) or extended, fucosylated, tetrasaccharide variant sialyl-Lewis-x (sLe^x) attached to Thr 57; and second, a sulphate attached to Tyr 48. Despite synthesis of glycosulphopeptides[17], no pure protein ligand has been available. We felt that the reconstitution of a bacterially-expressed protein to a form that displays activity that

mimics PSGL-1 would be an important test of our chemical methodology. Examination of structural parameters[15] of PSGL-1's interaction with P-selectin revealed a PTM-to-PTM spacing that might be achieved in a mimic containing a sulphotyrosine mimic group at position 439 and a sialylated glycan (siaLacNAc or sLe^x) mimic at position 43 (Fig. 2) in the LacZ-type reporter enzyme SSβG. This was achieved through design and synthesis (Supplementary Information) of novel Cys-modifying reagent Tys-MTS that allowed introduction of the group 'Tys', and novel trisaccharide and tetrasaccharide reagents siaLacNAc-alkyne, sLe^x-alkyne (Supplementary Fig. S4). The flexibility of the method also allowed truncated glycan structures to be incorporated at position 439. *In vitro* assay[18] (Supplementary Information) of these differently modified putative PSGL-1 mimics dissected contributions to binding to P-selectin (Fig. 3).

Consistent with previous observations[17], protein modified with just sulphated tyrosine mimic 'Tys' shows only modest P-selectin binding. While the glycan chain is being built (no glycan→GlcNAc→LacNAc), binding is largely unimproved. However, extension of glycan to trisaccharide, giving the combined presence of siaLacNAc and 'Tys', resulted in a doubly-PTM-modified PSGL-1 mimic with strong synergistic binding. This was amplified further still in a mimic bearing an additional fucosyl residue (in a protein containing both tetrasaccharide sLe^x and 'Tys'); once again, binding was synergistic and highlighted the critical importance of both suitable glycan and 'Tys' moieties. The flexibility of chemical modification allowed removal of the sulphated tyrosine mimic by disulphide cleavage to yield a PSGL-1 mimic bearing siaLacNAc or sLe^x only, which showed significantly reduced binding.

The resulting protein mimics bearing 'Tys' plus siaLacNAc or sLe^x were designed not only as mimics of PSGL-1 but also as enzymes with galactosidase (*lacZ*-type) activity capable of visualizing widely-used probe X-Gal;[11] this bestows upon them not only binding ability but also *lacZ*-type reporter activity. Hence, these mimics were called PSGL-LacZ and PSGL*-LacZ, respectively (Fig. 2). PSGL-LacZ and PSGL*-LacZ were explored as probes of *in vivo* inflammatory brain
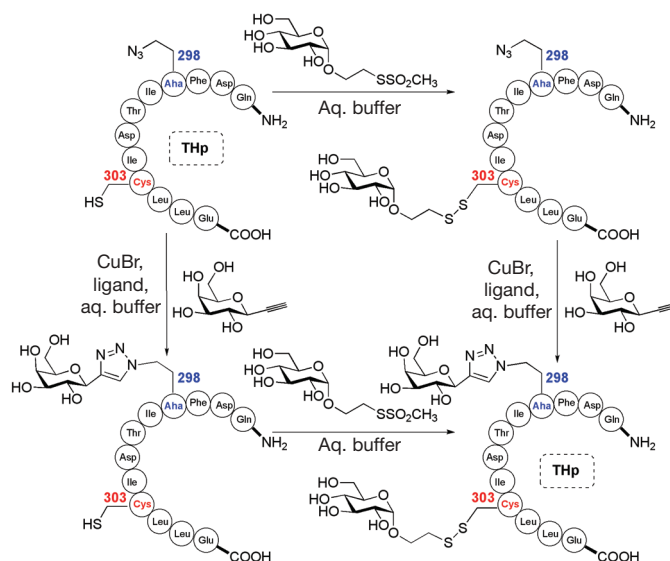

Figure 1 | Differential multi-site chemical protein modification. Complete and entirely selective protein modification is seen through the combination of the glyco-MTS glycosylation reaction with the glyco-CCHC reaction. The mutual compatibility of these two methods (orthogonality) is shown by the ability to conduct these two sequential glycosylations in either order in the differential glycosylation of a mammalian THp fragment: galactosylation at natural *N*-linked glycosylation site 298 (blue tag site) and glucosylation at *O*-linked site 303 (red tag site). The general approach can be applied to other proteins (see Supplementary Fig. S2a for SSβG), and can utilize alternative chemical tags such as Hpg (see Supplementary Information).
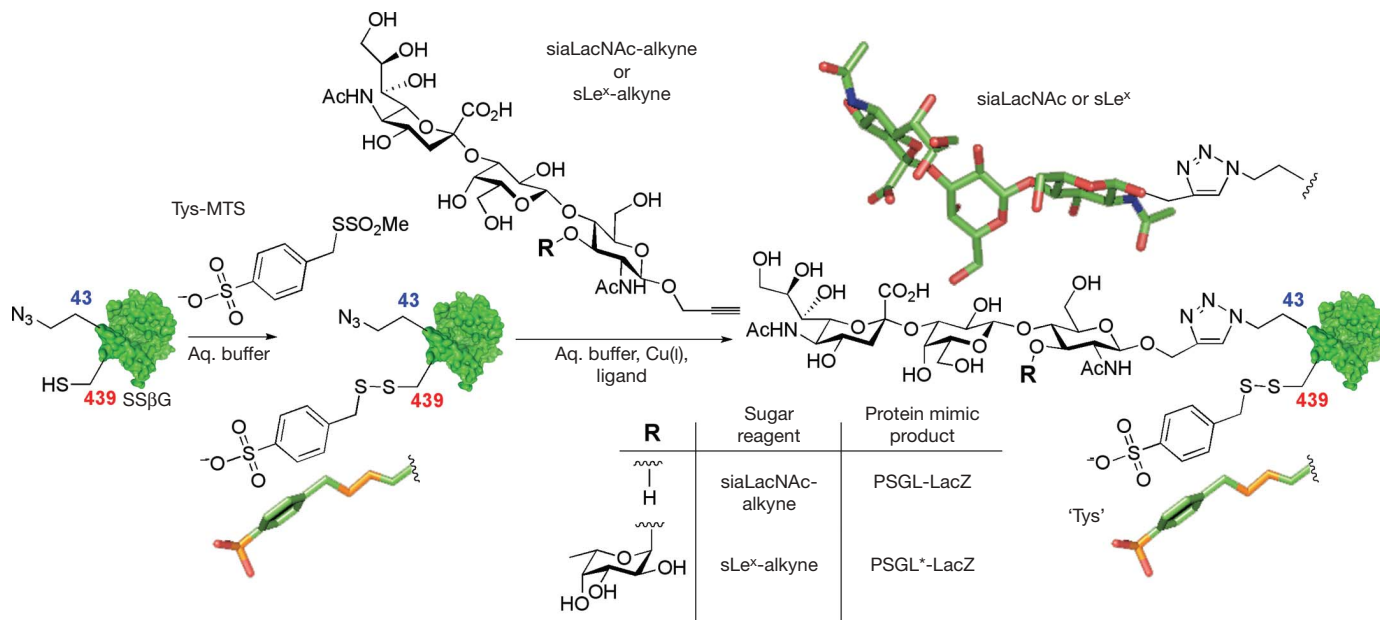
Figure 2 | Creating functional mimics of PSGL-1. Putative PSGL-1 mimics were created using sequential differential chemical modification: first, disulphide ligation (at red tag site) with the sulphotyrosine mimic group 'Tys', and then glyco-CCHC (at blue tag site) with either trisaccharide siaLacNAc (R = H) to create PSGL-LacZ or tetrasaccharide sLe[x] (R = α-L-fucosyl) to create PSGL*-LacZ. Stick models of modifications shown next to chemical strctures. Protein scaffold (not to scale) shown as surface model (green) of PDB structure 1gow.

lesions. Normal brain is not thought to contain pre-stored P-selectin[19], in Weibel-Palade bodies, but chronic inflammatory lesions in the brain are known to be associated with expression of P-selectin. PSGL-LacZ and PSGL*-LacZ were tested in rat models for both acute[20] and chronic[21] interleukin-1β (IL-1β)-induced inflammation. Eight days after injection of replication-deficient, IL-1β-expressing adenovirus, the majority of vessels in the lesion were P-selectin positive (Fig. 4) using PSGL-LacZ or PSGL*-LacZ. Even in the acute inflammation model, highly localized expression of P-selectin was detectable 10 min after intracerebral injection of 1 ng IL-1β within and adjacent to the injection site; the contralateral hemisphere showed no P-selectin expression. This suggests that PSGL-LacZ and PSGL*-LacZ are potent reporter systems for P-selectin; but this detection of early P-selectin expression after acute injury also suggests that low-level constitutive expression of P-selectin is present in cerebral vasculature, and casts doubt on experiments using less

sensitive methods[22]. It should be noted that this detection of acute inflammation using PSGL-LacZ or PSGL*-LacZ surpasses antibody-based methods in terms of sensitivity[23]. The greater sensitivity of fucosylated protein PSGL*-LacZ allowed its use at concentrations fivefold lower than those of PSGL-LacZ. All controls (uninduced, no PSGL-LacZ, sTyr-mimicry only) showed no P-selectin detection. Even under conditions of shear stress, probed through direct carotid injection (Fig. 4), *in vivo* detection of inflammation was seen. The power of these synthetically reconstituted mimics PSGL-LacZ, PSGL*-LacZ was next examined in detection of disease.

Malaria is endemic in 106 countries of the world with ~400 million infected individuals[24]; mortality is often the result of cerebral malaria[25]. P-selectin mediated recruitment of platelets[26] is one of the underlying histopathological hallmarks of cerebral malaria[25], implicated in progression and survival prognosis. As a further test of their ability to act as reporters for *in vivo* disease, PSGL-LacZ
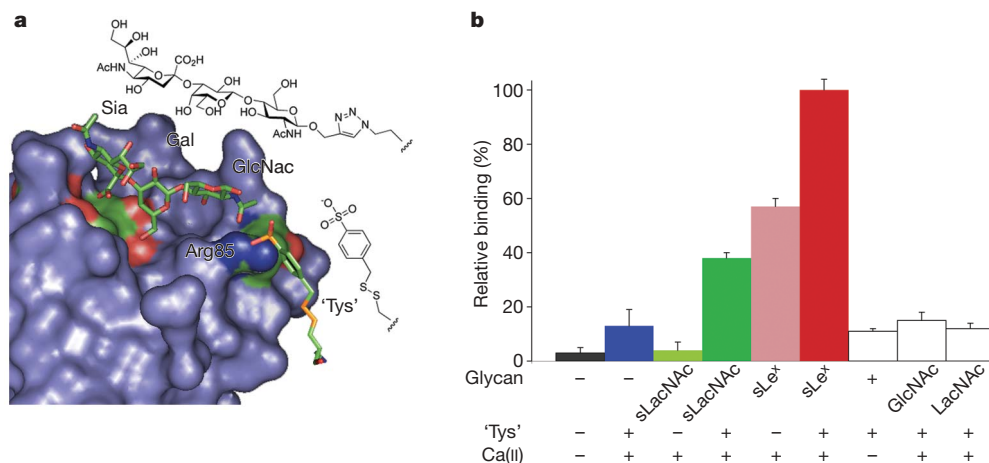

Figure 3 | Binding of mimics to human P-selectin. a, EGF region of P-selectin (grey) modelled with TIM-barrel protein SSβG (only modified side chains shown) differentially chemically modified with trisaccharide siaLacNAc (Sia–Gal–GlcNAc) at position 43 and sulpho-tyrosine mimic 'Tys' at position 439. **b**, Relative binding (Supplementary Information,

condition B) shows the successful additivity of the chemically introduced PTM mimics 'Tys', sia(s)LacNAc and sLe[x] in model TIM-barrel protein to P-selectin, and allows dissection of the structural contributions: the combined dependence on the presence of the sialic acid residue, fucosyl residue, 'Tys' and Ca(II) for optimal binding is clear. Error bars, s.e.m.
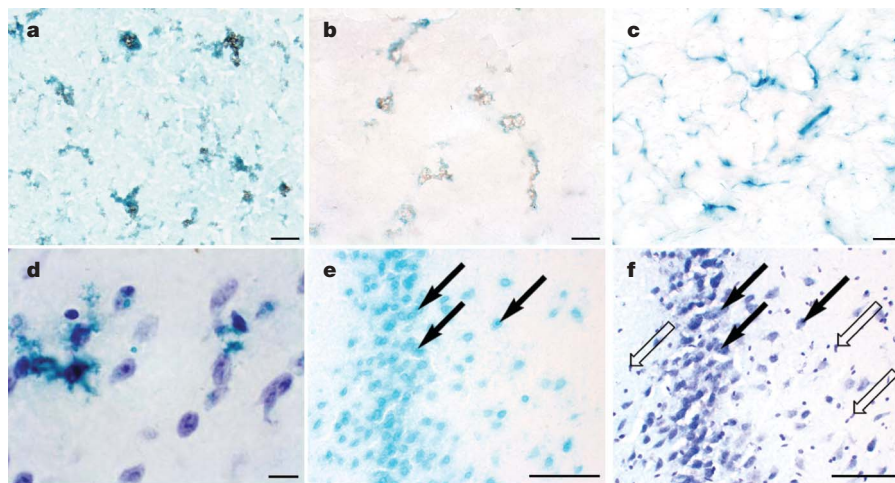
Figure 4 | Use of mimics PSGL-LacZ, PSGL*-LacZ and GlcNAc-LacZ as probes. a, X-Gal-stained tissue of chronically inflamed rat cortex in the presence of PSGL-LacZ. **b**, X-Gal-stained tissue of acutely inflamed rat cortex with PSGL-LacZ. **c**, X-Gal-stained vasculature following direct carotid injection. **d**, Detection of cerebral malaria pathology with PSGL*-LacZ in C57BL/6 mice 6 d after infection. (cresyl-counterstained). **e**, GlcNAc-modified probe GlcNAc-LacZ selectively binds to nuclei of a subpopulation of brain cells: selective X-Gal-staining of neuronal nuclei (filled arrows) reveals hitherto unknown GlcNAc-modified-protein-binding activity absent in glial cell nuclei (**f**) (open arrows, cresyl-stained only). Scale bars **a**–**c**, **e**, 50 μm; **d**, 10 μm.

and PSGL*-LacZ were used to assay a *Plasmodium*-infection malarial disease model; after just 6 days cerebral disease was readily detected (Fig. 4).

Finally, GlcNAc-modified LacZ protein probes were created. The role and extent of GlcNAc-modified proteins is a rapidly developing field[14], within which novel probes will delineate function and partners; 25 proteins may contain this PTM in mammalian brain alone[27]. No probes with site-selective GlcNAc modification have yet been used to map potential sugar-specific binding partners. Using GlcNAc-LacZ, specific binding was observed to the mouse innate immunity protein DC-SIGN-R2 (ref. 28; see Supplementary Information). When used as a probe of potential binding partners for brain proteins with GlcNAc modification, GlcNAc-LacZ selectively bound (Fig. 4e, f; Supplementary Fig. S7) to nuclei of a neuron subpopulation, showing no binding to glial cells. This hitherto unknown GlcNAc-dependent cell type selectivity suggests that neurons may display selective GlcNAc-binding proteins. This is intriguing in the light of previously suggested[29] GlcNAc-mediated nuclear shuttling in *Aplysia* neurons, and proposed roles for GlcNAc in the neuronal nuclear localization of Alzheimer's-associated protein Tau and its consequent implication in neurodegenerative disease[30]. The mechanism and identity of participants in this newly elucidated GlcNAc-specific process are being determined.

These results show that orthogonal protein modification strategies can create synthetic yet functional protein mimics to precisely unravel contributions from complex multi-component binding domains involving different PTMs. The compatibility, and orthogonality, of a metal-catalysed cycloaddition chemical modification with methods for chemical protein glycosylation[5] has expanded the diversity of methods for precise protein modification. These methods are mild enough to create proteins that retain original (for example, enzymatic reporter) function and are bestowed with additional (for example, protein-binding) function. This has allowed the chemical synthesis of reconstituted mimics of naturally-occurring proteins, such as PSGL-1, containing multiple PTMs, and the discovery of new PTM-dependent interactions.

METHODS

PSGL*-LacZ synthesis. A solution of SSβG-Aha43-Cys439 (0.4 mg ml^{-1}; 4 ml) in phosphate buffer (50 mM, pH 8.0) and reagent Tys-MTS (10 mg ml^{-1}, 30 μl) was stirred at RT for 1 h, and dialysed (pH 8.5, 50 mM phosphate; MWCO 12,000–14,000, 3 × 2 l). Resulting singly-'Tys'-modified protein was concentrated (~2 mg ml^{-1}), and reagent sLe^x-alkyne (2 mg, 2 mmol) added.

Copper(I)bromide (99.999%, 10 mg ml^{-1} acetonitrile, 500 μl) and *tris*(1-(O-ethylcarboxymethyl)-1*H*-1,2,3-triazol-4-ylmethyl)amine (50 mg ml^{-1} acetonitrile, 660 μl) were mixed. Resulting solution was added to protein solution (3 × 70 μl, 5 min intervals) with gentle agitation, briefly centrifuged (13,000*g*), supernatant concentrated (Vivaspin, 10,000 MWCO, Vivascience, cat no VS0101), washed (3 × 500 μl deionised water), concentrated (100 μl) and characterized.

In vivo probing. Wistar mice were anaesthetized (2.5% isoflurane in 70%N$_2$O:30%O$_2$). 1 ng of recombinant rat IL-1β (NIBSC) in 1 μl saline was injected stereotaxically, 1 mm anterior, 3 mm lateral to Bregma, 4 mm depth into left striatum. Animals recovered for 4.0–4.5 h, were re-anesthetized as above, probe solution (for example, PSGL*-LacZ, 5–10 mg ml^{-1}, 80 μl) was administered into the right carotid artery and allowed to circulate for 10 min before the animal was killed. Rats were deeply anaesthetized with sodium pentobarbitone. Trans-cardiac perfusions were performed using heparinized saline (10 ml) and formalin (1 ml). Brain tissue was removed and post-fixed for 4 h in formalin at 4 °C and cryoprotected (30w/v sucrose solution, 8 h, 4 °C) before embedding in TissueTek (Bayer). 10 μm slices were placed on gelatin-coated slides and dried.

Section probing. Tissue was removed, embedded in TissueTek and frozen over *n*-pentane. 10 μm slices of unfixed frozen brain tissue on 3-aminopropyl-triethoxysilane-treated microscope slides were immersed in neutral buffered formalin (pH 7, 10 min) and overlaid with PSGL*-LacZ (100 μl, 50 mg ml^{-1}, 2 mM CaCl$_2$-containing PBS), incubated (37 °C, 2.5 h), washed (PBS-Ca buffer) and fixed (formalin, 10 min) and washed (PBS).

Probe visualization. Slides were incubated (18 h, 37 °C) with X-Gal (1 mg ml^{-1} in 5 mM K$_2$Fe(CN)$_4$, 5 mM K$_3$Fe(CN)$_4$ containing PBS, from 40 mg ml^{-1} DMSO stock) then formalin fixed (10 min), washed (PBS then water) and images taken after dehydration and DePeX-fixation (Histological Equipment).

1. Wold, F. *In vivo* chemical modification of proteins (post-translational modification). *Annu. Rev. Biochem.* **50**, 783–814 (1981).
2. Davis, B. G. Mimicking posttranslational modifications of proteins. *Science* **303**, 480–482 (2004).
3. Dwek, R. A. Glycobiology: Toward understanding the function of sugars. *Chem. Rev.* **96**, 683–720 (1996).
4. Davis, B. G. Synthesis of glycoproteins. *Chem. Rev.* **102**, 579–601 (2002).
5. Davis, B. G., Lloyd, R. C. & Jones, J. B. Controlled site-selective glycosylation of proteins by a combined site-directed mutagenesis and chemical modification approach. *J. Org. Chem.* **63**, 9614–9615 (1998).
6. Kiick, K. L., Saxon, E., Tirrell, D. A. & Bertozzi, C. R. Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *Proc. Natl Acad. Sci. USA* **99**, 19–24 (2002).
7. van Hest, J. C. M., Kiick, K. L. & Tirrell, D. A. Efficient incorporation of unsaturated methionine analogues into proteins *in vivo*. *J. Am. Chem. Soc.* **122**, 1282–1288 (2000).

8. Fodje, M. N. & Al-Karadaghi, S. Occurrence, conformational features and amino acid propensities for the pi-helix. *Protein Eng.* **15**, 353–358 (2002).

9. Cubellis, M. V., Rozzo, C., Montecucchi, P. & Rossi, M., *Gene* **94**, 89–94 (1990).

10. Easton, R. L., Patankar, M. S., Clark, G. F., Morris, H. R. & Dell, A. Pregnancy-associated changes in the glycosylation of Tamm-Horsfall glycoprotein. *J. Biol. Chem.* **275**, 21928–21938 (2000).

11. Campbell, S. J., Carlotti, F., Hall, P. A., Clark, A. J. & Wolf, C. R. Regulation of the CYP1A1 promoter in transgenic mice: An exquisitely sensitive on-off system for cell specific gene regulation. *J. Cell Sci.* **109**, 2619–2625 (1996).

12. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise Huisgen cycloaddition process: Copper-catalyzed regioselective ligation of azides and terminal alkynes. *Angew. Chem. Int. Edn* **41**, 2596–2599 (2002).

13. Tornøe, C. W., Christensen, C. & Meldal, M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* **67**, 3057–3064 (2002).

14. Vosseller, K., Sakabe, K., Wells, L. & Hart, G. W. Diverse regulation of protein function by O-GlcNAc: a nuclear and cytoplasmic carbohydrate post-translational modification. *Curr. Opin. Chem. Biol.* **6**, 851–857 (2002).

15. Somers, W. S., Tang, J., Shaw, G. D. & Camphausen, R. T. Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P-and E-selectin bound to SLe(X) and PSGL-1. *Cell* **103**, 467–479 (2000).

16. Kansas, G. S. Selectins and their ligands: Current concepts and controversies. *Blood* **88**, 3259 (1996).

17. Leppänen, A., White, S. P., Helin, J., McEver, R. P. & Cummings, R. D. Binding of glycosulfopeptides to P-selectin requires stereospecific contributions of individual tyrosine sulfate and sugar residues. *J. Biol. Chem.* **275**, 39569–39578 (2000).

18. Nelson, R. M., Dolich, S., Aruffo, A., Cecconi, O. & Bevilacqua, M. P. Higher-affinity oligosaccharide ligands for E-selectin. *J. Clin. Investig.* **91**, 1157–1166 (1993).

19. Barkalow, F. J., Goodman, M. J., Gerritsen, M. E. & Mayadas, T. N. Brain endothelium lack one of two pathways of P-selectin-mediated neutrophil adhesion. *Blood* **88**, 4585–4593 (1996).

20. Sibson, N. R. *et al.* MRI detection of early endothelial activation in brain inflammation. *Magn. Reson. Med.* **51**, 248–252 (2004).

21. Ferrari, C. C. *et al.* Reversible demyelination, blood-brain barrier breakdown, and pronounced neutrophil recruitment induced by chronic IL-1 expression in the brain. *Am. J. Pathol.* **165**, 1827–1837 (2004).

22. Gotsch, U., Jager, U., Dominis, M. & Vestweber, D. Expression of P-selectin on endothelial cells is upregulated by LPS and TNF-alpha *in vivo*. *Cell Adhes. Commun.* **2**, 7–14 (1994).

23. Bernardes-Silva, M., Anthony, D. C., Issekutz, A. C. & Perry, V. H. Recruitment of neutrophils across the blood-brain barrier: The role of E- and P-selectins. *J. Cereb. Blood Flow Metab.* **21**, 1115–1124 (2001).

24. Schofield, L. & Grau, G. E. Immunological processes in malaria pathogenesis. *Nature Rev. Immunol.* **5**, 722–735 (2005).

25. Grau, G. E. *et al.* Platelet accumulation in brain microvessels in fatal pediatric cerebral malaria. *J. Infect. Dis.* **187**, 461–466 (2003).

26. Sun, G. *et al.* Inhibition of platelet adherence to brain microvasculature protects against severe *Plasmodium berghei* malaria. *Infect. Immun.* **71**, 6553–6561 (2003).

27. Khidekel, N., Ficarro, S. B., Peters, E. C. & Hsieh-Wilson, L. C. Exploring the O-GlcNAc proteome: Direct identification of O-GlcNAc-modified proteins from the brain. *Proc. Natl Acad. Sci. USA* **101**, 13132–13137 (2004).

28. Powlesland, A. S. *et al.* Widely divergent biochemical properties of the complete set of mouse DC-SIGN-related proteins. *J. Biol. Chem.* **281**, 20440–20449 (2006).

29. Elliot, S. P., Schmied, R., Gabel, C. A. & Ambron, R. T. An 83-kDa O-GlcNAcglycoprotein is found in the axoplasm and nucleus of *Aplysia* neurons. *J. Neurosci.* **13**, 2424–2429 (1993).

30. Lefebvre, T. *et al.* Does O-GlcNAc play a role in neurodegenerative diseases? *Expert Rev. Proteomics* **2**, 265–275 (2005).

naturejobs

Women in computer science are like "canaries in a coal mine" according to Lenore Blum, a computer scientist at Carnegie Mellon University in Pittsburgh, Pennsylvania. Her remarks, made in a talk at Harvard University and reported in *The New York Times* (17 April), were highlighting the fact that the number of women graduating in computer science in the United States is falling. Figures from the National Science Foundation say that 38% of computer-science graduates were women in 1985, but in 2003 women accounted for only 28%. Blum believes that this tailing off is the beginning of something more serious — she fears that the factors encouraging women to quit the field will soon be leading their male counterparts away too.

Some of those factors are actually misconceptions about the state of the job market. It is commonly thought that the dotcom bust and the trend for outsourcing information-technology jobs to cheaper labour markets in countries such as India, are restricting opportunities. But, in fact, government reports and industry experts expect demand for computer scientists to increase. Nevertheless, the decline in the number of women active in the field is worrying — and seems to stretch beyond academia. A Correspondence to *Nature* last year pointed out that none of the 41 authors of *Towards 2020 Science* — a report from Microsoft that examined how to integrate computing into the sciences — was female (M. E. Pollack *Nature* **441,** 25; 2006).

Blum suggests that switching the course emphasis away from programming efficiency to problem-solving and software design might resolve these issues. That approach has certainly worked at Carnegie Mellon, where the revamped curriculum has seen the intake of women students for computer science rise from 8% to nearly 40%. Another idea might be to emphasize that computer skills are essential to many disciplines, such as materials science, astronomy and bioinformatics. Having a hearty background in IT can help both men and women excel, not just in computer science, but across all scientific disciplines.

Paul Smaglik, *Naturejobs* **editor**

BIOSCIENCE IN THE SUN

Flush with funds, Florida is enjoying a state-backed bioscience boom, says **Gene Russo**.

It's hard to imagine a more affable tour guide than Bart Chernow. A newcomer to the University of Miami's growing medical school, the vice-president for special programmes and resource strategy seems genuinely smitten with the place. Showing off the school's lush courtyards and new facilities, including a 35,000-square-metre clinical centre that opened in January and a 16,700-square-metre biomedical research institute due to open next year, he stops to pick up a few bits of rubbish from an otherwise spotless pavement. He's both booster and caretaker.

With all the gusto of a storyteller, Chernow boasts about the medical school's promise and lauds its leadership (former US health and human services secretary Donna Shalala is the university president) and philanthropic success. But when it comes to discussing early plans for a new biotech park next to the medical school, his primary charge in his new position, Chernow is more cautious. He knows that success is by no means guaranteed. Biotechnology companies rarely make profits, and other locales in Florida, from Tampa to Palm Beach, already have major biotech park efforts under way. "What we would like to do is understate and outperform," he says.

High hopes under the palms

Although wary of tight US National Institutes of Health (NIH) budgets and the high-risk nature of banking on biotech firms, researchers in Florida are brimming with optimism. Buoyed by an influx of state funds doled out by former Governor Jeb Bush, the bioscience field in Florida is booming. Alongside Disney and orange-tree groves, the state government envisions a new industry bringing tens of thousands of square metres of lab space, well-paid jobs and, eventually, disease treatments and royalties

Already there are major signs of success. The state has convinced three highly regarded California-based institutes — Scripps, Torrey Pines and Burnham — to

Pascal Goldschmidt has lured high-profile researchers to Florida.

set up campuses in Florida. Scripps Florida is operating in temporary buildings with 230 staff while its state-of-the art facilities are being constructed, due to open in early 2009. Eventually the Scripps campus will include three buildings with 32,000 square metres of open lab space at a total cost of $186 million, paid for by Palm Beach County. The state has added about $310 million in start-up funds over ten years. Burnham plans to build up a 250-person operation over the next seven to ten years, mostly using newly recruited scientists. With more than $80 million of funding from state and local governments, Torrey Pines plans to open its 9,300-square-metre research centre in 2009 in Port St Lucie, with a projected 189 new jobs over the next ten years.

Scripps president Richard Lerner encouraged Torrey Pines founder and president Richard Houghten to pursue a Florida branch as well. "I'll have Jeb call you," said Lerner over dinner one night. A few weeks later the governor was on the phone. Early last year, Houghten found himself scouting sites in Florida. "I had a San Diego-centric view of the world," he says. "We flew out to Florida and loved it."

Florida universities have also had a boost, largely from state and local sources. The University of Florida has a new state-sponsored Center of Excellence for Regenerative Health Biotechnology (CERHB), a 26,000-square-metre, $85-million cancer-genetics building, an Emerging Pathogens Institute — which recently landed $55 million in state funding and is building a 9,300-square-metre lab space — and a burgeoning biomedical-engineering department that's seen $120 million of investment over the past four and a half years. The University of South Florida will house another state-funded centre of excellence, this one focused on biomolecular identification and targeted therapeutics. The university is collaborating with SRI International, an independent R&D organization, to build a new marine-technology research facility in St Petersburg. SRI expects to grow from 40 staff to 200 in the next ten years; the
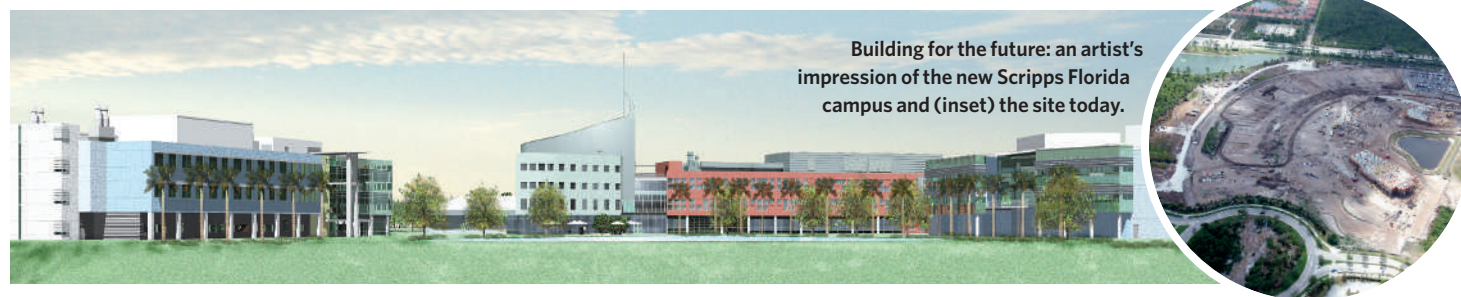
Building for the future: an artist's impression of the new Scripps Florida campus and (inset) the site today.

state gave $20 million and local government provided the site for a 2,700-square-metre facility.

The calibre of the recruits coming to Florida campuses is impressive. Miami's medical-school dean, Pascal Goldschmidt, convinced his former Duke University colleague Margaret Pericak-Vance, an expert in the genetics of neurodegenerative diseases, to come to Miami — bringing 40 of her team, including 12 faculty members. Goldschmidt also lured Marc Lippman, the chair of internal medicine at the University of Michigan who will become Miami's chair of medicine. Scripps recruits include chair of cancer biology John Cleveland, an 18-year veteran of St Jude Children's Research Hospital in Memphis, Tennessee, and vice-president of scientific operations Harry Orf, who was director of the molecular-biology laboratories at Massachusetts General Hospital, Boston, for 21 years.

Sunshine state's secret

Why Florida? In addition to the influx of state funding, the state has low corporate taxes and no personal income tax. Also, the environment and demographics are research-friendly. The numerous elderly citizens represent both potential study subjects for neurodegenerative disease, and sources of philanthropic support. For those studying the genetics of disease, the mix of races — from West Indian to Cuban to Latin American — matches the country's evolving demographics. Florida's warm, wet climate and influx of immigrants make it the perfect place to study emerging infectious disease. The acting director of the Institute of Emerging Pathogens, Richard Moyer, suggests that its strength lies in bringing together the resources of large, extensive departments to study the movement of pathogens in food, plants and animals — resources not found at Harvard or Johns Hopkins. And then there are the warm climate and great beaches.

The sunshine state seems to have everything. But it can't escape the major barrier facing bioscience researchers elsewhere: tight NIH budgets. Orf sees an upside to expanding in the era of shrinking federal funding: start-up packages will help fill the holes left by absent NIH funding, at least for a few years, although researchers will have to become self-sustaining. Institutes are already looking to philanthropy within the state. "People are beginning to realize that the real growth is in private philanthropy," says Moyer.

Given the emphasis that many institutions put on translational research, corporate partnerships will be key. In November, Scripps announced a five-year, $100-million collaboration with Pfizer on diseases including cancer, diabetes and mental illness. In exchange for funds, Pfizer will have the right to license therapies. Translational research is a big draw. "I have to turn people away," says Pat Griffin, head of the Scripps Translational Research Institute. "It's a rare

opportunity to learn translational research that isn't in industry."

Late last year the Moffit Cancer Center at the University of South Florida signed a collaborative agreement with Merck to build, with the help of $35 million in state and local funds, a for-profit institute called M2Gen. At the CERHB, the mission is to move bench research in the clinic by making drugs for phase I and II clinical trials. Clients include start-ups and independent researchers. "Not many universities have similar operations," says Richard Snyder, director of the University of Florida's biotherapeutic programmes at the centre. The emphasis on translational research, says Snyder, should improve the chances for NIH funds, as the agency's roadmap plan calls for such projects.

Can such ventures help transform Tampa, Palm Beach or Miami into the next big biotech cluster? Economist and consultant Joseph Cortright has his doubts. In analysing biotech successes and failures for the Brookings Institution in Washington DC two years ago, Cortright found that the industry, which rarely produces a profitable company, is highly concentrated in a few places, including San Francisco, Boston and San Diego, where institutes and universities are well established. "It's expensive, time-consuming and very risky," he says of the industry. Some companies, he notes, are saying they might set up shop in Florida, but, seeing the tens of millions garnered by Burnham, Torrey Pines and Scripps, they have their hands out as well. "The question is, when does Florida stop giving out money?" Cortright says.

Bright prospects for bright students

Also of concern is the supply of support staff. Recruitment for top positions remains strong. "In the first couple of years we worried no one would come to party," says Bill Ditto, chair of the University of Florida's growing biomedical-engineering department. "Now I'm crying with the quality of talent we have to turn away." But Orf says that support positions, such as technicians, are lacking. Scripps and others have started outreach efforts to foster an understanding of biotechnology in young Floridians, hoping to encourage a generation that is suited to the emerging bioscience and biotech sectors. At Snyder's centre, they are developing courses for University of Florida students focused on topics ranging from understanding regulatory compliance to assay development to working in science teams.

With ample money, top-flight talent and state-of the art facilities, Florida aims to be a major bioscience player for years to come. Whether the time and money invested will yield products, therapies and state revenue is an open question. "I'm cautiously optimistic," Chernow repeats with a smile. ∎

Gene Russo is assistant editor for *Naturejobs.*

Margaret Pericak-Vance (top) and Bill Ditto think Florida has potential.

MOVERS

Frank Gannon, director-general, Science Foundation Ireland, Dublin, Ireland

1994–present: Executive director, European Molecular Biology Organization, Heidelberg, Germany
1994–present: Senior scientist, European Molecular Biology Laboratory, Heidelberg, Germany
1990–94: Associate professor, Department of Microbiology, University College, Galway, Ireland

Unlike many biochemists, Frank Gannon didn't play with chemistry sets as a kid — he was too busy seeking his next intellectual challenge. Interested in law and languages, the University of Galway undergraduate decided to specialize in biochemistry only when a friend mentioned the unfamiliar topic. "When I don't know about something, that's where I want to go," says Gannon.

He continued with a PhD at Leicester, UK, in enzymology, because proteins were in vogue at the time. But for his postdoc, Gannon wanted a change of scenery as well as science. At the University of Wisconsin at Madison, he worked on oestrogen receptors, which were poorly understood at that time. Although offered a job at the Michigan Cancer Foundation, he felt he wasn't yet ready to run his own lab. His journey back to Europe led him to Pierre Chambon's lab in Strasbourg, France — one of Europe's top laboratories.

Eager to prove his own scientific worth, Gannon took on the challenge of creating a successful research programme at the University of Galway without money or resources. "Career-wise, this was not a sensible move," says Gannon. He wrote a manifesto to guide his career goals: create a world-class lab, bring biotechnology into the education system, and influence Irish industry — all of which he had accomplished by the time he left a dozen years later to lead the European Molecular Biology Organization (EMBO).

To help EMBO reach its full potential, Gannon worked to increase communications across countries and create initiatives to promote young-scientist awards as well as researchers in developing countries. These efforts helped spur Europe's restructuring of the research enterprise, including the creation of the European Research Council.

Now Gannon has made yet another career move. Returning to Ireland, he becomes director-general of Science Foundation Ireland (SFI) on 1 July. Patrick Fottrell, chair of the SFI board, believes Gannon's leadership and international standing will prove pivotal to Ireland's efforts to recruit and retain world-class scientists, as well as in attracting high-tech corporate research and development. To do so, he plans to maintain scientific quality at Ireland's universities, even as they expand postgraduate offerings.

Gannon gauges his career success by what he calls "the Rip Van Winkle test". When you move on from a job, is the organization as untouched by your presence as if you'd slept through it, or have you improved it?
Virginia Gewin

NETWORKS & SOCIETY

Postdoc redefined

"There is something very wrong here," Elias Zerhouni, director of the US National Institutes of Health (NIH), told the annual conference of the National Postdoctoral Association (NPA) in Berkeley, California, on 30 March. He was lamenting that most principal investigators are 40 or older when they get their first NIH grant. It's a problem the agency is trying to address with the K99 grant, which started last year.

Intended to bridge the gap between postdoc and independent investigator, K99s fund individual postdocs for two years and then travel with them to an independent appointment for three years. In November 2006, the NIH awarded 58 K99s; that should go up to 170 this year, a big commitment given its flat budget.

But the very grant designed to help young scientists gain a foothold shuts many of them out. To be eligible for a K99, researchers must have received their PhDs within the past five years. "Second postdocs are becoming a common thing, and as the length from the defence date grows, you're losing opportunities," says Scott Nowak, who is on his second postdoc at the Memorial-Sloan-Kettering Cancer Center in New York.

It may also fail to serve another crucial role: attracting promising foreign postdocs to first-time faculty positions in the United States. Non-US scientists are eligible for K99s, but the lure of alternative career paths and attractive options abroad is diverting foreign talent, according to Rajika Bhandari of the Institute of International Education, a non-profit research group. A 2007 report by the Lumina Foundation for Education estimates that by 2025, the United States will be short of 16 million workers holding an associates or higher degree.

Zerhouni defended K99 as a "defining experiment" in how best to prepare the next generation of scientists. He said it was not intended to make broad changes.

Pushed by the NPA, the NIH and the US National Science Foundation recently settled on a joint definition of postdoctoral scholar, stating that postdocs are in a period of "mentored advanced training" designed to move scientists along their "chosen career path". That establishes a research sponsors' duty as adviser and implicitly acknowledges that not all careers are academic, says Diane Klotz, chair of the NPA's board of directors. But Zerhouni, noting that postdocs fill highly specialized roles within labs, urged the NPA to do more.
Monya Baker

POSTDOC JOURNAL

All ears

I recently returned from a conference of about 500 participants where almost everyone works on corn (maize). The experience got me interested again in a field that's no longer my primary focus.

The conference was particularly engaging because my principal investigator assisted with the introductions. Also, my second child's impending arrival made for an easy conversation starter. I presented a poster with brand new data — on a mutant I'm characterizing — that I was quite excited about. But I also included on the poster an older photograph of corn ears with sterility defects that I thought the community at large might be interested in. I was amazed that the latter generated far more excitement than the former for most viewers.

Unlike earlier years, I received a lot of suggestions and lots of encouragement. It was great to have so much attention for my project, to have so many people wanting to hear about it: this was not the case on many previous occasions. Also encouraging, I met a researcher whose work intersected with mine. We talked extensively about his results and my previous work, which got him very excited. It seems that there is a budding interest in this field now that I've left it. Did I miss the wave or, maybe, did I start it? Either way, it's satisfying to see my work appreciated, even though it took some time.
Moira Sheehan is a postdoc in plant breeding and genetics at Cornell University.

MOVERS

Frank Gannon, director-general, Science Foundation Ireland, Dublin, Ireland

1994–present: Executive director, European Molecular Biology Organization, Heidelberg, Germany
1994–present: Senior scientist, European Molecular Biology Laboratory, Heidelberg, Germany
1990–94: Associate professor, Department of Microbiology, University College, Galway, Ireland

Unlike many biochemists, Frank Gannon didn't play with chemistry sets as a kid — he was too busy seeking his next intellectual challenge. Interested in law and languages, the University of Galway undergraduate decided to specialize in biochemistry only when a friend mentioned the unfamiliar topic. "When I don't know about something, that's where I want to go," says Gannon.

He continued with a PhD at Leicester, UK, in enzymology, because proteins were in vogue at the time. But for his postdoc, Gannon wanted a change of scenery as well as science. At the University of Wisconsin at Madison, he worked on oestrogen receptors, which were poorly understood at that time. Although offered a job at the Michigan Cancer Foundation, he felt he wasn't yet ready to run his own lab. His journey back to Europe led him to Pierre Chambon's lab in Strasbourg, France — one of Europe's top laboratories.

Eager to prove his own scientific worth, Gannon took on the challenge of creating a successful research programme at the University of Galway without money or resources. "Career-wise, this was not a sensible move," says Gannon. He wrote a manifesto to guide his career goals: create a world-class lab, bring biotechnology into the education system, and influence Irish industry — all of which he had accomplished by the time he left a dozen years later to lead the European Molecular Biology Organization (EMBO).

To help EMBO reach its full potential, Gannon worked to increase communications across countries and create initiatives to promote young-scientist awards as well as researchers in developing countries. These efforts helped spur Europe's restructuring of the research enterprise, including the creation of the European Research Council.

Now Gannon has made yet another career move. Returning to Ireland, he becomes director-general of Science Foundation Ireland (SFI) on 1 July. Patrick Fottrell, chair of the SFI board, believes Gannon's leadership and international standing will prove pivotal to Ireland's efforts to recruit and retain world-class scientists, as well as in attracting high-tech corporate research and development. To do so, he plans to maintain scientific quality at Ireland's universities, even as they expand postgraduate offerings.

Gannon gauges his career success by what he calls "the Rip Van Winkle test". When you move on from a job, is the organization as untouched by your presence as if you'd slept through it, or have you improved it?
Virginia Gewin

NETWORKS & SOCIETY

Postdoc redefined

"There is something very wrong here," Elias Zerhouni, director of the US National Institutes of Health (NIH), told the annual conference of the National Postdoctoral Association (NPA) in Berkeley, California, on 30 March. He was lamenting that most principal investigators are 40 or older when they get their first NIH grant. It's a problem the agency is trying to address with the K99 grant, which started last year.

Intended to bridge the gap between postdoc and independent investigator, K99s fund individual postdocs for two years and then travel with them to an independent appointment for three years. In November 2006, the NIH awarded 58 K99s; that should go up to 170 this year, a big commitment given its flat budget.

But the very grant designed to help young scientists gain a foothold shuts many of them out. To be eligible for a K99, researchers must have received their PhDs within the past five years. "Second postdocs are becoming a common thing, and as the length from the defence date grows, you're losing opportunities," says Scott Nowak, who is on his second postdoc at the Memorial-Sloan-Kettering Cancer Center in New York.

It may also fail to serve another crucial role: attracting promising foreign postdocs to first-time faculty positions in the United States. Non-US scientists are eligible for K99s, but the lure of alternative career paths and attractive options abroad is diverting foreign talent, according to Rajika Bhandari of the Institute of International Education, a non-profit research group. A 2007 report by the Lumina Foundation for Education estimates that by 2025, the United States will be short of 16 million workers holding an associates or higher degree.

Zerhouni defended K99 as a "defining experiment" in how best to prepare the next generation of scientists. He said it was not intended to make broad changes.

Pushed by the NPA, the NIH and the US National Science Foundation recently settled on a joint definition of postdoctoral scholar, stating that postdocs are in a period of "mentored advanced training" designed to move scientists along their "chosen career path". That establishes a research sponsors' duty as adviser and implicitly acknowledges that not all careers are academic, says Diane Klotz, chair of the NPA's board of directors. But Zerhouni, noting that postdocs fill highly specialized roles within labs, urged the NPA to do more.
Monya Baker

POSTDOC JOURNAL

All ears

I recently returned from a conference of about 500 participants where almost everyone works on corn (maize). The experience got me interested again in a field that's no longer my primary focus.

The conference was particularly engaging because my principal investigator assisted with the introductions. Also, my second child's impending arrival made for an easy conversation starter. I presented a poster with brand new data — on a mutant I'm characterizing — that I was quite excited about. But I also included on the poster an older photograph of corn ears with sterility defects that I thought the community at large might be interested in. I was amazed that the latter generated far more excitement than the former for most viewers.

Unlike earlier years, I received a lot of suggestions and lots of encouragement. It was great to have so much attention for my project, to have so many people wanting to hear about it: this was not the case on many previous occasions. Also encouraging, I met a researcher whose work intersected with mine. We talked extensively about his results and my previous work, which got him very excited. It seems that there is a budding interest in this field now that I've left it. Did I miss the wave or, maybe, did I start it? Either way, it's satisfying to see my work appreciated, even though it took some time.
Moira Sheehan is a postdoc in plant breeding and genetics at Cornell University.

MOVERS

Frank Gannon, director-general, Science Foundation Ireland, Dublin, Ireland

1994–present: Executive director, European Molecular Biology Organization, Heidelberg, Germany
1994–present: Senior scientist, European Molecular Biology Laboratory, Heidelberg, Germany
1990–94: Associate professor, Department of Microbiology, University College, Galway, Ireland

Unlike many biochemists, Frank Gannon didn't play with chemistry sets as a kid — he was too busy seeking his next intellectual challenge. Interested in law and languages, the University of Galway undergraduate decided to specialize in biochemistry only when a friend mentioned the unfamiliar topic. "When I don't know about something, that's where I want to go," says Gannon.

He continued with a PhD at Leicester, UK, in enzymology, because proteins were in vogue at the time. But for his postdoc, Gannon wanted a change of scenery as well as science. At the University of Wisconsin at Madison, he worked on oestrogen receptors, which were poorly understood at that time. Although offered a job at the Michigan Cancer Foundation, he felt he wasn't yet ready to run his own lab. His journey back to Europe led him to Pierre Chambon's lab in Strasbourg, France — one of Europe's top laboratories.

Eager to prove his own scientific worth, Gannon took on the challenge of creating a successful research programme at the University of Galway without money or resources. "Career-wise, this was not a sensible move," says Gannon. He wrote a manifesto to guide his career goals: create a world-class lab, bring biotechnology into the education system, and influence Irish industry — all of which he had accomplished by the time he left a dozen years later to lead the European Molecular Biology Organization (EMBO).

To help EMBO reach its full potential, Gannon worked to increase communications across countries and create initiatives to promote young-scientist awards as well as researchers in developing countries. These efforts helped spur Europe's restructuring of the research enterprise, including the creation of the European Research Council.

Now Gannon has made yet another career move. Returning to Ireland, he becomes director-general of Science Foundation Ireland (SFI) on 1 July. Patrick Fottrell, chair of the SFI board, believes Gannon's leadership and international standing will prove pivotal to Ireland's efforts to recruit and retain world-class scientists, as well as in attracting high-tech corporate research and development. To do so, he plans to maintain scientific quality at Ireland's universities, even as they expand postgraduate offerings.

Gannon gauges his career success by what he calls "the Rip Van Winkle test". When you move on from a job, is the organization as untouched by your presence as if you'd slept through it, or have you improved it?
Virginia Gewin

NETWORKS & SOCIETY

Postdoc redefined

"There is something very wrong here," Elias Zerhouni, director of the US National Institutes of Health (NIH), told the annual conference of the National Postdoctoral Association (NPA) in Berkeley, California, on 30 March. He was lamenting that most principal investigators are 40 or older when they get their first NIH grant. It's a problem the agency is trying to address with the K99 grant, which started last year.

Intended to bridge the gap between postdoc and independent investigator, K99s fund individual postdocs for two years and then travel with them to an independent appointment for three years. In November 2006, the NIH awarded 58 K99s; that should go up to 170 this year, a big commitment given its flat budget.

But the very grant designed to help young scientists gain a foothold shuts many of them out. To be eligible for a K99, researchers must have received their PhDs within the past five years. "Second postdocs are becoming a common thing, and as the length from the defence date grows, you're losing opportunities," says Scott Nowak, who is on his second postdoc at the Memorial-Sloan-Kettering Cancer Center in New York.

It may also fail to serve another crucial role: attracting promising foreign postdocs to first-time faculty positions in the United States. Non-US scientists are eligible for K99s, but the lure of alternative career paths and attractive options abroad is diverting foreign talent, according to Rajika Bhandari of the Institute of International Education, a non-profit research group. A 2007 report by the Lumina Foundation for Education estimates that by 2025, the United States will be short of 16 million workers holding an associates or higher degree.

Zerhouni defended K99 as a "defining experiment" in how best to prepare the next generation of scientists. He said it was not intended to make broad changes.

Pushed by the NPA, the NIH and the US National Science Foundation recently settled on a joint definition of postdoctoral scholar, stating that postdocs are in a period of "mentored advanced training" designed to move scientists along their "chosen career path". That establishes a research sponsors' duty as adviser and implicitly acknowledges that not all careers are academic, says Diane Klotz, chair of the NPA's board of directors. But Zerhouni, noting that postdocs fill highly specialized roles within labs, urged the NPA to do more.
Monya Baker

POSTDOC JOURNAL

All ears

I recently returned from a conference of about 500 participants where almost everyone works on corn (maize). The experience got me interested again in a field that's no longer my primary focus.

The conference was particularly engaging because my principal investigator assisted with the introductions. Also, my second child's impending arrival made for an easy conversation starter. I presented a poster with brand new data — on a mutant I'm characterizing — that I was quite excited about. But I also included on the poster an older photograph of corn ears with sterility defects that I thought the community at large might be interested in. I was amazed that the latter generated far more excitement than the former for most viewers.

Unlike earlier years, I received a lot of suggestions and lots of encouragement. It was great to have so much attention for my project, to have so many people wanting to hear about it: this was not the case on many previous occasions. Also encouraging, I met a researcher whose work intersected with mine. We talked extensively about his results and my previous work, which got him very excited. It seems that there is a budding interest in this field now that I've left it. Did I miss the wave or, maybe, did I start it? Either way, it's satisfying to see my work appreciated, even though it took some time.
Moira Sheehan is a postdoc in plant breeding and genetics at Cornell University.

The inside track from academia and industry

Talents and technologies

Scientists need to concentrate on ways to capitalize on their intellectual property.

Michael Alvarez

Few would argue with the idea that the fundamental role of universities is education; since 1088, the University of Bologna in Italy (the oldest such institution) has been in the business of educating students and awarding degrees. But in today's world the role of universities has expanded, and their relationship to global labour and industrial markets is more immediate and direct.

In some cases, companies search universities far and wide for specific technologies, whereas in others they are interested in acquiring the skills of well-trained scientists through employment — grabbing the goose, not the golden egg. As demand for these valuable commodities has increased, offices have evolved that support the licensing of technologies and the career development of talented scientists.

Technology-transfer offices are common at research universities worldwide, but faculty and trainee knowledge of their location, function, scope and scale sometimes lags behind. This is unfortunate, because these offices can introduce scientists to opportunities for collaboration and alternative career pathways.

Even scientists without plans to patent and commercialize their research or collaborate with companies can benefit from exposure to such offices.

> **"Technology-transfer offices can introduce scientists to alternative career pathways."**

Although many scientists may work with industrial interdependencies in mind, a deeper understanding of both the academic and industrial sides of these relationships can highlight the value of supporting such interactions. These may not be beneficial to a particular research project, but such visibility can help faculty members provide effective guidance for their trainees.

For instance, how does a researcher know whether or at what stage a particular discovery satisfies the criteria for being both patentable and patent worthy? And, if a given university has decided those conditions have not been met, what other options are available to the scientist who wants to patent or commercialize a process or product based on their technology? The field of intellectual-property (IP) management is complex, and researchers should consult experts, sooner rather than later, to help determine the best options for a given innovation or discovery.

First, it is important to know when, how and in what instances technology-licensing offices should be consulted before the need arises. This prepares scientists to recognize circumstances that might warrant the use of such offices. Early consultation also affords familiarity with the commercialization process, which can help generate ideas and approaches to science, and perhaps applications that a researcher might not have considered.

In addition, awareness of how tech-transfer offices work can lead to greater understanding of where and how scientists are employed in the exchange and transfer of technologies, and can provide more clarity about the vague idea of 'going into industry'. This knowledge may be useful to young scientists exploring career options, a great way to gain exposure in various positions both in academia and in companies and government agencies that actively license university-based technologies.

Many scientists may think that professional careers in IP are limited to attorneys, or that they need formal legal training to contribute in this arena. In fact, these domains are open to professionals from diverse scientific, technical and business backgrounds, and are not limited to law firms or tech-transfer offices.

In technology-based corporations, many departments, such as R&D, marketing, legal and finance, work together to develop, manage, research, evaluate, acquire, protect and license technologies or other forms of IP. Success depends on these activities, and there are many rewarding roles for

> **"How does a researcher know whether a particular discovery satisfies the criteria for being both patentable and patent worthy?"**

scientists across this spectrum.

As for career centres, they have long existed in most US universities, serving both undergraduates and alumni. For graduate researchers, however, such facilities are fairly new, perhaps because the belief has been that these trainees would follow in the footsteps of their advisers and take up faculty positions, perhaps without fully understanding the alternatives available to them.

Combining career education with research or clinical training ensures that both the graduate and postdoctoral experiences are technically sound, and that trainee scientists understand their full range of career options. Having knowledge of where and how both technologies and talents are being used leads to a greater sense of the potential of each. And as our understanding of tech-transfer offices and career centres expands, so too does our ability to derive returns on our investments in research and training.

Michael Alvarez is director of Stanford University's School of Medicine Career Center, California.