nature

www.nature.com/nature

Vol 450 | Issue no. 7170 | 29 November 2007

The paradox of Pakistan

How to avoid the mistakes of the past.

Next year will be the 50th anniversary of Pakistan's first fully fledged military coup. In 1958, it fell to General Mohammad Ayub Khan to "save the nation" from what he called "discredited politicians", and later to offer himself up for election to consolidate his power. Today's general is a different one, but the justification for continued martial law sounds depressingly familiar.

Such governance may be undemocratic, but both science and education tend to receive more investment when the generals — backed by generous aid from the United States — are in power than when a elected party is in control. It is no accident that many of the country's scientifically most productive institutions were established during the US-backed army rule of General Ayub Khan, again under Zia ul-Haq's rule in the 1980s and now under Pervez Musharraf.

Elected governments led by both Benazir Bhutto and Nawaz Sharif placed science among their lowest priorities. That reflected mistaken thinking that science is an unnecessary luxury in so poor a country, as well as the tendency of these rulers to fill science and education posts with friends in need of patronage. But it is widely accepted that science should have an important role in countries such as Pakistan, helping to develop a skilled population, build robust institutions and assist rational policy-making.

Washington's relationship with Pakistan has been dictated mostly by US priorities in foreign policy. Today, Musharraf is favoured because of his opposition to the Taliban and al-Qaeda. Back in the 1980s, Pakistan under military rule was rewarded by the Reagan administration for taking a frontline role in the war against the Soviet-backed takeover of Afghanistan.

Although Pakistan has pressed ahead with its nuclear-weapons programme, the country's administrations have proved reluctant to spend much of their own money on non-military research, relying instead on US largesse. But this creates problems, as one former minister of agriculture puts it: "When the Americans need us, they shower us with riches. We spend like there is no tomorrow. When they leave, we are unable to maintain our new scientific palaces, and our institutions begin to crumble."

And so, with yet another US-backed Pakistani military ruler in trouble, will history repeat itself? General Musharraf has shown that he is a ruthless military dictator. In the past four weeks, he has shut down the country's independent media, locked up thousands of his political opponents and sacked an increasingly independent-minded judiciary.

But there is another side to the eight years of his rule. In that time, some 2,000 students have been sent abroad for PhD training; there has been a 60-fold increase in the science budget. Pakistan has 12 million Internet users, and mobile phones are in the hands of 65 million people. And a large university building programme is under way. Some of this is down to Atta-ur-Rahman, Musharraf's right-hand man and head of the country's Higher Education Commission.

"Pakistan has proved reluctant to spend much of its own money on non-military research."

Yet, if the past is to be our guide, then entities that will see their budgets cut after elections on 8 January could include the university expansion scheme and the science ministry. The money set aside for maintaining information, computer and telecommunications infrastructure might also be redeployed. An innovative scheme to invite foreign faculty members to Pakistan might be in trouble. And some of the students sent abroad for PhDs will be reluctant to return home if the research environment reverts to what it used to be.

There are, of course, many important differences between today's Pakistan and the military dictatorships of old: for the moment, General Musharraf, who was expected to step down as head of the army this week, may remain a player in any future political scenario. If this happens, Rahman is likely to stay in his job and continue with the reforms. And Musharraf has appointed the highly regarded Shams Kassim Lakha, former president of the Aga Khan University in Karachi, to head the Ministry of Science and Technology.

But Musharraf's survival in any position of power is by no means assured. There is talk of his removal by the army in a counter-coup. There have also been several attempts on his life. None of his military-president predecessors had any lasting democratic role.

The bottom line is that both Pakistan's incoming rulers and their foreign supporters want a stable and secure Pakistan. Many of the existing reforms to science and higher education are crucial to both these aims. If these reforms were allowed to stagnate or die, no one's interests will served, least of all those of Pakistan. ■

An inconvenient truth

Research on human embryonic stem cells must go on.

The top item on prime-time television news in Germany on 21 November featured a statement from research minister Annette Schavan. She was responding to the publication of two studies in which scientists had reprogammed mature adult human cells to behave in a similar way to embryonic stem cells. The findings, she said, vindicated her preference for adult stem-cell research and reprogramming over work on human embryonic stem cells. After all, who needs embryonic cells if it is possible to flick a switch in skin cells to make them a source of virtually any type of cell for perfectly matched tissue replacement?

In the studies in question, researchers at two laboratories reprogrammed mature skin cells, giving them the characteristics of human embryonic stem cells so that they could be coaxed to differentiate

©2007 Nature Publishing Group

into other cell types, such as neurons or heart cells (K. Takahashi *et al. Cell* doi:10.1016/j.cell.2007.11.019; 2007 and J. Yu *et al. Science* doi:10.1126/science.1151526; 2007).

Around the world, opponents of human embryonic stem-cell research such as Schavan have leapt on these results to justify their support of tight controls on the work. But this is exactly the wrong time to constrain research on human embryonic stem cells, which for one thing will be required to help scientists work out how best to coax adult cells into becoming new tissues. Both labs say that they could not have made their breakthrough without the work that has been done on embryonic stem cells.

Stem-cell scientists have acclaimed the latest results, while acknowledging that they represent only a first step, and that many issues need to be resolved before there is any chance of applications in the clinic. James Thomson of the University of Wisconsin-Madison, one of the scientists who first isolated human embryonic stem cells and co-author on one of the recent studies, chose this moment to expand publicly on his qualms about using human embryonic stem cells. And Ian Wilmut of the University of Edinburgh, UK, whose team created Dolly the sheep, the first cloned mammal, says that he is abandoning plans to work on human embryonic stem cells.

Many stem-cell scientists share this general unease, both because of the dilemma of working with embryos and because women must donate eggs for the process, in a highly invasive procedure. But they have nonetheless gone ahead with such work because they see it as scientifically necessary if clinical benefits are to be derived any time soon from our growing understanding of cellular differentiation.

These scientists are not oblivious to the ethical issues and they are not merely indulging personal fascination. They have not denied the importance of doing research on adult stem-cells and reprogramming in parallel. It would be a relief for them if all the scientific problems had been solved in the papers published last week — abandoning work on human embryonic stem cells would allow them to operate with a clear conscience and without having to defend their work all the time.

"This is exactly the wrong time to constrain research on human embryonic stem cells."

From the researchers' viewpoint, the debate surrounding human embryonic stem-cell research has some parallels with that on animal research. Many would be delighted to abandon the bureaucracy, cost and general inconvenience of doing contentious work. Where genuine alternatives are available, researchers will grasp them. Just as soon as there is no scientific need to work on embryonic stem cells, researchers will design their experiments to use much easier material. But that moment has not yet arrived. ■

The big picture

Europe needs to find a way to prioritize and build large scientific facilities.

European nations remain unable to agree on a general mechanism for the prioritization and construction of scientific facilities that are large enough to require continent-wide collaboration. One such case is the current stalemate over a proposed neutron-science facility, the European Spallation Source (ESS).

Although X-rays from synchrotrons can be used to probe structures, neutrons are particularly sensitive to light atoms and so can provide a fuller picture of important crystals and molecules. According to some estimates, Europe is home to roughly two-thirds of the world's 6,000 neutron scientists. But they lack a state-of-the art neutron source on a par with the Spallation Neutron Source currently being commissioned at the Oak Ridge National Laboratory in Tennessee.

Germany decided back in 2002 that it would not provide the funds necessary to host the facility. Now Sweden says that it would like to host it, and a joint bid by Spain and Hungary would use European Union (EU) infrastructure funding to finance a site in one of the two countries. But in the absence of an effective framework to share the construction and operating costs for such a facility between the nations that would use it, construction is unlikely to begin any time soon.

The larger European nations have effective systems of their own for selecting and building research facilities. In Britain, for example, the scientific research councils cooperate to create a facilities road map. But since the European Commission developed its own research policy, supplanting the patchwork of discipline-specific research collaborations that had developed between European nations, there has been no effective mechanism for the construction of larger facilities to serve the whole continent.

The seventh Framework Programme for funding research, for example, which runs until 2013, makes no provision for the construction of facilities. The European Strategy Forum on Research Infrastructures, an ad hoc group of senior research administrators from constituent countries, has already agreed on a list of European facilities, including the ESS, that scientists want to see built. But there is currently scant prospect of this list being converted into an actual construction schedule. That's mainly because EU member states — especially major ones such as Britain and Germany — are reluctant to release any of their existing national resources into a Europe-wide facilities pot.

The research ministers of the larger member states (backed by their own senior research administrators — and by many senior researchers in their own countries) are loathe to endorse a wishlist whose very existence might result in money that is currently spent on national grants and facilities being siphoned off to pay for major facilities in far-off lands.

This reluctance is perfectly understandable — but it has to be overcome if Europe is going to build large, general-purpose research facilities, such as the ESS. Another committee — the European Research Area Board — is currently being established, and should be well-positioned to prioritize the construction of such facilities and then to promote these priorities. But perhaps what is needed most is a more constructive attitude towards such collaborations from the larger EU member states. ■

©2007 Nature Publishing Group

RESEARCH HIGHLIGHTS

Taming the transposon

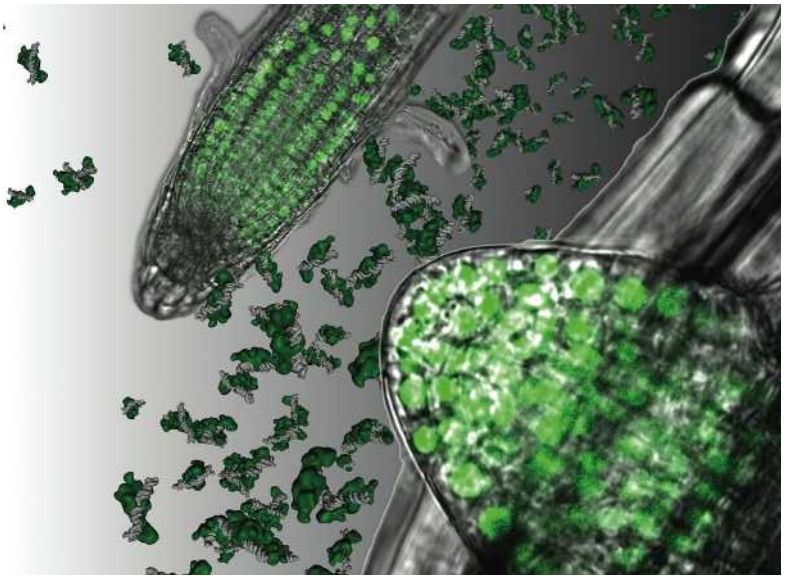
Science **318**, 1302–1305 (2007)

Two proteins that help plants to develop optimally for their light environment may have evolved from 'jumping genes', called transposons, researchers have found.

Many plants use proteins known as phytochromes to sense and respond to red (and far-red) light. Two proteins involved in phytochrome signalling, FAR1 and FHY3, are encoded by DNA sequences similar to the sections of transposons that enable them to jump around the genome. These sections encode enzymes known as transposases.

Haiyang Wang at the Boyce Thompson Institute for Plant Research in Ithaca, New York, and his colleagues report that *FAR1*- and *FHY3*-like genes are intermingled with transposase genes in many flowering plants — and that the FAR1 and FHY3 proteins turn on other genes required for phytochrome signalling. They suggest that transposons may have provided a supply of 'control' genes in evolution, enabling flowering plants to respond sensitively to light.

The image shows FHY3 protein (green) accumulated in the nuclei of *Arabidopsis thaliana* root cells.



SCIENCE

NANOELECTRONICS

Carbon patchwork

Nano Lett. **7**, 3394–3398 (2007)

Graphene — single sheets of graphite-like carbon — has been hailed as a microelectronic wonder material. Richard Kaner and his team at the University of California, Los Angeles, have come up with a novel way of making films of the two-dimensional electrical conductor that can be patterned into circuits only a few nanometres thick.

Previous approaches to such manufacture have involved stripping down graphite layer by layer or heating silicon carbide to extreme temperatures. By oxidizing graphite and then using water to separate the product into individual flakes the team has come up with a simpler and cheaper method. Graphite oxide flakes are sprayed onto a heated surface and reduced back into graphene using hydrazine. Overlapping flakes create a continuous patchwork, from which the team has made tiny transistors.

STRUCTURAL BIOLOGY

Sneaking in

Nature Struct. Mol. Biol. doi: 10.1038/nsmb1342 (2007)

Researchers have solved the structure of the protein that allows the measles virus to bind and enter host cells. The finding may provide a template for the design of drugs that block the virus's entry.

Christopher Garcia and his colleagues at Stanford University School of Medicine in California crystallized the virus's haemagglutinin protein. They report that it is folded like another class of protein, the neuraminidases, despite limited sequence and functional similarity to that group.

Haemagglutinin's structure resembles a propeller, with six blades around a cavity. Amino acids that have a crucial role in binding to host-cell receptors are located outside the rim of the cavity.

POPULATION BIOLOGY

Polar bare

Proc. R. Soc. B doi:10.1098/rspb.2007.1307 (2007)

When a declining population becomes too thinly spread, the sheer unlikelihood of meeting a potential mate will reduce reproductive success and hasten the decline. The possibility of this phenomenon — the Allee effect — affecting polar bears has been modelled by Péter Molnár of the University of Alberta in Canada and his colleagues.

In Canada, selective hunting of male polar bears has reduced the ratio of males to females available to mate — that is, without cubs. Using population data from Lancaster Sound, in Nunavut, collected annually between 1993 and 1997, the team predicts at what point, as the overall population falls, male scarcity would lower fertilization rates. Should the Lancaster Sound population decrease by two-thirds and maintain its current sex ratio, Molnár and his team predict that only 81% of females would mate successfully — compared with an estimated 99% today.

R. GREEN/TAXI/GETTY

QUANTUM PHYSICS

Mini mysteries

Nature Photon. doi:10.1038/nphoton.2007.227 (2007)

Researchers have designed a quantum dot that emits high-frequency streams of single photons, a breakthrough in the field of quantum cryptography.

Streams of photons, the particles that carry light, are fundamental components in quantum information science, and could be used to exchange cryptographic keys that foil eavesdroppers.

Stefan Strauf and his team at the University of California, Santa Barbara, measured an output of 4 million photons per second from their semiconductor device. This outstrips previous methods for generating single photons, which have managed up to about 200,000 photons per second. A more powerful source laser plus a perfect detector, they say, promises rates higher than 100 million per second — close to making the exchange of quantum information practical.

©2007 Nature Publishing Group

PHYSICS

Picking the packing

New J. Phys. **9,** 406 (2007)

Mathematicians still cannot explain why spheres take on a range of packing arrangements — some loose, some tight— when they fill a fixed volume randomly. But physicists can now control the arrangement formed by small ferromagnetic beads.

Geoffroy Lumay and Nicolas Vandewalle at the University of Liege in Belgium report tuning the beads' packing by applying magnetic fields. Turning up the magnetic force increases repulsion between the beads and thus forces them into a looser configuration. As the magnetic forces are removed, the beads slip into tighter arrangements. This technique could become important for engineers designing emulsions because knowing the number of droplets in a given volume permits an emulsion's viscosity, among other properties, to be accurately determined.

ECOLOGY

Blue-sky thinking

Geophys. Res. Lett. **34,** L20707 (2007)

The growing frequency of wildfires in the western United States has been attributed to increasing temperatures and drought conditions, often linked to global warming. But two geographers in North Carolina believe that there is a third important factor in the northern Rocky Mountains: fewer storms.

Paul Knapp of the University of North Carolina at Greensboro and Peter Soulé of Appalachian State University pored over more than a century's weather records. They noticed that major mid-latitude cyclones, which bring precipitation and cooler temperatures, are arriving later in the year and occurring less often.

The trend was gradual for most of the twentieth century but became more pronounced in the 1980s — corresponding to sharp changes in wildfire statistics, which have been kept over six decades. This relationship explains about a fifth of the rise in wildfire activity, the authors estimate.

PSYCHOLOGY

Mind and memory

Science **318,** 1257 (2007)

A trait called 'theory of mind' involves being able to infer the current mental states of others. This has been thought to require the ability to imagine a personal experience occurring in other contexts on the basis of remembered events. But the link has proved difficult to test.

Shayna Rosenbaum and her co-workers, from the York University and Rotman Research Institute in Toronto, Canada, did just that — and debunked the assumption. They gave a range of tests to two patients with rare brain injuries that cost them the ability to recall personal experiences but not the facts they learned before becoming brain damaged. The tests assessed how well they could reason about the thoughts and feelings of others and their scores were indistinguishable from those of 14 control subjects.

ANTHROPOLOGY

Jade trade

Proc. Natl Acad. Sci. USA doi:10.1073/pnas.0707304104 (2007)

Lingling-o, three-pointed jade ornaments worn as ear pendants across Southeast Asia (pictured below), were spread by seafaring traders in the South China Sea. And, suggest Hsiao-Chun Hung, of the Australian National University in Canberra, and her colleagues, they may have brought Austronesian languages with them.

PROC. NATL ACAD. SCI. USA

By working out the elemental composition of 144 jade ornaments, which were mostly recovered in archaeological digs, the authors show that at least 116 examples were carved out of stone from eastern Taiwan. Given where the ornaments were found, this implies that jade was traded throughout a 3,000-kilometre-wide region encompassing the Philippines, and parts of Malaysia, Vietnam, Cambodia and Thailand, between 500 BC and AD 500.

Because the distribution of *lingling-o* reflects that of Austronesian languages, which are spoken by about 350 million people in Southeast Asia and Oceania today, the researchers propose that itinerant craftsmen may have carried the jade to other Austronesian-speaking settlements or helped spread the languages with their trade.

JOURNAL CLUB

Ursula Euteneuer
Adolph-Butenandt Institute, University of Munich, Germany

Unexpected links between cellular organelles continue to unfold.

As a graduate student, I was intrigued by centrioles. Their beautiful ninefold symmetry, occurrence in pairs and positioning at the heart of the cell —as a constituent of the microtubule-organizing centrosome — made them appear significant. But how they are built and replicated, and even their true purpose in cell division, remained enigmatic.

Adding to my fascination was the fact that, in my cultured cells, one of the centrioles often acted as an anchor, or basal body, for a cilium-like appendage — even though the endothelial cells I used were not thought to ever have cilia.

Once considered a biological oddity, these stubby, non-motile 'primary' cilia are now known to be present in most cells of the human body (see www.bowserlab.org/primarycilia/cilialist.html) and probably serve as essential sensors whose disturbance is linked to a growing list of diseases.

I have always wondered whether there is an additional functional link between centrioles and primary cilia. A recent paper (A. Robert *et al. J. Cell Sci.* **120,** 628–637; 2007) shows that a protein involved in the biogenesis of primary cilia is also a bona fide constituent of the centrosome. Moreover, this centrosomal/ciliary component is tied into the regulation of the cell cycle.

This finding joins a series of recent discoveries of proteins that are both centrosomal and ciliary. Taken together, these studies are revealing a truly novel functional link: centrioles are there to make primary cilia; primary cilia act as sensors of external stimuli; and, as other studies have shown, external stimuli can regulate cell proliferation. So the mysterious organelles of my graduate student days are being demystified, but the story that unfolds turns out to be even more fascinating than I expected.

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

©2007 Nature Publishing Group

NEWS

Loopholes in oversight rules revealed

Documents released under the Freedom of Information Act raise questions about how the US Department of Energy (DOE) oversees investigations of misconduct at its science facilities. During such oversight, government officials do not always retain records of the investigations, it has emerged.

DOE officials overseeing an investigation into alleged misconduct at Oak Ridge National Laboratory in Tennessee approved the investigatory panel's findings at a meeting that government officials left without keeping copies of key investigative records. This is according to a 13 November ruling by attorneys at the DOE — issued in response to a Freedom of Information Act request by this reporter.

The story begins in March 2006, when a reviewer for the journal *Nature Physics* alleged that a manuscript submitted by Oak Ridge researchers in the group headed by electron microscopist Stephen Pennycook contained inappropriately manipulated data. An investigation panel convened by the lab's deputy director for science and technology, James Roberto, exonerated Pennycook and others of wrongdoing in July 2006, but other scientists expressed doubts (see *Nature* **444,** 129; 2006). In January this year, the DOE said that Oak Ridge, which is run by the not-for-profit company UT-Battelle under a 5-year, US$6.3-billion contract, had reopened the investigation. "Once again, the committee concluded there was no evidence of misconduct," an Oak Ridge spokesman told *Nature*.

According to the November ruling, DOE officials approved this first finding at a meeting last year with lab officials where they were shown copies of a panel's investigation report in binders marked "do not duplicate", and reviewed the process that the panel used to arrive at its finding of "no misconduct". The lab officials then took the binders back, the ruling says. It adds that although the DOE officials felt that it was "very important" to confirm the panel was objective and composed of people who did not work at Oak Ridge, they apparently did not keep a record of who those investigators were. Last week, Oak Ridge named Paul Peercy of the University of Wisconsin-Madison and David Williams of the University of Alabama in Huntsville, as two panel members.

Government officials oversaw a misconduct investigation at Oak Ridge National Laboratory.

OAK RIDGE NATL LAB.

"Copies of the report were returned as is normal practice with confidential documents," says Oak Ridge's spokesman, Mike Bradley. "Department of Energy officials were not at any time restricted from access to the report," he adds.

The freedom-of-information ruling cites a standard clause in the contract between the DOE and UT-Battelle as evidence that the company "clearly intended to retain control over the report of investigation". That clause says that records relating to investigations "conducted under an expectation of confidentiality" will be considered the contractor's property. In 2005, Raymond Orbach, director of the DOE Office of Science widened the clause to cover misconduct reports from all US national labs run by the DOE.

In the Oak Ridge case, the DOE did later receive a copy of the investigation report, which the recent ruling orders Orbach's office to consider for public release — overruling his deputy, Patricia Dehmer, who had refused to do so. But Oak Ridge says that it was sent not for oversight but "for informational purposes only".

"Can a government agency fulfil its oversight responsibilities with this degree of access?" asks C. K. Gunsalus, an attorney and misconduct expert at the University of Illinois at Urbana-Champaign. She says that other science-funding agencies employ oversight experts who routinely receive and retain not only investigation reports, but the data at issue and the names of people involved, and analyse these in detail — a process that for a complex case might well take longer than one meeting.

Bioethicist Arthur Caplan of the University of Pennsylvania, Philadelphia, says he can't understand why the DOE introduced a rule relinquishing control of investigation reports. "Why would they give up supervisory powers?" he asks.

Officials in Orbach's office plan to meet this week to discuss the matter. ■

Eugenie Samuel Reich

13 November ruling
www.oha.doe.gov/cases/foia/tfa0213.pdf

Deforestation on the agenda at climate meeting

Deforestation issues must be included in global talks on carbon-emissions control, experts say.

European companies seeking to offset their greenhouse-gas emissions under the Kyoto Protocol are pumping billions of dollars into clean-energy projects in the developing world. But the protocol does not include funding initiatives to prevent deforestation, which is responsible for some 20% of global carbon emissions.

Eyeing an economic opportunity that could put money in the hands of those who preserve native forests rather than chop them down, tropical countries are now banding together to alter the rules after the Kyoto accord expires in 2012. The first step is to ensure that deforestation is on the agenda at Bali, Indonesia, where international negotiators will gather next week for the latest round of United Nations climate-change talks.

"All of the big countries [with tropical forests] have gotten together to tell the world that they support the same fundamental idea," says Doug Boucher, who works on the issue for the Union of Concerned Scientists, an environmental watchdog based in Cambridge, Massachusetts. Boucher points out that Indonesia has become an international leader on the issue despite being virtually absent from the debate just six months ago. "We may look back on this in a couple of years as having been a turning point," he says.

The idea that deforestation must be addressed in any cohesive response to global warming is not new. Although reforestation projects were allowed during the Kyoto deliberations a decade ago, the question of halting deforestation was dropped, largely because of technical questions. How does one verify a decline in deforestation? And how can anyone be sure that the problem hasn't just moved elsewhere?

A decade after the Kyoto talks, advocates say the issue is ready

©2007 Nature Publishing Group

Safer way to make human stem-like cells revealed

Shinya Yamanaka of Kyoto University in Japan has refined his method for making human skin cells behave like embryonic stem cells so that it does not use a cancer-causing agent.

Late last month, Yamanaka reported making pluripotent cells — cells that can turn into any of the roughly 220 cell types in the body — by using retroviruses to carry four genes into human skin cells (K. Takahashi *et al. Cell* 131, 861–872; 2007). The four genes reprogrammed the cells into a state similar to that of cells in the early embryo. But one of the genes used to make the cell lines is *c-myc*, which can cause cancer.

Yamanaka has now shown that he can make these 'induced pluripotent stem cells' in both humans and mice by using just three factors — not including *c-myc* (M. Nakagawa *et al. Nature Biotechnol.* doi:10.1038/nbt1374; 2007).

Risk analysis for Boston lab slammed as 'shoddy'

A US government safety assessment supporting the location of an infectious-disease lab in Boston was "not sound and credible", says a report issued last week by the US National Research Council.

Construction of the $178-million Boston University National Emerging Infectious Diseases Laboratory is due to finish in 2008, but one state and one federal lawsuit are challenging its opening. The facility will house research on deadly pathogens such as the Ebola and monkeypox viruses.

The report's criticisms are aimed at a document, released by the US National Institutes of Health (NIH) in July, that evaluated the facility's potential threat to its neighbours. It finds that the worst-case scenarios proposed by the NIH, including an Ebola outbreak caused by an infected lab worker, were not relevant to assessing the true risk. Pathogens with more potential to spread, such as influenza, should have been chosen, the report says. It adds that the NIH assessment suffers from shoddy risk analysis and modelling. The NIH says it will consider the report.

AIDS worker's misconduct ban belatedly comes to light

Scott Brodie, a former researcher at the University of Washington in Seattle, allegedly committed misconduct when doing AIDS research there in the late 1990s and early 2000s, and was banned from working for the university, according to *The Seattle Times*.

The news emerges only now because the newspaper had to win a court case to get a copy of the 2003 report of the investigation. Brodie, under the name John Doe, reportedly sued the university and the newspaper to prevent release of the report.

Brodie left the university in summer 2003 and is currently employed by drug company Schering-Plough, based in Kenilworth, New Jersey. A spokesman for the company says that it has just learned of the investigation.

According to one of the university investigators, geneticist Denny Liggitt, the report was handed to the National Institutes of Health's Office of Research Integrity.

European satellite system is back on the launch pad

European transport ministers have revived plans for the Galileo satellite-navigation project. The decision comes almost six months after the collapse of a public–private consortium to build the €3.4-billion (US$5-billion) network of 30 satellites.

Under the new agreement, the European Union will provide an additional €2.4 billion in unused agricultural funds to finance the system. The decision came despite opposition from Spain, which sought a larger stake in the venture.

Now that funding is confirmed, it will be up to the member states to decide how contracts will be awarded for the project, says Kai-Uwe Schrogl, secretary-general of the European Space Policy Institute in Vienna. "This is now a new game," he says.

So far, only a single test satellite in the Galileo system, GIOVE-A, has been launched. A second, GIOVE-B, has faced repeated delays and is expected to launch in spring next year.

Europe's GIOVE-B is being readied for launch.

ESA

Californian stem-cell directors to be investigated

California officials may investigate potential conflicts of interest in the state's $3-billion stem-cell initiative.

State Controller John Chiang has asked for the Fair Political Practices Committee to look into accusations against Robert Klein, the chairman, and John Reed, a board member, of the California Institute for Regenerative Medicine (CIRM) in San Francisco. A public advocacy group has also called for both to resign after it emerged that Reed, on Klein's advice, asked the CIRM to reconsider a grant decision involving a researcher at the Burnham Institute in La Jolla.

Reed is president of the Burnham Institute and intervened after the CIRM decided to rescind a previously awarded $638,000 grant to a researcher who was not an on-site, full-time employee of the Burnham Institute.

Klein says he now realizes that Reed should never have written the letter, and that he looks forward to the investigation. "We welcome guidance so that we can learn," he says.

Clarification

The News story 'Loopholes in oversight rules revealed' (*Nature* **450,** 590; 2007) referred to a ruling issued by US Department of Energy attorneys in response to a Freedom of Information Act request. We would like to make it clear that the documents that were the subject of that request have not been released.

Endangered species cling on as rulings are overturned

The US government has revised seven decisions made by a former official who worked on the Endangered Species Act. Julie MacDonald resigned from the Fish and Wildlife Service in May after it emerged that she pressured agency scientists, and solicited advice on decisions from a friend in the online game *World of Warcraft*.

Preble's meadow jumping mouse (*Zapus hudsonius preblei*) will not now lose its listing, but will be considered threatened only in Colorado (it is also found in Wyoming). Twelve endangered species of Hawaiian picture-wing fly will be able to stretch out in 'critical habitats' larger than the 0.4 hectares per species designated for them by MacDonald. And decisions affecting several other species, including the white-tailed prairie dog (pictured) and the arroyo toad, have been deemed flawed. But research needed to revise some of these decisions will be carried out only if funding can be found.

D. HOSKING/FLPA

©2007 Nature Publishing Group

NEWS

Loopholes in oversight rules revealed

Documents released under the Freedom of Information Act raise questions about how the US Department of Energy (DOE) oversees investigations of misconduct at its science facilities. During such oversight, government officials do not always retain records of the investigations, it has emerged.

DOE officials overseeing an investigation into alleged misconduct at Oak Ridge National Laboratory in Tennessee approved the investigatory panel's findings at a meeting that government officials left without keeping copies of key investigative records. This is according to a 13 November ruling by attorneys at the DOE — issued in response to a Freedom of Information Act request by this reporter.

The story begins in March 2006, when a reviewer for the journal *Nature Physics* alleged that a manuscript submitted by Oak Ridge researchers in the group headed by electron microscopist Stephen Pennycook contained inappropriately manipulated data. An investigation panel convened by the lab's deputy director for science and technology, James Roberto, exonerated Pennycook and others of wrongdoing in July 2006, but other scientists expressed doubts (see *Nature* **444,** 129; 2006). In January this year, the DOE said that Oak Ridge, which is run by the not-for-profit company UT-Battelle under a 5-year, US$6.3-billion contract, had reopened the investigation. "Once again, the committee concluded there was no evidence of misconduct," an Oak Ridge spokesman told *Nature*.

According to the November ruling, DOE officials approved this first finding at a meeting last year with lab officials where they were shown copies of a panel's investigation report in binders marked "do not duplicate", and reviewed the process that the panel used to arrive at its finding of "no misconduct". The lab officials then took the binders back, the ruling says. It adds that although the DOE officials felt that it was "very important" to confirm the panel was objective and composed of people who did not work at Oak Ridge, they apparently did not keep a record of who those investigators were. Last week, Oak Ridge named Paul Peercy of the University of Wisconsin-Madison and David Williams of the University of Alabama in Huntsville, as two panel members.

Government officials oversaw a misconduct investigation at Oak Ridge National Laboratory.

OAK RIDGE NATL LAB.

"Copies of the report were returned as is normal practice with confidential documents," says Oak Ridge's spokesman, Mike Bradley. "Department of Energy officials were not at any time restricted from access to the report," he adds.

The freedom-of-information ruling cites a standard clause in the contract between the DOE and UT-Battelle as evidence that the company "clearly intended to retain control over the report of investigation". That clause says that records relating to investigations "conducted under an expectation of confidentiality" will be considered the contractor's property. In 2005, Raymond Orbach, director of the DOE Office of Science widened the clause to cover misconduct reports from all US national labs run by the DOE.

In the Oak Ridge case, the DOE did later receive a copy of the investigation report, which the recent ruling orders Orbach's office to consider for public release — overruling his deputy, Patricia Dehmer, who had refused to do so. But Oak Ridge says that it was sent not for oversight but "for informational purposes only".

"Can a government agency fulfil its oversight responsibilities with this degree of access?" asks C. K. Gunsalus, an attorney and misconduct expert at the University of Illinois at Urbana-Champaign. She says that other science-funding agencies employ oversight experts who routinely receive and retain not only investigation reports, but the data at issue and the names of people involved, and analyse these in detail — a process that for a complex case might well take longer than one meeting.

Bioethicist Arthur Caplan of the University of Pennsylvania, Philadelphia, says he can't understand why the DOE introduced a rule relinquishing control of investigation reports. "Why would they give up supervisory powers?" he asks.

Officials in Orbach's office plan to meet this week to discuss the matter. ■

Eugenie Samuel Reich

13 November ruling
www.oha.doe.gov/cases/foia/tfa0213.pdf

Deforestation on the agenda at climate meeting

Deforestation issues must be included in global talks on carbon-emissions control, experts say.

European companies seeking to offset their greenhouse-gas emissions under the Kyoto Protocol are pumping billions of dollars into clean-energy projects in the developing world. But the protocol does not include funding initiatives to prevent deforestation, which is responsible for some 20% of global carbon emissions.

Eyeing an economic opportunity that could put money in the hands of those who preserve native forests rather than chop them down, tropical countries are now banding together to alter the rules after the Kyoto accord expires in 2012. The first step is to ensure that deforestation is on the agenda at Bali, Indonesia, where international negotiators will gather next week for the latest round of United Nations climate-change talks.

"All of the big countries [with tropical forests] have gotten together to tell the world that they support the same fundamental idea," says Doug Boucher, who works on the issue for the Union of Concerned Scientists, an environmental watchdog based in Cambridge, Massachusetts. Boucher points out that Indonesia has become an international leader on the issue despite being virtually absent from the debate just six months ago. "We may look back on this in a couple of years as having been a turning point," he says.

The idea that deforestation must be addressed in any cohesive response to global warming is not new. Although reforestation projects were allowed during the Kyoto deliberations a decade ago, the question of halting deforestation was dropped, largely because of technical questions. How does one verify a decline in deforestation? And how can anyone be sure that the problem hasn't just moved elsewhere?

A decade after the Kyoto talks, advocates say the issue is ready

©2007 Nature Publishing Group

Brazil to boost science spend

"Don't you come back into my office asking for more money until 2010." Brazilian president Luiz Inácio Lula da Silva's playful admonishment to his science minister came after the president's announcement last week of a remarkable US$28-billion package for science and technology over the next three years. The spending promised is equivalent to 1.5% of the country's GDP — currently, science receives about 1% of GDP.

The investment is part of a federal plan to improve academic research and to counter the lack of technological innovation in the industrial sector, which is a special cause of concern. Most researchers in Brazil are still employed in the public sector, which has helped to energize the academic community, with the number of scientific publications increasing by around 9% a year since 2000. Now, 2% of the world's scientific publications are Brazilian — ranking it 15th in the world.

But despite this, the private sector is losing out because of an absence of qualified researchers. Brazil is responsible for just 0.1% of all the patents registered globally each year, according to a 2007 report by the Organisation for Economic Co-operation and Development. To address this, the government's new science plan includes extensive grants and tax breaks for research projects conducted by private companies. Funding for postgraduate qualifications will also continue to increase to expand the still insufficient scientific workforce, particularly in key fields including biofuels, nuclear technology, climate change and Amazon-related studies. The new plan also includes measures to reduce bureaucracy and excessive legislation.

Although the scientific community supports the measures, many researchers remain sceptical. Some fear that the still-evident cogs of bureaucracy will block a lot of funding before it can be spent usefully. "There was a period where projects would be approved, funding would be allocated, and then the money never came — nobody would get anything, it was surreal," says Luiz Eugênio Mello, head of the Brazilian Society of Experimental Biology. "That memory is still very fresh. But overall, the plan has good intentions and good ideas."

A common complaint concerns the heavy taxation and regulatory control over imports of research equipment. "A European researcher who needs a product can get it the next day. Here we wait three to six months for it," Mello says. "There's no way we can be competitive like that."

No surprise then that the most warmly applauded moment during Lula's speech last week came when he vowed to streamline the import of scientific instruments. On paper at least, the future of science in Brazil looks promising. ■

João Medeiros

M. AIRIO/AFP/GETTY

President Lula: science for industry.

---

to take the stage in Bali. Advances in aerial and satellite monitoring should help in answering the first question, and the tropical nations have developed various policy solutions to the second.

Costa Rica and Papua New Guinea led the Coalition for Rainforest Nations in a proposal to revive the question at the climate talks in Montreal in 2005. Recently, Indonesia has joined Brazil and other major countries in the 'Forestry 11' coalition, which includes 11 nations representing some 85% of the world's tropical forests, according to the Union of Concerned Scientists.

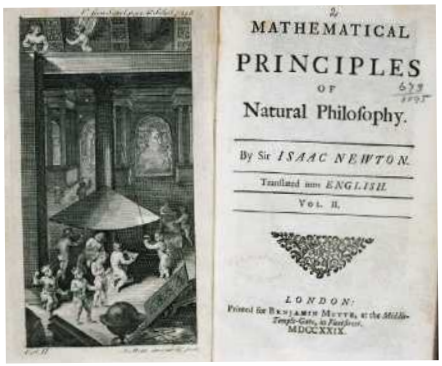
The proposals vary, but all tend to focus on verified emissions reductions, as opposed to upfront payments on projects intended to achieve certain targets. The Coalition for Rainforest Nations, for instance, proposes that deforestation be tracked on a national level, with carbon credits being sold only after a decrease in national emissions in that sector had been verified.

Although these nations have not endorsed a single idea, Boucher says, they are united in their efforts to ensure that deforestation has a clear place on the agenda for post-Kyoto deliberations.

"Deforestation is front and centre," says Steve Ruddell of Forecon, a consultancy based in New York that works on carbon markets and forestry issues. "It's still not a done deal, by any means, but I think the time is right." ■

Jeff Tollefson

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



CORBIS

Now in Arabic...

Hundreds of science books, including classics by Isaac Newton and Stephen Hawking, will be translated into Arabic for the first time. The ambitious plan by a non-profit group in Abu Dhabi has the backing of the Crown prince and funding from the Abu Dhabi Authority for Culture and Heritage.

The project, called Kalima ("word" in Arabic), is an attempt to address the fact that, although there are more than a quarter of a billion Arabic speakers worldwide, only a few hundred books are translated into Arabic each year. The group is working with more than 20 publishers throughout the Arab world. It plans to help them acquire, translate, publish and distribute about 100 books in Arabic every year. Around a quarter of these will be science titles.

"There is a particularly large gap in the Arab library of books in the natural-science category," says Karim Nagy, the Egyptian entrepreneur and book collector who directs the project. "We have therefore purposely placed a heavier weighting on it."

One book already translated is *A Briefer History of Time*, Stephen Hawking's revision of his best-seller (see above). Next year, Kalima will translate books by Niels Bohr, Werner Heisenberg, Max Planck and Richard Feynman into Arabic, and prepare Arabic versions of recent works by Roger Penrose, Steven Weinberg and Freeman Dyson. Other scientists to be translated include Stephen Jay Gould, Stephen Wolfram and James Watson. Eventually, Nagy hopes also to begin translating Arabic books into English and other languages. ■

Jascha Hoffman

©2007 Nature Publishing Group

SPECIAL REPORT

Monuments and instruments

The architecture of the buildings in which researchers work can have a crucial effect on the fruits of their labour, **Emma Marris** finds.

R. SOBOL/ZUMA PRESS/NEWSCOM

"Vision — and quirkiness." Those were the qualities that the Ray and Maria Stata Center at the Massachusetts Institute of Technology (MIT) was meant to embody, according to an MIT statement at the time of its dedication in 2004. Now it also embodies a spirit of leakiness. "When I was an undergraduate visiting the lab, I thought that all the random buckets and X marks on the carpet must be for some high-tech rover experiment," says student Shuonan Dong. "When I became a graduate student, I realized that the buckets were just buckets, to collect rain." In October, MIT sued the building's architect and a contractor over "design and construction failures" in the building, which houses the Computer Science and Artificial Intelligence Laboratory, among other things.

The architect involved is no run-of-the-mill designer, but one of the world's most renowned — Frank Gehry, creator of such monuments as the Guggenheim Museum Bilbao in Spain and the Walt Disney Concert Hall in Los Angeles. Gehry is unimpressed with the suit: "MIT is trying to use it to get a million and a half in insurance money," he says. "If there are leaks, I am sure they are fixable. The staff and faculty who use the building love it. They are not complaining."

In fact, some are complaining. But the leaks are not their only, or even their most pressing, complaint. There are scientists who object to the way their new surroundings seek to shape their work and even their ideas. One irritation is the building's internal openness. According to Gil Alterovitz, a researcher affiliate working in bioinformatics: "It was picked to be a design showcase, but for the researchers, maybe they don't want to be on display. Everyone can see what you're doing, and there is a sign that has been up for a year or two that says 'don't tap on the glass, we feel like fish'."

Although the Stata building is unique in some respects (what other lab has a façade that resembles "a party of drunken robots", in the words of no less a critic than the architect himself), in its enforcement of transparency it is part of a trend among scientific buildings seeking architectural prestige.

Openness has become something of an architectural obsession. At London's Natural History Museum, for example, the new Darwin Centre buildings expose curators to visitors through high glass walls. And the response is not uniformly favourable. At the inward-looking aquarium of the James H. Clark Center at Stanford University, researchers put posters up on the glass to get some privacy, says Kenneth Kornberg, who runs an architecture firm in California and Tokyo that specializes in research space.

For a new building given to — or imposed on — existing faculty, an architectural concept embodying a specific attitude to research can be a problem. But with a new building for a new institution, it can be a way of setting the agenda. Janelia Farm, the new flagship lab for the Howard Hughes Medical Institute outside Washington DC, is "based on randomness and the opportunity for interactivity", according to its architect Rafael Viñoly. Its labs are glass-walled. The verdant lawns on its flat roofs seem to go on for ever (and the fact that local deer plunged off the edges when the building first opened testifies to the power of the illusion).

"It was picked to be a design showcase, but maybe the researchers don't want to be on display."

Bob McGhee, who helped design Janelia Farm, knows that those open, glass-walled labs can turn people off. But this is part of the point. "The facility helps you decide which kind of people will really work there," he says. It becomes a way for potential faculty to feel out whether they are ready for the kind of culture the team behind Janelia Farm wants to create. Janelia's communality is built into the design in other ways, too: the site gives pride of place to that great interactive crucible of scientific breakthroughs, a pub.

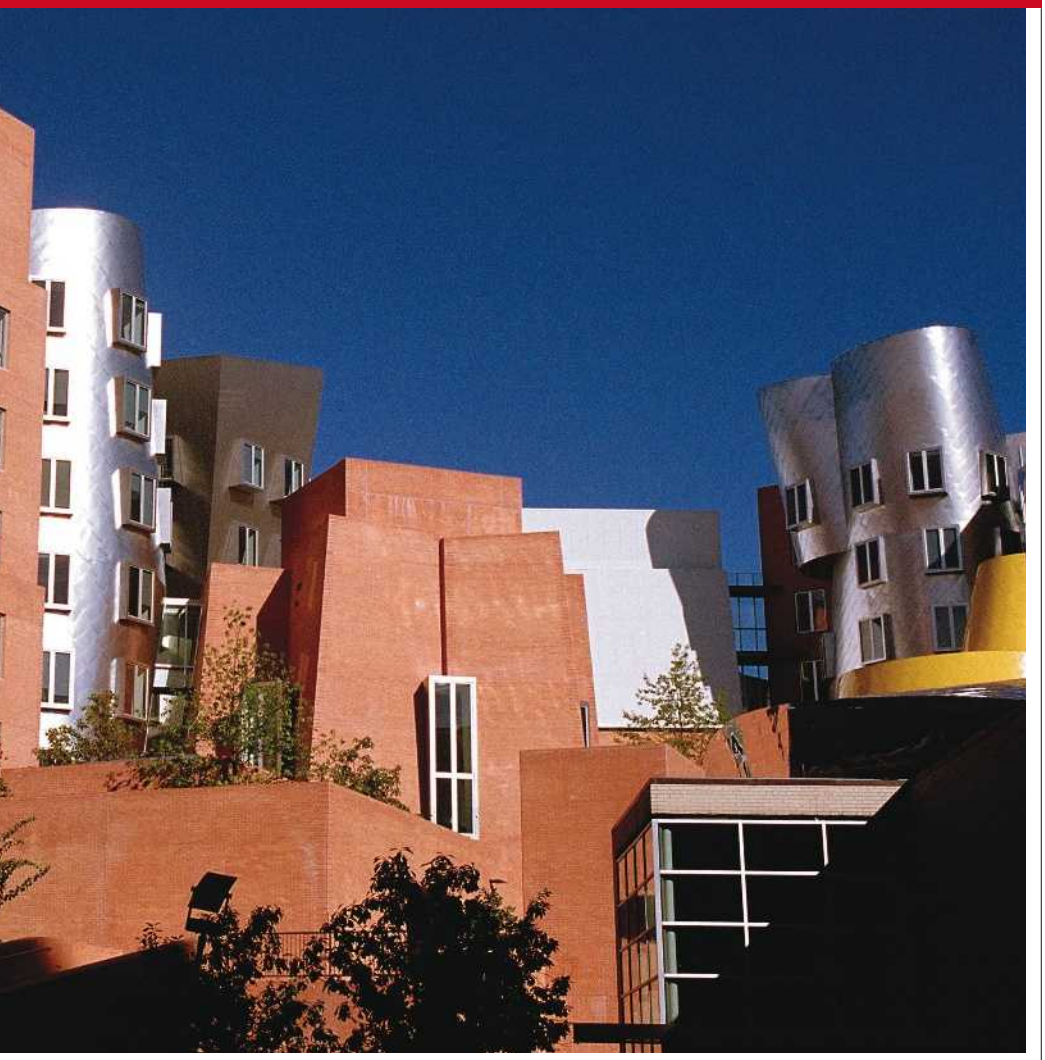
Architecture has always been used to make statements about what science is and how it should be done. Peter Galison, a historian of science at Harvard University, has traced the evolution of physical laboratory space from the temples of learning inhabited by pre-war "gentlemen physicists" through the factory-like spaces of the Second World War and beyond into today's more eclectic styles. The buildings, he says, reveal how their makers see scientists, and scientists inside the buildings take cues from the buildings when thinking about themselves.

These cues are not always imposed from outside. Take the central building at Fermilab in Batavia, Illinois, which the lab's director Robert Wilson wanted to be "a soaring centralized monument," says Galison, "not a bunch of sprawling tin shacks". Wilson wanted a grandeur to match that of the lab's undertakings in fundamental physics.

Fermilab's distinctly flashy Wilson Hall, as it was eventually named, was an exception in the 1970s. But various factors are now at work to promote a greater role in lab architecture for striking designs and the big-name architects who provide them. One is the role of large-scale philanthropy. Private donors are often more excited by the idea of a concrete-and-steel artwork masterminded by a celebrity in wire-framed specs than by cookie-cutter lab space. It speaks to the importance and contemporaneity of the science they are funding.

Another factor may be a broader sense of what an academic is. "Scientists now move in different circles," says Galison. "They are all moving back and forth between basic and

©2007 Nature Publishing Group



The Ray and Maria Stata Center at MIT has drawn both praise and complaint.

applied science. They don't see themselves as part of a corps of employees." And so you start to see buildings with all the extras — carpets, wooden banisters, art. These buildings "try to be places where people other than scientists are going to be comfortable, like venture capitalists and people starting companies", Galison says.

And then there are quite specific issues. McGhee explains that, when designing Janelia Farm, he realized how 'dry' lab space — desks with computers — is supplanting traditional wet lab space, with its complex plumbing and ventilation requirements. Wet labs are creatures unto themselves, and not all architects are experienced in designing them. Sometimes this problem is solved by splitting up the job — a famous architect will design the shell and another firm will design the lab space.

One such building is the Buck Institute for Age Research in Novato, California (see page 603). I. M. Pei, a comparatively rare example of a big-name architect who has worked on several scientific buildings, including the iconic National Center for Atmospheric Research in Boulder, Colorado, designed the travertine exterior, with its floating staircases. Kornberg and his company worked on the interiors, an arrangement that in general he is not necessarily sold on. "That hybrid does not always produce the best buildings," he says. "Would you have one architect do the shell and another design the interior of your house?"

The move towards greater dry lab space increases the possibilities for involving non-specialist architects in lab building, and for flexibility in the buildings themselves. Gehry thinks that people working in the Stata building who don't like it should alter their environments as they see fit. "I was trying to give them the body language of a building that wasn't precious, so they could push and kick and nail into it," he says. Viñoly's attitude that "the building should be like an instrument, like the equipment the scientists use", also bespeaks flexibility — as problems and mindsets change, so architecture should be willing to change with them.

But there will always be room for disgruntlement — and delight. Now a graduate student, Dong sings the praises of the Stata. "I love the open space, and I love the Dr Seuss style. Many of my friends work in practical buildings with a grid-like layout, where it never rains indoors. But my oh my, how boring that must be!" ■

Additional reporting by Heidi Ledford and Jennifer Meyer.

©2007 Nature Publishing Group

POLAR BEARS DYING IN EARLY ICE MELT
Oldest and youngest bears in Hudson Bay die off after warm springs.
www.nature.com/news

Europe looks to draw power from Africa

The power needs of Europe, the Middle East and North Africa could be met by an ambitious idea to network renewable energies across the region. The cornerstone of the plan, developed by a group of scientists, economists and businessmen, involves peppering the Sahara Desert with solar thermal power plants, then transmitting the electricity through massive grids.

Prince Hassan bin Talal of Jordan was scheduled to present this green-energy idea, dubbed DESERTEC, to members of the European Parliament in Brussels on 28 November.

The vision is ambitious: it would require roughly 1,000 100-megawatt power plants, using mirrors to concentrate energy from the Sun's rays, throughout the Middle East and North Africa to meet the region's projected energy needs. A high-efficiency electricity grid, yet to be built, would then ferry the power around and across the Mediterranean Sea and northern Europe.

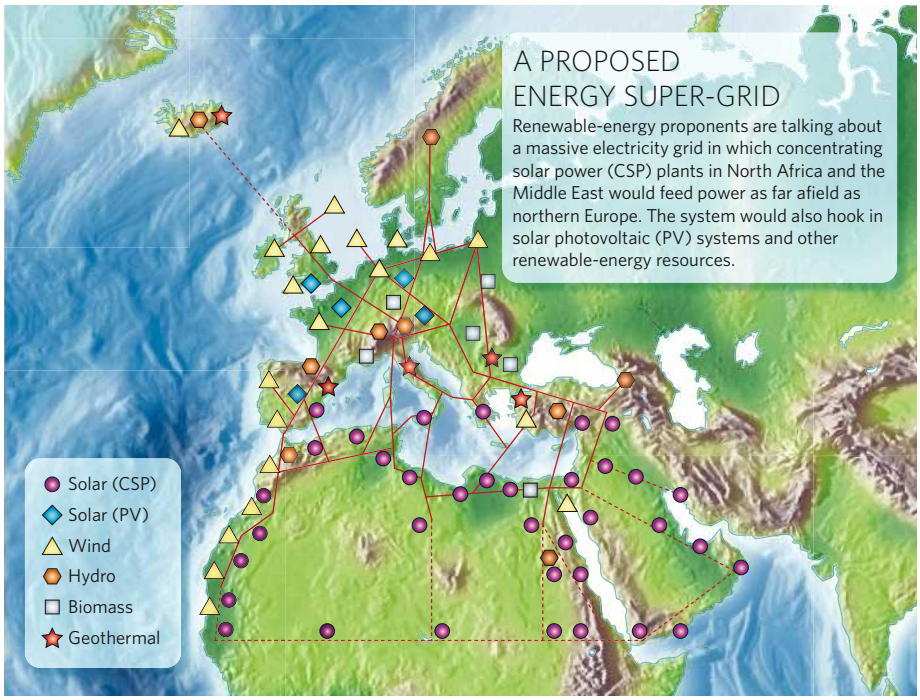
"The technology for the DESERTEC concept is available and can offer unlimited, cheap and carbon-dioxide-free energy to Europe," says Gerhard Knies, a retired physicist based in Hamburg, Germany. Knies is co-founder of the Trans-Mediterranean Renewable Energy Cooperation (TREC), which came up with the DESERTEC idea.

The European Union has a binding target to get 20% of its energy from renewable sources by 2020, so the idea is gaining support in some areas. "It makes a lot of sense to profit from the high amount of solar radiation in the deserts," says Robert Pitz-Paal, head of the solar research department at the German aerospace agency DLR. But with a price tag of almost €400 billion (US$595 billion), it remains to be seen if DESERTEC will be adopted politically.

"Unless it's extremely cheap, it won't stop people using easy-to-get fossil fuels," says Jon Gibbins, an energy engineer at Imperial College London. "We didn't stop using coal in the last century because of oil."

Nationalistic concerns may also be a stumbling block, with European politicians reluctant to be dependent on Africa. For instance, Hermann Scheer, a member of the German parliament, is in favour of renewable-energy approaches, but is pushing for European energy autonomy through small, decentralized power plants on European soil.

The vision of covering the Sahara with solar panels to generate electricity for Europe goes back to Frank Shuman, a Philadelphia-based inventor who built a prototype solar thermal plant in Egypt in 1913. But the idea never took off, and today solar power in the region comes from relatively small solar-cell installations on houses and other individual buildings.



SOURCE: TREC

Power sharing: how the proposed renewable-energy network might look.

In 2003, Knies co-founded TREC and began presenting the idea at conferences. He eventually got the attention of the German environment ministry, which has commissioned three technical studies to evaluate the concept. Germany aims to get 50% of its energy from renewable sources by 2050, and is a leader in developing solar thermal power. "The ministry was interested in knowing if it's feasible to import solar energy from North Africa and to penetrate the market there," says Ralf Christmann, an officer at the ministry.

> **"It makes a lot of sense to profit from the high amount of solar radiation in the deserts."**

The DESERTEC scenario foresees a mix of renewable energies, from wind to geothermal to biomass power (see map). But the core is solar thermal power, which uses solar energy stored in a special heat-retaining fluid to drive a turbine and create power. First demonstrated in 1982 with a 10-megawatt plant in California's Mojave Desert, solar thermal plants can now produce electricity at a cost of about 15–20 eurocents per kilowatt-hour. According to the DLR, further improvements in technology and scale could bring that down to less than 10 eurocents per kilowatt-hour, making it more competitive with coal.

Initial solar thermal plants are being planned in Algeria, Egypt and Morocco, with more under construction in Spain and Italy.

Bringing electricity from Africa to Europe presents another challenge. The DLR says that €45 billion of the overall budget should be invested by 2050 to place high-voltage d.c. transmission cables throughout the region. Such a line already exists between Norway and the Netherlands.

The DESERTEC group is asking parliamentarians to set up a €10-billion fund to finance the development of solar thermal plants over the next 7 years, and to establish a political framework for the idea. But although the project may not take off on the scale its supporters hope for soon, solar thermal power could still pick up elsewhere. "Right now, 1,000 megawatts of solar thermal energy are being built in California and Nevada deserts, and we are planning an additional 5,000 megawatts," says Dan Kammen, director of the Renewable and Appropriate Energy Laboratory at the University of California, Berkeley. "Exploiting solar energy from deserts is a good idea worldwide." ■

Emiliano Feresin

©2007 Nature Publishing Group

ON THE RECORD

"Since I'm not in academia, I only publish papers when I think I've found something cool."

Garrett Lisi, surfer, snowboarder, physicist and author of the much discussed grand-unification preprint "An Exceptionally Simple Theory of Everything" endears himself still further to the academic mainstream.

ZOO NEWS

Climbing the walls

Keepers at St Louis Zoo in Missouri remain baffled as to how a cheetah managed to scale a 3-metre wall and briefly escape her enclosure — despite the fact that this is the third time one of the cats has achieved this feat at the same exhibit since 2000.

M. JACOB, ST. LOUIS ZOO

SCORECARD

▲ **Plastic snakes**

Officials in the Australian city of Tamworth have enlisted toy snakes in an attempt to scare away tens of thousands of starlings messing cars and terrorizing residents with their droppings.

▼ **Irish salmon**

Northern Ireland's only salmon farm has been wiped out after a 26-square-kilometre swarm of jellyfish destroyed 100,000 fish, causing damage valued at more than US$2 million.

SHOWBIZ NEWS

Queen chancellor

Brian May's stellar academic career continues apace — fresh from defending his doctoral thesis on astrophysics, the former Queen guitarist has been appointed chancellor of Liverpool John Moores University.

Sources: The Daily Telegraph, BBC, Reuters, Associated Press

SIDELINES

Plant-disease controls sap outbreak responses

Microbiologists in the United States are expressing concern about a government proposal to limit research on several plant pathogens because of their potential to be used as bioweapons. The researchers say that the plan to subject rice and citrus disease agents to the same restrictions as Ebola virus and anthrax are ill-conceived and will limit the response to a natural outbreak.

The US Department of Agriculture (USDA) plans to add four plant pathogens to the government's list of 'select agents'. Created to keep infectious diseases out of the hands of would-be terrorists, the current list includes 81 human, animal and plant pathogens and toxins. Researchers studying substances included on the list must confine their work to high security labs, file mountains of paperwork and submit to background checks.

The latest draft includes pathogens of rice, citrus, rye and deciduous trees that have wreaked havoc elsewhere, but have not crossed US borders. The USDA is accepting public comment on the proposal until 3 December.

Many scientists think that plant diseases should not be lumped with deadly human pathogens such as smallpox. Most plant-disease research is already subject to strict regulation and quarantine, they point out. Adding more plant diseases to the select-agent list could have a chilling effect on research, says Jan Leach, a microbiologist at Colorado State University in Fort Collins, who studies *Xanthomonas oryzae*, a bacterium that causes leaf blight of rice. "Some of us will just drop out. We won't work on those pathogens anymore," she says.

Her lab halted research on one strain after it was added to the list. But if the proposed changes go forward, all strains would be subject to the extra regulations. The disease — endemic in Asia and Africa — poses minimal threat to the United States because of different climate and farming techniques, she says.

In Florida, the select-agent law stymied the state's response to a 2005 outbreak of citrus greening, says Wayne Dixon, a senior plant pathologist at Florida's agriculture department in Gainesville. Now, the USDA plans to remove from the list one strain that is widespread in Florida, while adding another that is currently hitting orange trees in Brazil. Dixon worries that if the Brazilian strain hits Florida, research and epidemiology will again be handcuffed by the disease's select-agent status. For example, researchers who perform lab tests on potentially infected tree samples must destroy them if the sample reveals a positive result.

Michael Firko, a USDA official who was involved in drawing up the new select-agent list, says the agency attempts to balance national security with scientific freedom by updating the list biennially and seeking input from researchers. He cites the delisting of plum and soya bean diseases in 2005 and the current proposal to remove one strain of citrus greening. Yet for many plant pathologists, these changes don't happen often enough or fast enough. "We had our hands tied for two years," says Dixon. ■

Ewen Callaway

L. M. ALAVREZ/AP

Southern Gardens Citrus in Clewiston — one of Florida's biggest producers of oranges — has so far lost almost 100,000 trees to the bacterial disease citrus greening (infected leaves shown left). Giving such pathogens 'select-agent' status can handcuff research and epidemiology, delaying response to outbreaks.

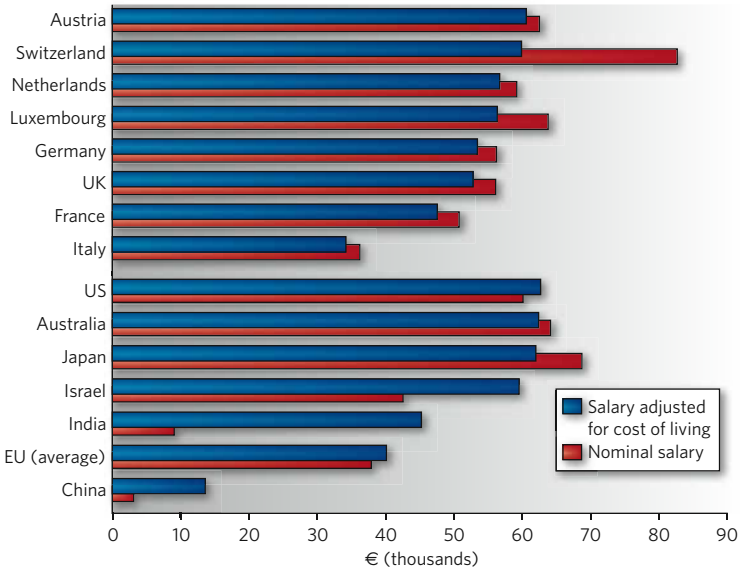
©2007 Nature Publishing Group

USGS

THE LONG WALK OF THE SALAMANDERS
Genetic studies reveal ancient travels.
www.nature.com/news

GRAPHIC DETAIL

The real value of a scientist's wage



If money is the measure of all things, Austria may be the most attractive country in Europe for a working scientist. And India is close to Europe's leading science nations, even though the average gross salary for researchers there is a meagre €9,200 (US$13,600). That's according to data comparing scientists' salaries, after taking into account the cost of living.

The European Commission report includes data on the 2006 salaries of thousands of scientists working in the public and private sectors across 38 countries. The information was gathered from an online survey as well as from national databases.

The cost-of-living figures show that only Austria, Switzerland, the Netherlands, Luxembourg and Israel offer salaries similar to those in the United States, Australia and Japan.

Unweighted, Switzerland (€82,700) has the highest gross salary worldwide. Bulgaria (€3,600) and China (€3,200) have the lowest salaried scientists of the countries in the study.

There are striking differences within the member countries of the European Union (EU). On average, EU researchers can expect €22,500 less than their colleagues in the United States, Japan and Australia — after the cost of living is factored in. "This huge disparity certainly contributes to our top people seeing better opportunities elsewhere in the world," warns EU research commissioner Janez Potočnik, who announced the results on 13 November.

Quirin Schiermeier

http://ec.europa.eu/eracareers/pdf/final_report.pdf

©2007 Nature Publishing Group

Australia prepares to ratify Kyoto Protocol

Kevin Rudd, Australia's new prime minister, has wasted little time in getting to work on an election pledge to ratify the Kyoto Protocol on greenhouse-gas emissions. Rudd led the Labor party to victory in Saturday's election. The next day, he reiterated his promise to ratify the protocol by Christmas. Ousted leader John Howard refused to ratify Kyoto during his 11-year tenure.

Rudd also confirmed that he plans to head Australia's delegation to the United Nations Climate Change Conference in Bali next week. John Connor, chief executive of the Climate Institute of Australia in Sydney, says the country's changed stance is already causing ripples. "It will be a huge boost to the Bali negotiations to have one of the blockers change into a real supporter," he says.

Australia is currently on track to meet its generous Kyoto target of an 8% increase over 1990 emissions by 2012.

T. BLACKWOOD/AFP/GETTY

Kevin Rudd, Australia's new prime minister.

Italy reappoints institute head despite controversy

The Italian government has confirmed the reappointment of Enrico Garaci as head of the ISS, Italy's national health-research institute in Rome, despite public criticism of his distribution of research money without peer review (see *Nature* 450, 320; 2007).

The government accepted his nomination by health minister Livia Turco on 16 November. But a few days later Turco gave Garaci an effective slap on the wrist by halting procedures he had begun for distributing her ministry's €3-million (US$4.5-million) fund for stem-cell research. Since the summer, Garaci had failed to issue a call for research proposals, and had earmarked beneficiaries without external review.

Turco has now said the money available will be increased to €8 million, "for which there will be an open call and evaluation by external referees".

Argentina sets up a ministry of science

Argentina's newly elected president Cristina Fernandez de Kirchner moved quickly to advance science in her administration, which begins on 10 December.

Lino Barañao was named as the first minister of the newly formed Ministry of Science, Technology and Productive Innovation. Widely respected for his research on animal reproduction and his management of scientific projects, Barañao was president of the National Agency for the Promotion of Science and Technology.

"This is a very good indication for the future direction of science in Argentina," says Juan Pablo Paz, a physicist at the University of Buenos Aires. "It is better than appointing a politician in the old-fashioned style."

Indonesia edges closer to sharing bird-flu samples

A deal is being negotiated that could see Indonesia end its policy of withholding samples from human cases of avian flu.

Until now, Indonesia has refused to share its samples with the World Health Organization (WHO), saying it is unfair that

©2007 Nature Publishing Group

ownership of the samples passes to the WHO collaborating centres, and that it does not benefit from any resulting papers or patents.

Indonesia says it will share samples under a material transfer agreement that allows research use, but gives Indonesia sovereign ownership of the samples. The country also wants access to vaccines developed using its samples. An international meeting on 20–23 November in Geneva, Switzerland, ended without agreement. But a statement, still being thrashed out by negotiators, is expected to open the way to concessions.

Indonesia has had the most deaths from the H5N1 bird-flu virus of any country — 91 so far — the most recent on 6 November.

China secures contract for Afghan copper mine

Afghanistan last week announced that a Chinese firm has won the bid for the first post-Taliban natural-resource development project: the US$30-billion Aynak copper deposit outside Kabul (see *Nature* 449, 968–971; 2007).

The state-owned China Metallurgical Group will now negotiate the final terms of the deal, in which China is pledging $2.9 billion to Afghanistan to create a mine in five years that will tap the estimated 13 million tonnes of ore.

F. J. JACKSON/ALAMY

Afghanistan's rich copper resources are set to be mined by a Chinese company.

China's operation of Aynak will be closely watched by scientists, aid agencies and economists because of environmental and socio-economic concerns raised over mining projects elsewhere.

India commits to boosting biotechnology research

For the first time, India has appointed a biologist as head of its largest research agency. The announcement coincides with the unveiling of a national strategy for biotechnology, supported by a 65-billion-rupee (US$1.6-billion) commitment over the next 5 years.

Samir Brahmachari, former director of the Institute of Genomics and Integrative Biology in Delhi, is the new chief of the Council of Scientific and Industrial Research (CSIR). The CSIR manages 41 labs with a staff of more than 18,000 scientists and has been without a permanent director since December 2006 (see *Nature* 447, 238–239; 2007).

The strategy approved by the cabinet on 14 November calls for one-third of the government's research budget to be spent on biotechnology — a 450% increase over the previous 5 years — in partnership with private-sector funding. The plan will create 50 biotech 'centres of excellence' by 2012.

Correction

The News story 'Field trials aim to tackle poverty' (*Nature* **449,** 957; 2007) implied that the idea of using mobile-phone-based reminders and time-based incentives to improve tuberculosis drug-regimen adherence originated at the Abdul Latif Jameel Poverty Action Lab at the Massachusetts Institute of Technology (MIT). In fact, the idea came from Viji Draviam of Harvard Medical School and Nishanth Sastry of MIT and the University of Cambridge, UK.

©2007 Nature Publishing Group

BUSINESS

Fever pitch

The push for new anti-inflammatory drugs has pharmaceutical companies flocking to a previously abandoned therapy. **Heidi Ledford** reports on the resurrection of interleukin-1 blockers.

In 1977, researchers isolated a novel protein from human blood that had been exposed to a pathogenic bacterium. Some called the protein a pyrogen because it triggered fever; others called it an activating factor for its ability to stimulate white blood cells. It also damaged joints and influenced liver proteins. Many were sceptical that it was just one protein.

"People didn't believe that a single molecule could do all these things," says Charles Dinarello, an immunologist at the University of Colorado at Denver and one of the first to study the protein, which was eventually called interleukin-1. "I was ridiculed quite a bit."

Thirty years later these arguments are long settled, but the list of functions for interleukin-1 continues to grow. It is now blamed for destroying insulin-producing cells in diabetics and triggering inflammation in Alzheimer's disease. Inhibitors of interleukin-1 are credited with easing the symptoms of rheumatoid arthritis and gout, and alleviating a serious hereditary inflammatory disease.

So it is hardly surprising that the protein is attracting the attention of companies ranging from Swiss pharmaceutical giant Novartis to pharmaceutical antibody developer Xoma, in Berkeley, California, which has just 300 employees. But while the potential market for interleukin-1 blockers now looks bright, many of these companies abandoned previous interleukin-1 programmes years ago. "The drug companies more or less gave up on interleukin-1," says Jürg Tschopp, an immunologist at the University of Lausanne, Switzerland. "But recently, every drug company I know has reactivated their programme."

> "The drug companies more or less gave up on interleukin-1."
> — Jürg Tschopp

That may be because drug firms see a pressing need for additional anti-inflammatory therapies. Non-steroidal anti-inflammatory drugs such as aspirin and COX-2 inhibitors come with unwanted side effects, including gastrointestinal bleeding and, in the case of some COX-2 inhibitors, an increased risk of heart attack and stroke. Another class of anti-inflammatory drugs inhibits the inflammation-inducing protein tumour necrosis factor alpha (TNF-α). These inhibitors are prescribed to treat chronic inflammatory conditions such as rheumatoid arthritis and inflammatory bowel disease. Sales of the top three TNF-α inhibitors reached US$10 billion in 2006.

But TNF-α inhibitors also heighten the risk of opportunistic infections, and simply don't work for some patients, including up to 30% of those with rheumatoid arthritis. "Everybody has focused on TNF-α for the past ten years," says Raymond Dionne, a pharmacologist at the US National Institutes of Health (NIH) in Bethesda, Maryland. "A lot of people swear by it, but really wish that something else was available."

Left on the shelf

It was the TNF-α inhibitors that silenced interleukin-1 research programmes nearly a decade ago. Amgen, a biotechnology company headquartered in Thousand Oaks, California, inherited an early formulation of an interleukin-1 inhibitor when it bought a smaller biotech company. Amgen would eventually shape that drug into Kineret (anakinra) — the only interleukin-1 inhibitor currently on the market. But by the time Amgen turned its attention to Kineret, TNF-α inhibitors were showing promise, says rheumatologist Alan Solinger, who oversaw Amgen's clinical trials for both Kineret and the company's TNF-α blocker Enbrel (etanercept) when he worked for the company. He is now vice-president of clinical immunology at Xoma. "The predecessor to Kineret sat on the shelves for about four years before Amgen realized it had an asset that could be developed," says Solinger. "By then, the TNF agents had already moved ahead."

Kineret is a recombinant form of a protein naturally produced by the body to keep interleukin-1 in check by preventing it binding to its receptor. Clinical trials showed that the drug performed well against rheumatoid arthritis and caused no side effects. But Kineret requires a daily injection and it is not an ideal interleukin-1 blocker. "This drug is very short lived," says Tschopp. "You inject it and it's gone."

Trials showed that TNF-α inhibitors outperformed Kineret in treating rheumatoid arthritis, but some say that this comparison unfairly weakened the general case for interleukin-1 blockers. "Enbrel and the others were excellent TNF blockers, but Kineret was not as good an interleukin-1 blocker," says Solinger. "Everybody moved over to the TNF bandwagon, including Amgen."

Rheumatoid arthritis is a potential target for the next generation of interleukin-1 blockers.

The US Food and Drug Administration (FDA) approved Kineret in November 2001. A month later, Amgen announced its acquisition of Immunex, the creators of Enbrel. Over time, Amgen shifted money away from marketing Kineret and focused on Enbrel, which generated almost $3 billion in 2006 sales, making it one of Amgen's top three products.

Although commercial interest faded, research still pointed to the medical importance of interleukin-1. Its processing into a mature protein requires the activation of a protein complex called the inflammasome. Tschopp and his colleagues found that the inflammasome is activated by the uric acid crystals that accumulate in the joints of patients with gout[1], and a small clinical trial showed that Kineret provided quick relief to those suffering an acute attack of gout.

Other researchers have linked a hereditary inflammatory illness called neonatal-onset multisystem inflammatory disease to the inflammasome. Another small clinical trial showed that Kineret relieved symptoms of this disease, including hearing loss, brain lesions and rashes[2]. More recently, Kineret was shown to partially relieve the symptoms of type 2 diabetes[3]. And animal studies have suggested that type 2 diabetes may respond to interleukin-1 blockade but not to TNF-α inhibitors.

Kineret itself has not proved popular, however, despite the promising trials. "No one cares about Kineret," says Eric Schmidt, an analyst at investment bankers Cowen and Company,

©2007 Nature Publishing Group

SPL

based in New York. "Maybe that's not reflective of the value of this pathway. Maybe Kineret is constrained by daily dosing. But it's just not competitive."

Trials race

Instead, companies are racing to make new interleukin-1 blockers. Amgen has an anti-interleukin-1 antibody in phase II clinical trials for rheumatoid arthritis. Novartis, which abandoned its first interleukin-1 project in 2000, now has an anti-interleukin-1 antibody in phase III trials for treatment of a rare inflammatory disorder called Muckle–Wells syndrome and phase II trials for rheumatoid arthritis. Previous trials showed remission of Muckle–Wells syndrome for up to six months after a single injection of antibody.

Several smaller biotech companies have also entered the fray. Regeneron, based in Tarrytown, New York, submitted a biologics licence application to the FDA in May for its interleukin-1 blocker Rilonacept (IL-1 Trap) for treatment of Muckle–Wells syndrome and another inherited inflammatory disorder. And Xoma is testing an anti-interleukin-1 antibody in phase I studies in patients with type 2 diabetes.

For Solinger, bringing a viable interleukin-1 therapy to market would be the culmination of a long scientific odyssey. Thirty years ago, he worked down the hall from the NIH laboratory in which Dinarello first purified the 'activating factor'. "That's one reason why I'm excited about this," he says. "Fairly soon we should know if we have a winner or not." ■

1. Martinon, F. *et al. Nature* **440,** 237–241 (2006).
2. Goldbach-Mansky, R. *et al. N. Engl. J. Med.* **355,** 581–592 (2006).
3. Larson, C. M. *et al. N. Engl. J. Med.* **356,** 1517–1526 (2007).

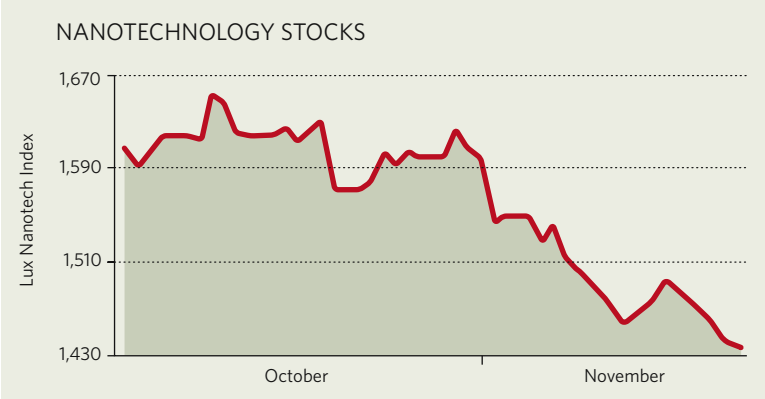
IN BRIEF

EUROPE SAYS NO Regulators have indicated that two strains of genetically modified maize (corn) are unlikely to receive approval for commercial cultivation in the European Union (EU) — setting back seed companies' hopes of a more liberal regulatory environment for genetically modified agriculture in Europe. EU environment commissioner Stavros Dimas said on 22 November that the two strains, insect-resistant Bt 11 from Syngenta and insect- and herbicide-resistant corn 1507, sold by Dow and Pioneer, would be rejected on the grounds of their possible adverse impact on butterflies and other insects.

BIOTECH BUY British drug company GlaxoSmithKline is to buy Reliant, a New Jersey biotechnology company that specializes in cardiovascular medicines, for US$1.65 billion. The agreed sale, announced on 21 November, means that Reliant will cancel a plan, announced in August, for an initial public stock offering. Analysts say that the move — the latest in a string of large biotechnology purchases by major drug companies this year — reflects Glaxo's need to line up new revenue streams in the wake of reported safety problems that are hitting sales of its diabetes drug Avandia.

ROAD TEST Daimler says it will put 100 prototypes of its new Smart Fortwo plug-in all-electric cars on the streets of London by the end of this year, as part of a plan to test the feasibility of the vehicles. The cars will initially run off nickel hydride batteries, but will switch to lithium ion ones as they become available, company officials told the *Financial Times*. The cars will have a range of 110 km and take three-and-a-half hours to recharge, the official added. The German car maker hopes to start commercial production by 2010.

MARKET WATCH



With the collapse of both the dollar and the housing market undermining confidence about the general economic outlook in the United States, 2007 has been a pretty grim year for technology stocks there. Nanotechnology is no exception, and the Lux Nanotech Index (symbol LUXNI on the American Stock Exchange) is lurching unsteadily towards the end of the year.

The index, which tracks the performance of US companies that supply nanotechnology products and services and of some major industrial corporations that rely heavily on them, has dipped with the rest of the market. Indeed, the index is at the same level now as it was four years ago.

"There's been a lack of any discernible trend" in the market for nanotech stocks in recent months, says Peter Hebert, founder of New York-based Lux Research, which compiles the index.

Hebert says that more and more nanotechnology development is taking place inside the large industrial corporations that use the technology — especially in Japan — rather than in small specialist companies. That reflects the strategic importance of nanotechnology to these big companies, but it doesn't do much for the investor seeking to make a quick profit out of a successful technology.

One bright spot was the 1 November initial public offering of stock in Nanosphere, an Illinois-based molecular-diagnostics company. The offering raised US$98 million. Hebert calls this "a pretty good achievement" and adds that it is good news for companies such as Harris and Harris, a New York-based venture-capital firm that has backed several dozen small nanotech companies. ■

Colin Macilwain

©2007 Nature Publishing Group

R. VENTE

A NEW ANGLE ON 'OLD'

Are ageing and disease two sides of the same coin? **Erika Check Hayden** reports from an institute that is betting that they are.

"Don't forget," Robert Hughes reminds a colleague in the hallway, "beer on the landing later to celebrate my impending death." Hughes isn't ill, but rather kicking off festivities for his 45th birthday. And working at the Buck Institute in Novato, California, an independent institute devoted to ageing research, tends to alter the way one thinks about birthdays.

Their work makes the Buck's 15 principal investigators peculiarly, if playfully, aware that they are adults getting older. But their institute is on the other side of the ageing divide. "It is a toddler becoming a juvenile," says Dale Bredesen, the Buck's director. Youth, like age, has its problems, and after a rough infancy, the Buck has emerged as a player with the potential to change the way people think about ageing.

The Buck was founded on the premise that ageing and disease are manifestations of the same biological processes, and they can be understood only by working across disciplines. It is a modern take, but it has its supporters, including the US National Institutes of Health (NIH). In 2005, the agency named the Buck as one of five national Nathan Shock Centers of Excellence in the Basic Biology of Aging. And in September, it gave the institute US$25 million to create a new 'interdiscipline' called geroscience: defined as the study of connections between ageing and age-related disease.

Now, the Buck is embarking on a growth spurt that will add 10 labs in 3–4 years. Plans have been drawn up for a long-term goal of three new buildings on its sunny hilltop campus and 20 more investigators, bringing the total number of labs to 45. But the expansion will depend on funding, which, despite recent votes of confidence from the NIH, may be hard won. Peers in research on ageing see the Buck as a pioneer, testing the links between ageing and disease. "It's clearly an experiment worth doing," says molecular biologist Stephen Helfand of Brown University in Providence, Rhode Island. But research into ageing has historically faced scepticism. "We're all watching to see whether it is working," says Helfand. "Are they going to do something good and bold, or are they going to be doing the same old boring things we're doing?"

> **"We don't need a National Cancer Institute and a National Institute on Aging in separate buildings."**
> — Gordon Lithgow

To understand the novelty of the Buck's approach, historical perspective helps. In the 1950s, ageing was thought to be intractable. The rather crude tools available to those who studied it didn't engender respect. Caloric restriction was known to extend lifespan in some organisms but was enigmatic. There weren't a lot of ways to study ageing beyond simple observation. "People were essentially grinding up old and young rats" and coming up with descriptions rather than learning about mechanics, says Judith Campisi, a cellular and molecular biologist who splits her time between the Buck and the Lawrence Berkeley National Laboratory.

Around 1961, researchers established that ageing was a bona fide biological process, separate from, and as important as, the pathologies associated with it. By 1974, the NIH had set up a specific branch — the National Institute on Aging — signalling that the field had earned a place in mainstream biomedical research. But soon after, the advent of genetic engineering and modern molecular biology upended the tidy distinctions between ageing and disease.

A new generation of researchers embraced molecular biology. In the 1990s, they discovered that knocking out certain genes endowed organisms such as yeast, worms and flies with unusually long lifespans[1]. But oddly enough, researchers started to find that mutating these genes to slow ageing could also stave off ailments such as cancer and diabetes[2], which

©2007 Nature Publishing Group

claim many more old people than young but that aren't usually seen as related to the ageing process itself. These studies raised the question: could ageing and disease be part of the same process after all?

Some researchers have reverted to the earlier course, says the Buck's Gordon Lithgow. "We're coming around in a circle and saying, it's the same thing. We don't need a National Cancer Institute and a National Institute on Aging in separate buildings with separate funding streams, for example," says Lithgow. "If ageing is the biggest causal factor for adult cancer, they should be together."

And thanks to a trust fund left by Beryl Buck, a nurse who died in 1975, the institute can pursue this integrated approach. Buck left behind a trust to benefit the citizens of Marin County, a wealthy enclave of towns, including Novato, just across the Golden Gate Bridge from San Francisco. The fund was to be used, in part, "to extend help towards the problems of the aged".

The trust's funds ballooned from $11 million to several hundred million dollars after its core stock holdings shot up in value. And dispensation of the funds fell into a messy legal battle between San Francisco and Marin County. But even before this brawl resolved in Marin County's favour in 1986, a board of advisers convened by the Buck Trust met to figure out how to meet Beryl Buck's goal of aiding Marin County's aged and needy. "The joke was that the trust was for the poor, elderly people of Marin — both of them," recalls Jack Rowe, who was studying ageing at Harvard Medical School in Boston, Massachusetts, at the time, and led the board of advisers. The group settled on the idea of an independent research institute focused on gerontology, and modelled on places such as the Salk Institute for Biological Studies in San Diego.

A design for life

The advisers set about duplicating two Salk characteristics: its enviable location and remarkable architecture (see page 592). The trust bought a 200-hectare tract of land atop scenic Mount Burdell, an undeveloped hilltop dotted with oak trees and boasting views of both the Pacific Ocean and the San Francisco Bay. The board commissioned renowned architect I. M. Pei to design the institute's $59-million headquarters. And despite Marin County residents' objections to the development, construction started in 1996.

To recruit a director, the advisers once again looked down the California coast — this time to La Jolla, where Bredesen, then a neuroscientist and neurologist, was working at yet another private institute, the Burnham.

Bredesen agreed to accept the job, on one condition: the Buck would have to revise its mission. "They wanted to set up an institute related to neurology and neurological disease, and I said, 'That's not right. That's not where we are in ageing today,'" Bredesen says. "It was clear that we needed to think about ageing and its relation to disease." Bredesen envisaged a self-sufficient, interdisciplinary research institute built on three cores: the biology of ageing, the study of disease and the development of technology.

"I'm totally uninterested in the fountain of youth."
— Robert Hughes

The idea that growing old and growing ill are two sides of the same coin remains controversial. Backers of the concept include David Sinclair of Harvard Medical School who made headlines with findings that a chemical in red wine called resveratrol extends lifespan and might prevent diabetes-like symptoms in mice[3] (see also page 712). "I don't see ageing as a disease, but as a collection of quite predictable diseases caused by the deterioration of the body," Sinclair says.

But others don't see it that way. The University of Michigan's Richard Miller says that Sinclair's characterization is "missing the point in a subtle but important way". Ageing is a major cause of many diseases, but not the only one, Miller argues. And, he adds, ageing has some effects that aren't considered disease states. "It's important to make a distinction between ageing and disease," Miller says.

Still, those who differ agree that interfering with the ageing process could help patients who are suffering from age-related disease. Sinclair is already running a clinical trial using resveratrol to prevent diabetes in humans.

Ageing researchers' shift towards treating disease, rather than attempting to lengthen lifespan, has legitimized the field — and attracted scientists who would have eschewed ageing research in the past. Research on ageing has always been dogged by a flake factor — the suspicion that scientists pursuing the fountain of youth were merely involved in a self-serving and impossible quest. Today, researchers studying ageing have largely distanced themselves from that perception.

Hughes, for instance, studies protein–protein interactions and uses drug-screening technologies to look for potential therapies for protein-misfolding conditions, such as Huntington's disease. "I'm totally uninterested in the fountain of youth," Hughes says. "What's interesting to me is the idea that pharmaceutical drugs that extend lifespan will be an interesting new class of drugs to treat disease."

Risky business

To Hughes and his colleagues at the Buck, the new geroscience grant offers a means to turn that idea into reality. "The essential thing that will come from this grant is speed," says Lithgow. The geroscience grant enables risky work that, by and large, will fail. But in a few cases, it might succeed, leapfrogging years of painstaking studies and hitting pay dirt much faster.

Lithgow and Campisi are going after checkpoint genes, cellular traffic cops that survey for damage and prevent faulty cells from dividing. These have long been a target in cancer research because the products of such genes prevent cancer.

Last year, Lithgow's lab reported that deactivating checkpoint genes in the nematode worm *Caenorhabditis elegans* extended its adult lifespan[4]. Deactivating such genes in mammals results in an early death as cells that can divide become cancerous, but practically no *C. elegans* cells divide. To Campisi and Lithgow, the finding suggests that losing certain checkpoint genes in non-dividing cells could preserve the cells' health. They think that this concept could extend to, for example, the brain cells of human stroke patients, as adult human neurons don't divide either.

Normally, it would take years of work even to begin to address such a question — further studies in worms, then bridging studies in other

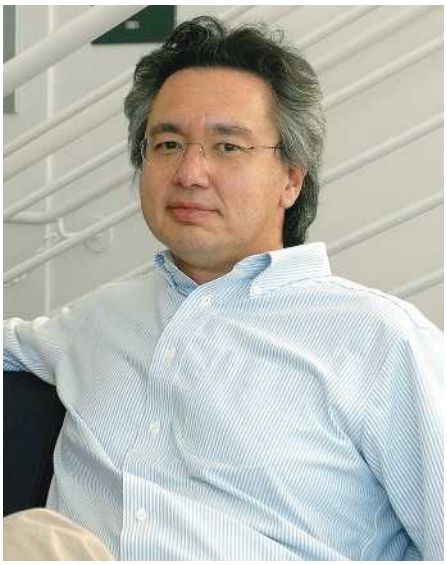
S. ALVAREZ & G. LITHGOW

Mutations in any of dozens of genes can extend the lifespan of the nematode worm *C. elegans* up to sixfold.

©2007 Nature Publishing Group

R. VENTE





Pioneering spirit: (from left) Gordon Lithgow, Judith Campisi and Robert Hughes of the Buck Institute for Age Research are pursuing new models of ageing.

model organisms. Finally, perhaps, someone would fund an extension into human cells.

But Campisi and Lithgow vaulted over these hurdles and set up a three-way collaboration with their Buck Institute colleague, Xianmin Zeng, who specializes in human embryonic stem-cell research. Zeng will grow human neurons from stem cells, and Lithgow and Campisi will study cell-division proteins in the resulting neuronal tissue. Such a leap would probably be panned as too speculative by review committees that mete out NIH funding. "It would be difficult to get grants like these reviewed in the traditional way," Lithgow says.

The geroscience grant is a boon for the Buck's brand of high-risk research. But it does not guarantee success. The Buck is up against a few obstacles — for example, it must adhere to numerous building restrictions, limiting the potential for expansion. And although its location is idyllic, Marin County is isolated; the nearest universities, Berkeley and the University of California, San Francisco, are both 50 kilometres away.

"Their slight weakness is that the Buck is a standalone institute not connected to other universities or hospitals," says Leonard Guarente, who studies ageing at the Massachusetts Institute of Technology, Cambridge, and serves on the Buck's board, "but their strength is they have very good scientists and they do good work."

But good funding doesn't necessarily follow good work. And although certain types of ageing research have captured the popular imagination, that hasn't translated into bountiful support for basic studies such as those that go on at the Buck. The Buck draws $6 million a year from the Buck Trust, and the $25-million geroscience grant will be spread over five years. With a $30-million annual operating budget and a big expansion planned, that leaves a gap of more than $20 million a year to fill through other grants and private funding.

Public profile

But money is flowing to the multibillion-dollar industry of anti-ageing products, many of which are of a dubious nature. And controversial personalities, such as former computer scientist Aubrey de Grey, are attracting public interest and dollars with sexier goals than that of the Buck's.

De Grey is convinced that all ageing is caused by cellular damage; fix that damage and humans can live for ever. This tantalizing idea has convinced private citizens to donate more than $4.5 million to support his Methuselah Mouse Prize, an incentive for researchers looking to extend mouse lifespan. But mainstream researchers are split on whether his simplification of the ageing process — and his ability to draw large donations, such as a $3.5-million gift in 2006 from American entrepreneur Peter Theil — is helping or hurting their cause. An invited talk by De Grey at the Buck, for instance, divided faculty members between those who label his work 'pseudoscience' and want the institute to distance itself from him, and those who are glad to see the public interested. "Aubrey has his weak points, and he grandstands, and that turns people off," says Campisi. "But I think the fact that he has stimulated a wealthy donor to give money to basic research on ageing is a good thing for all of us."

"I don't see ageing as a disease, but as a collection of predictable diseases caused by the deterioration of the body." — David Sinclair

Still, donors seem more excited by the hunt for a fountain of youth than they are about working out the details of normal ageing. "When I say I study ageing, people say, 'You must be rich.' But people aren't rallying to try to understand ageing. They're rallying to cure their son of diabetes," says Helfand. "The disease people get the money." The National Institute on Aging's $1-billion budget this year is smaller than that of the major disease-specific branches of the NIH, including those that fund cancer, cardiovascular and brain research. And NIH budgets overall are stagnant — a trend that has hit independent institutes such as the Buck perhaps harder than universities and medical centres, which can count on alumni and patients.

But the Buck's vision is to change that picture. If geroscience works, it will produce drugs that may one day cure or prevent afflictions such as cancer and Alzheimer's disease. And that will benefit regular people — maybe even people such as Hughes, who, upon celebrating his 45th birthday, estimates that he is halfway through his own lifespan.

"I don't think the point of what I'm doing here is to live to 120, but to live for as long as I'm going to live and be healthy, and to help the field think more clearly about real things affecting real people in real time," Hughes says. ■

Erika Check Hayden writes for *Nature* from San Francisco.

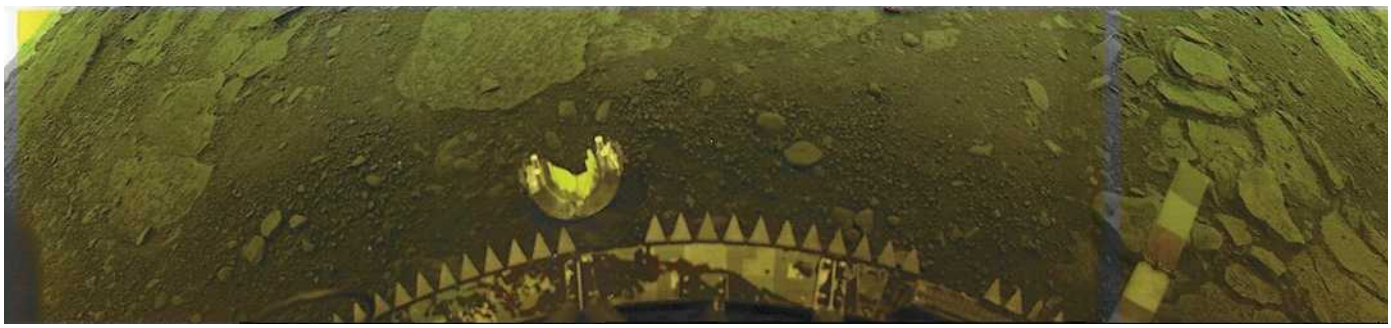
1. Kenyon, C. *Cell* **120,** 449–460 (2005).
2. Pinkston, J. M., Garigan, D., Hansen, M., Kenyon, C. *Science* **313,** 971–975 (2006).
3. Baur, J. *et al.* Nature **444,** 337–342.
4. Olsen, A., Vantipalli, M. C., Lithgow. G. J. *Science* **312,** 1381–1385 (2006).

©2007 Nature Publishing Group

G. BRAASCH/CORBIS

THE GIRL NEXT DOOR

D. P. MITCHELL



©2007 Nature Publishing Group

Hanging bright in the morning sky, Venus's allure is obvious; but its blasted surface looks too hot to handle. **Eric Hand** investigates the difficulties of returning to the closest planet — and new plans to reap the rewards of doing so.

The panorama is distorted and claustrophobic, overcast by a thunder-in-the-air gloom. Flat rocks fill the view, framed by the menacing, serrated teeth of the spacecraft and a patch of yellow sky. The lens cap lies discarded in the dust as if the cameraman had been mugged. Or melted.

It was 1982 when Venera 13, one of many Soviet missions to Venus, landed in this hellish place. The four cell-phone-quality panoramas it and its sister probe sent back (see opposite page, bottom) are still the only colour pictures of the surface that researchers have. As the morning or evening star, Venus is the brightest and loveliest of the planets, but as a subject for science it is barely an also-ran. The European Space Agency's Venus Express, launched in 2005, is the first mission to Venus since NASA's Magellan mission, launched in 1990, was brought to a fiery end in the planet's atmosphere in 1994. Between the two, Mars got all the love, with a mission almost every year. NASA's planetary database now contains about 25 terabytes of martian data; America's space-age Venus data total only 400 gigabytes — a tenth of the return from the two Mars rovers launched in 2003.

"Venus has been like the forgotten planet," says Håkan Svedhem, project scientist on the Venus Express team. "Mars has completely taken over."

There are various explanations for the neglect, but as Venus Express is showing (see articles starting on page 629) a dearth of beguiling questions is not one of them. The runaway greenhouse effect that boiled away the oceans the planet seems to have started off with draws a lot of attention and speculation. But scientists are also interested in the rocks below. Venus, unlike Mars but like Earth, retains enough internal heat to drive large-scale planetary processes — flows in the mantle and the crust. But the ways these processes have played out on Earth and Venus are vastly different — in part, perhaps, because of that vicious greenhouse effect and the absent oceans.

To address these questions, planetary scientists want a surface mission that picks up where the Venera probes left off. "You reach a point where you want to do more than fly-bys and orbiters," says David Grinspoon, astrobiology curator at the Denver Museum of Nature and Science in Colorado. "We need to push for these more comprehensive missions that explore the deep atmosphere and surface *in situ*." Now Grinspoon and other venusian diehards may be getting a few crucial rungs higher on NASA's to-do list. "From my perspective, we can't ignore Venus any longer," says NASA's director of planetary exploration James Green. "We're not going to ignore it. It's gathering momentum." Earlier this month, Green announced that NASA would consider a 'flagship' mission, its highest, multi-billion-dollar mission class. He has set up a team to study the science questions to be asked and the technology that would be needed.

One explanation for the neglect of Venus is simple. Human exploration is a stated priority for NASA; human exploration of the surface of Venus is next to inconceivable. Mars is by far the most plausible outpost beyond the Moon — and it has the added bonus of being considerably more likely than Venus to offer a discernible record of past life.

Extraterrestrial life and human exploration excite the public. And if that drives science policy, then so be it, says Steve Squyres, a planetary scientist at Cornell University in Ithaca, New York, and lead scientist for NASA's Mars rovers. "We are doing science with public money and lots of it. We have to keep in mind not just our own academic interests," he says. Although he says he is equally interested in both planets — as a young professor he worked on the Magellan mission to Venus — his academic career is now centred squarely on Mars. "To a certain extent, you have to go where the data are," he says.

> **"We can't ignore Venus any longer."**
> — James Green

Which leads to the other advantage of Mars. Scientists study Mars because they can. Mars orbiters can map its surface across the whole electromagnetic spectrum using radar, a wide range of infrared and visible wavelengths, ultraviolet and even γ-rays. So far, the only thorough mapping of Venus's shrouded surface has been by radar, first by some of the Venera probes, then by Magellan.

And Mars science isn't limited to orbiters. The thin atmosphere makes slowing down landers hard, but once you're down, life's not too difficult: Squyres's rovers have trundled along for years now. It's far easier for engineers to design instruments to survive the cold of Mars than it is to immerse a probe in Venus's crushing atmosphere and scalding temperatures.

That hasn't stopped Kevin Baines. "I've been trying since 1992 to get this country back to Venus," says Baines, a planetary scientist at the Jet Propulsion Lab in Pasadena, California, who believes that he holds the record for the greatest number of NASA-rejected Venus mission applications. His proposals in the 1990s included an infrared imager that would map the surface, at night, using spectral gaps in the atmospheric greenhouse to glimpse the hot surface below — a technique Venus Express is now using in the hopes of finding evidence of contemporary volcanism. Spotting active eruptions might kindle interest — Mars's volcanoes, though monumental, are thought to be almost entirely dormant — and might also help explain the chemical structure of the atmosphere. But so far, Venus Express has not found any lava hot spots. And an instrument that was supposed to locate tell-tale plumes of sulphur rising from volcanoes has failed to send back any useful data.

With Venus Express pursuing much of the science of his rejected orbiters, Baines has moved on to balloons — something the Soviet Union deployed on Venus more than 20 years ago with the Vega probes. In terms of ballooning, the charms of Venus and Mars are reversed. The hardly-there atmosphere on Mars is deeply inimical to the idea of floating lazily from place to place (which has not stopped such missions being suggested). The thick atmosphere on Venus, on the other hand, is ideal. A balloon 55 kilometres up in the venusian atmosphere would operate at comfortable temperatures and pressures, ferried around the world by ample winds.

Baines says his silvery-green balloon would be 7 metres across and five layers thick. With a coating of Mylar to reflect the Sun and a layer of Teflon to protect against sulphuric acid, a balloon could last months, perhaps a year, he says. A high priority for such a mission would be to measure the abundance of various isotopes of noble gases, which would allow scientists to deduce new details of the planet's history. Baines is also considering dropping 5-kilogram probes, like ballast, which would take pictures of the surface as they fell to a crash landing.

The probes might target specific spots on the surface. Most of the surface of Venus is basalt, and apparently rather young basalt at that — eruptive evidence of the heat within. The age is known because Magellan found very few large impact craters on the surface compared with the surfaces of Mars and the Moon. Fewer craters, other things being equal, mean younger crust. On Earth, large parts of the crust are kept young by plate tectonics. On Venus, though, there is no evidence of the faults that a surface made of shifting plates should have.

One proposed explanation for the oddity is that instead of producing plate tectonics as on Earth, convection currents in Venus's mantle knead the crust continuously, thickening and thinning it in a way that leaves it crater-free. Another, more dramatic, hypothesis is that every so often the crust gets so thin and cool and dense that it cannot support itself on the mantle. This 'catastrophic resurfacing' theory holds that, 750 million years ago, all the crust sank at

©2007 Nature Publishing Group

once, turning the entire planet into a sea of magma. "That's a good way to get rid of a lot of heat," says Steve Mackwell, director of the Lunar Planetary Institute in Houston, Texas, and a master of deadpan delivery.

Without new data, it is hard to decide between theories. Dropping probes into the right places might help — and might reveal clues to what the surface was like before. If, for example, they were to find some ancient granite, a rock that on Earth is produced by water-assisted recycling of the crust, they would be opening a window into a time when Venus was more like Earth, a time when it might have had continents, oceans and plate tectonics. "That would be a revelation," says Baines.

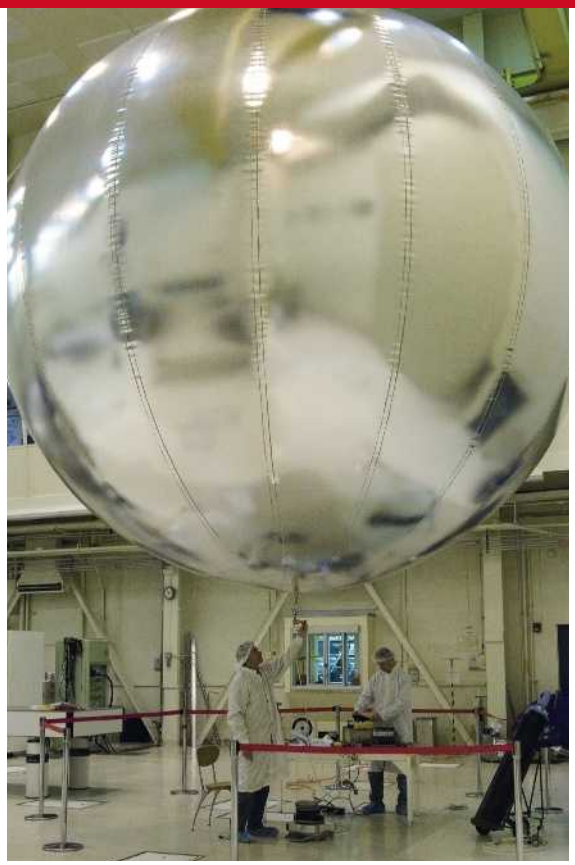
The nuclear fridge option

A lander or rover might also search out minerals or structures associated with now-closed chapters in the planet's history. Although tricky, a surface mission is not impossible, says Larry Esposito of the University of Colorado at Boulder. There are three choices: insulate your probe and do the science quickly (as the Venera probes did, and Baines's drop-probes would); use temperature-resistant electronics; or refrigerate the craft. The technologies of deep-sea submersibles that investigate black-smoker chimneys on Earth could translate well to venusian probes of the first sort, Esposito says, whereas heat-resistant electronics, which might be required for, say, a network of passive seismometers listening for quakes and eruptions, would be expensive to develop. A rover mission, or anything else intended to lead an active life of longer than a day, would need refrigeration, which requires a continuous power source. That means nuclear power of some sort. And that means a very expensive mission.

> "We put Venus up on this pedestal, and it disappointed us."
> — David Grinspoon

A nuclear rover might stretch the budget of even a flagship mission (the Mars Science Laboratory, a non-nuclear, non-refrigerated proposition, is currently budgeted at US$1.7 billion). A simple lander, though, or a balloon, might be cheap enough to fall within NASA's $700-million New Horizons mission category. And a bargain basement $425-million Discovery-class orbiter is still a possibility. Vesper, a Venus orbiter that would explore the carbon chemistry of the atmosphere, was proposed for the next round of Discovery missions by Gordon Chin, a planetary scientist at NASA's Goddard Space Flight Center in Greenbelt, Maryland. It has survived to the last stage of the selection process, but as Baines and Chin know from experience, a long history as bridesmaids has yet to see a Venus mission win its own special day.

With a flagship mission on the cards, might NASA be less likely than ever to approve a smaller Venus mission? Baines thinks otherwise.



NASA/JPL-CALTECH

Balloons would be oddly apt for studying Venus.

Venus orbiters and balloons could act as scouts for a major mission, he says. "If we can get those to work, people will be more comfortable spending the big money of a flagship mission."

Other nations may also quicken the pace of discovery. Although Europe has no firm plans to follow on from Venus Express — itself something of an afterthought in the wake of Mars Express — Japan has an orbiter slated for 2010. And Russia, which has not launched a successful planetary mission since the fall of the Soviet Union, has proposed a Venera-D mission that might include a lander with a launch window of 2016–18.

A farewell to fantasy

Engineering a viable Venus probe, and paying for it, is only part of the challenge. To justify a flagship mission, NASA's Green says it is up to scientists to ask compelling questions that demand new research. Beyond that, they need to engineer a new narrative, one that rallies other scientists and excites the public.

In pre-space-age science fiction, that wasn't a problem: Venus was a young, oceanic planet shrouded in cloud that acted as a counterpoint to the ancient clear-skied deserts of Mars. Sometimes it was a swampy home to dinosaurs, sometimes to frog people.

Mars, though, largely through chance, has managed to hold on to the narrative that has held sway since the days of Percival Lowell and his canals — the story of a once-watery planet transformed into desert, with vestiges of the life of its early wetter days possibly still preserved. Venus, on the other hand, lost its allure, its swamps and oceans evaporated by the evidence of prohibitive surface temperatures that emerged in the 1960s. (The dream lingered on — the Venera landers were designed to float if they landed in the sea.) "Venus suffers from not fulfilling our pre-space-age expectations," says Grinspoon. "We put Venus up on this pedestal. And it disappointed us."

In the past decade, an alternative narrative for Venus with obvious appeal has sprung up: Venus as a cautionary tale of greenhouse warming run amok. "Now, I think, the environment is a bigger concern," says Esposito. "It's causing people here on Earth to pay more attention to Venus." Chin goes so far as to say that the case for Venus exploration has been helped by Al Gore's Nobel prize for campaigning on global warming.

Gore himself might be wary of exaggerating the links too much. Earth's atmosphere is 0.04% carbon dioxide; that of Venus is 95%. If humans burned all of Earth's fossil fuel in one go, the atmosphere would still be only about 0.2% carbon dioxide, says James Kasting, a planetary scientist at Pennsylvania State University in University Park. Besides, Kasting says, models show that distance from the Sun is the critical factor in pushing a carbon-dioxide greenhouse over the edge into an ocean-evaporating runaway state. Only in a billion years, when the Sun is 10% hotter, will such a thing be likely on Earth, he says.

But perhaps Venus doesn't have to be of direct relevance to Earth's future to be exciting. Perhaps its very difference will be the key to its importance. For a long time the study of Earth-like planets has been limited to just Earth and its two nearest neighbours, siblings with life histories to compare and contrast. But those days are coming to an end. Current missions such as France's Corot, and future missions such as NASA's Kepler, due to launch in 2009, may discover a far greater range of Earth-sized planets outside the Solar System.

To understand the thin streams of data from those far-off worlds, an understanding of the full range of possibilities for an Earth-like planet's evolution will be important. "What if we find a couple dozen planets — and they're all dead planets?" asks Sara Seager, an astrophysicist at the Massachusetts Institute of Technology. "We're going to want to understand them." Studying Venus, she says, "might help us understand what leads to a habitable planet and what leads to a dead planet." In the context of the farthest planets ever studied, the story of the planet closest to hand may take on a new importance. ■

Eric Hand writes on physical sciences in *Nature*'s Washington DC office.

See also News & Views, page 617.

©2007 Nature Publishing Group

Each co-author should sign to reduce risk of fraud

SIR — Although I am in favour of all efforts to reduce scientific fraud, I do not feel that the proposal made in your Editorial 'Who is accountable?' (*Nature* **450,** 1; 2007) — namely, for researchers to sign a policy statement — will have any real effect.

Researchers must already navigate numerous formalities during the publication process. Signing such statements could soon become so routine that they would not give it a second thought. During investigations into scientific misconduct, these signatures would not constitute evidence of anything. We already have rules about integrity and conduct; disciplinary committees gain no leverage by pointing to additional signatures on additional declarations.

I support the push for declarations that require careful thought on the part of each co-author. Once each co-author is required to describe his or her contribution and to sign a final description of the division of labour, then it will be clear who is responsible for any particular aspect of the work. Researchers can limit their liability to fraud charges only by limiting their claim to the paper's achievements.

Only by such an active process can investigators use these declarations as meaningful statements. Given the additional benefit of recording each researcher's contribution to impact, this system might warrant the extra burden.

Kevin Yager
Polymers Division, National Institute of Standards and Technology, 100 Bureau Drive, MS 8541, Gaithersburg, Maryland 20899-8541, USA

Readers are welcome to comment at http://blogs.nature.com/nautilus/2007/10/accountability_of_co-authors.html

Moon is the best witness to early Solar System history

SIR — *Nature* usually recognizes emerging areas of scientific importance and adheres to high scientific and journalistic standards, so you are probably embarrassed by your amateurish and out-of-touch Editorial 'Because it's there' (*Nature* **450,** 2; 2007). You remark "Although there are interesting scientific questions about the Moon, few, if any, are of the first order." This insults planetary scientists who have been working hard to understand the early history of our Solar System.

The Moon is recognized as the best available witness to 4.5 billion years of Solar System history; the early record is clearer and more complete there than on any other planetary body. Its importance is well documented in the recent report published by the National Research Council of the National Academies, *The Scientific Context for Exploration of the Moon* (http://books.nap.edu/catalog.php?record_id=11954).

Carle M. Pieters, George A. Paulikas
National Research Council Committee on the Scientific Context for Exploration of the Moon, Washington DC, USA

Modifying theories is the fabric of scientific progress

SIR — Your News story 'Long-held theory is in danger of losing its nerve' (*Nature* **449,** 124–125; 2007) discussed recent technical criticisms of work by Henri Korn, Don Faber and colleagues that was published in the early 1980s. Korn and his co-authors have replied to this News story in Correspondence ('One-vesicle hypothesis has been extensively discussed' *Nature* **449,** 781; 2007).

I would like to add my view, as I find it disturbing not only that you publish suggestions of data fabrication, but also that you revive old allegations against Korn that have already been dismissed in 1989 and 2004.

Classical work by Bernard Katz and colleagues showed that neurotransmitter release is quantized at the neuromuscular junction. These early workers formulated a mathematical approach to investigate the statistics underlying neurotransmitter release, termed quantal analysis. There were two important parameters, the release probability, p, and n, which was interpreted as the number of available vesicles. Because p was small and n was larger, Poisson statistics seemed to be an adequate approximation to neurotransmitter release statistics. When quantal analysis was applied in the central nervous system, n did not seem to be large any more and binomial statistics thus emerged as the model of choice.

This posed a problem regarding the biological interpretation of the parameter n, as it was small in relation to the number of vesicles. Korn, Faber and colleagues, guided by experiment and modelling, proposed that n represents the number of synaptic contacts rather than the number of vesicles. This fundamental insight changed the way in which neuroscientists think about synaptic transmission.

Like every influential theory, this proposal has been extensively scrutinized and refined over the past quarter of a century. It turns out that the model parameters n and p have more complicated biological interpretations and are not stable over time or across synapses. In some systems, even the very concept of quantal transmission seems to be violated.

However, finding exceptions and modifications to scientific theories does not invalidate the theories or make them worthless. Rather, this is the very fabric of the scientific process. To instil a sense of scandal into this process degrades science and misleads the public about the scientific process.

Felix E. Schweizer
Department of Neurobiology, David Geffen School of Medicine, University of California at Los Angeles, 650 Charles E. Young Drive South, Los Angeles, California 90095-1763, USA

Good thinking should not be wasted on bad ideas

SIR — Your Editorial 'Watson's folly' (*Nature* **449,** 948; 2007) seemed to contradict itself. Appropriately, you censured Watson for his past and recent racist pronouncements, but with the same pen you chided organizations for cancelling his speaking engagements, scolding them because "science is about nothing if not openness and critical debate". But the issue at hand really has nothing to do with science. In the most forgiving light, it is social polemics, and in the worst it is simple racism.

Watson's actions may be viewed as folly by racists, but to moral individuals his words are an indecency. Not every social conjecture is a valid scientific hypothesis; and humanity should not waste good thinking on hateful, immoral ideas. Watson's admitted statements were not folly: they were reminders of the extensive obstacles that many peoples still face in their quest for equal opportunity in democratic societies.

The organizations that cancelled Watson's engagements are socially responsible. Those engagements were never planned as forums for debate of Watson's racist attitudes. Decisive action by those organizations prevented any perception that his assault on human dignity was permissible.

James L. Sherley
Programs in Regenerative Biology and Cancer, Boston Biomedical Research Institute, Watertown, Massachusetts 02472, USA

Readers are welcome to comment at http://blogs.nature.com/news/thegreatbeyond/2007/10/james_watsons_race_row.html.

Contributions to Correspondence may be submitted to correspondence@nature.com. They should be no longer than 300 words, and ideally shorter. Published contributions are edited.

©2007 Nature Publishing Group

BOOKS & ARTS

The great debate

Today's take on the stem-cell field could recede rapidly in tomorrow's rear-view mirror.

Fundamentals of the Stem Cell Debate: The Scientific, Religious, Ethical and Political Issues
edited by Kristen Renwick Monroe, Ronald B. Miller and Jerome Tobis
University of California: 2007. 226 pp. $50.00, £29.95 (hbk); $19.95, £11.95 (pbk)

STR/AFP/GETTY

Surgeons collect stem cells in Thailand, where experimental treatments are unregulated.

Christopher Thomas Scott

How best to portray stem-cell research, an area of biology that is on a tear? Add to the science a swarm of disciplines busily dissecting its impact — law, ethics, policy and commerce. Then, if you aren't susceptible to searing headaches, go global. Fortunately, there are brave writers out there confronting the challenge, supplying a steady stream of books that grapple with all of this. The difficulty is discovering the titles that make the best sense of it.

The editors of *Fundamentals of the Stem Cell Debate* may have had this in mind when compiling their volume. And who better to invite as contributors than a group of scientists and academics who do it for a living?

I never start a book anywhere but at the beginning, but I suggest that readers start in the middle of this one, with Lawrence Goldstein's chapter 5. Goldstein, director of the stem-cell research programme at the University of California, San Diego, hits the right notes with his concise history, clear technical explanation and straight talk about science and politics. He emphasizes that we don't settle policy disagreements by allowing governments to do some things and not others — no governing would get done. It's the same for science: if we fund just the research that some find acceptable, no research will get done.

Fundamentals of the Stem Cell Debate is a "comprehensive overview for a wide audience", say the publishers. Many of the surveys, though, skirt a deep treatment of the science, probably because it is fluid territory. For example, more could be said on whether adult stem cells can differentiate into different tissues and, if they do, what it might mean for future therapies. Peter J. Bryant and Philip H. Schwartz have a crack at it in the first two chapters. Thick with citations and written like review articles, these could stymie casual readers. But the authors do a solid job of summarizing the field, and the references are valuable for further scholarship.

The book breaks new ground by briefly mentioning the ethics of clinical trials. Therapeutic applications of stem cells are the next set of big issues for bioethics. The evaluation of preclinical results, safety, informed consent, choosing human subjects and oversight are a few of the topics now appearing in journals. No overview would be complete without another argument about the moral status of the embryo, and there are chapters on ethics and differing religious perspectives.

The chapter on religion is a well-written summary of the principal positions on embryonic stem-cell research around the world, but I wish authors would attempt to place the US religious framework into an international context. On some cases — and I suppose it's true here — the positions of Eastern religions are still emerging, so there is scant literature to review.

This book joins a constellation of releases that examine stem cells through different lenses. Readers should pick a few titles to get the full picture of how this formerly quiet corner of biology has exploded into new realms. For an account of the controversies surrounding patents and law, read Russell Korobkin's new book, *Stem Cell Century* (2007). If you hanker for activism, try Eve Herold's *Stem Cell Wars* (2006), an insider account of politics and law-making. New books on stem-cell ethics are overdue, but older compilations such as *The Human Embryonic Stem Cell Debate* (2001) and *God and the Embryo* (2000) offer lively debates from a different perspective. Popular science has a few to choose from: I like Ann Parson's *The Proteus Effect* (2004). Another, Michael West's *The Immortal Cell* (2003), prefaces each chapter with a bible verse.

It is odd we have a profusion of general books about stem cells but still no introductory textbook. Likely reasons are that the field hasn't yet coalesced and the subject cuts across the tried-and-true (and profitable) developmental and cell biology texts, now lumbering into twilight editions. A good introduction should include chapters on ethics, law and society — a hybrid that may not fit neatly into a marketing niche.

In the meantime, readings from *Fundamentals of the Stem Cell Debate* could be listed on an introductory curriculum, together with seminal research papers and chapters from foundational texts such as the fifth edition of Bruce Alberts' *Molecular Biology of the Cell* (2007), and the eighth edition of Scott Gilbert's *Developmental Biology* (2006).

The challenge for the serious reader is to keep up with all these breathtakingly rapid developments. ■

Christopher Thomas Scott is director of the Stanford University Program on Stem Cells in Society, 701 Welch Road, Suite A1105, Palo Alto, California 94304. He is author of *Stem Cell Now: A Brief Introduction to the Coming Medical Revolution*.

©2007 Nature Publishing Group

NATURE|Vol 450|29 November 2007

Is artificial life moving any closer?

Genesis Redux
edited by Jessica Riskin
University of Chicago Press: 2007. 389 pp. $25, £16

Beyond Human
by Gregory Benford and Elisabeth Malartre
Tor/Forge: 2007. 272 pp. $24.95

Greg Bear

Three vinyl toys stand on my office bookshelf: the Golem, Frankenstein's monster and the Terminator — a reminder that automatons of the imagination have been with us for a long time. The synthesis of life in both the past and future is discussed in two new books that open a window on to artificial life, enhanced life and robots. *Genesis Redux* gathers together papers from a rousing academic conference held in Stanford in 2003, and *Beyond Human* offers a critical but enthusiastic view from a physics and biological perspective.

The strength of *Genesis Redux* lies in its scholarship and range of topics. Clockworks, mechanical toys and their influence on biological concepts are presented in fascinating detail. Joan Landes introduces us to the Hoffmanesque works of Jacques de Vaucanson's feminine flautist and (excreting) duck, and to the flayed, preserved and posed cadavers, the *écorchés*, of Jean-Honoré Fragonard: there is a dancing fetus and a very naked man staring in horror, jawbone in hand. Landes delivers a lively analysis of our reactions to the abject and uncanny, the *frisson* so beloved by fans of Dr Frankenstein.

Philosophers and scientists of the day — before the younger Darwin — were already deeply concerned with the analogies that could be drawn between mechanism and creature. Were they useful, or misleading? The question remains in play. And life and thought are difficult to simulate (as Drosselmeyer, the toy-maker and magician from *The Nutcracker*, would testify) — but even harder to implement.

Less obviously connected with artificial life is an essay by Elliott Sober on sex ratios in human populations, and whether they point to a benevolent creator or to mere chance. Therein lies a profound question: if life is the result of a creative act, then is all life 'artificial'? What if there is, as godless evolution surely implies, self-creation? Are we not then our own artefacts, mechanisms of joy and magic — and does that wrap us in another illusion, like the beauty of a clockwork flute-player?

Norton Wise discusses the masculine and feminine Victorian categorization of men as prime mover — 'engine' — and women as 'mechanism'. And yet no mention is made of Mary Shelley, the most influential engine of artificial life in history. Likewise, in most of these discussions a paucity of references to key literature prevails. H. G. Wells's *The Time Machine* is cogently discussed, but not *The Island of Doctor Moreau*, *The New Accelerator* or *The Food of the Gods*. Completely neglected is Robert A. Heinlein, creator of Mike, the first computer graphic artificial intelligence, in *The Moon Is a Harsh Mistress* — though many of Mike's children peer out of its illustrations.

Bernadette Bensaude-Vincent gives a brief history of nanotechnology and its key analogies, drawn from molecular biology. In 1961, Richard Feynman was intrigued by the possibility of increasingly tiny assemblers. Eric Drexler hopes to emulate organic molecules in more enduring form. But the present bull market for nanotech seems to have devolved into innovations in materials science, rather than a quest to create tiny self-replicating machines. The same thing happened with artificial intelligence and neural network theory — high aspirations became mired in frustration, then subsumed by lower-level research and industrial buzzwords, lofty goals redefined.

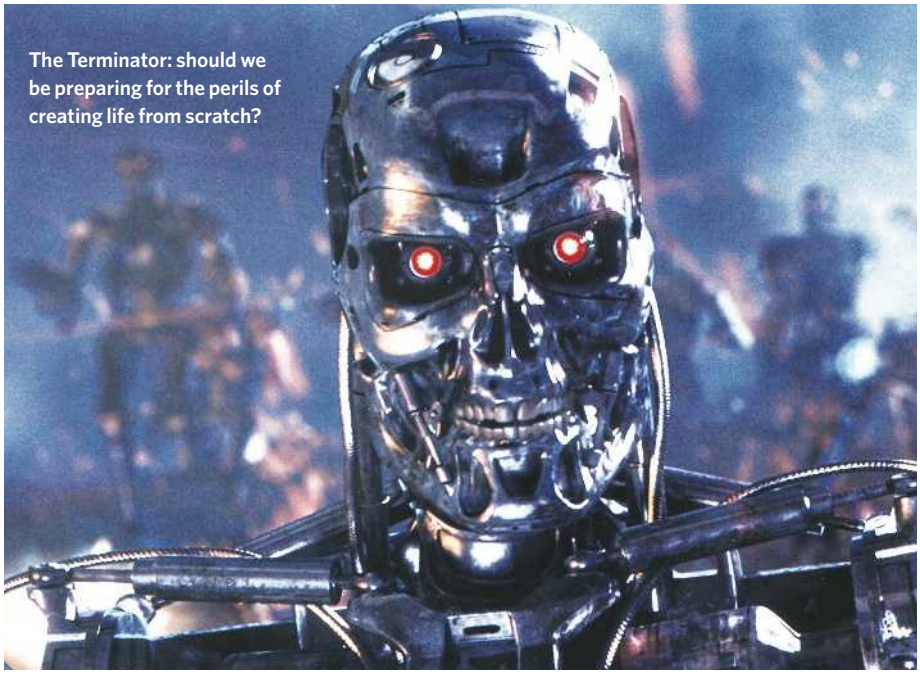
Genesis Redux takes the time to shed light on areas I would not naturally consider, and thus enlightens and expands the topic. Its cautious perspective — the enthusiasms of the past considered in the sober light of history — provides a useful counterpoint to our next book.

In *Beyond Human,* Gregory Benford and Elisabeth Malartre (a husband-and-wife team) take a more personal, promotional approach, and deftly give credit to the actual creators and thinkers: scientists, engineers, philosophers, writers and film-makers. Interviews with pioneers and participants in enhancement research, robotics and engineering, and informed perspectives on the theory, economics and actuality of life extension, give *Beyond Human* the flavour of a tourist guide to the future, conducted by natives, rather than a missionary's critical, hands-off examination.

The authors detail the experiences of our true cyborgs — disabled people whose high-tech prosthetics may make them more competitive than their able peers. They try to answer technical questions about the smallest size attainable for nanobot fingers and the practical limits of very small manipulators, and by implication, tiny intelligence devices. Their limiting answers could stimulate and irritate researchers into devising clever workarounds.

As biology and synthetic biology converge, and true artificial life becomes more and more feasible, the power for change will exceed that of nuclear energy. The fire of life is more subtle and pervasive and, through mass production and the internet, is being made available not just to well-funded government institutions but to private individuals. Training and research at secondary school will be essential for the biotech future, but could put dangerous tools into the hands of those whose aim is mass destruction — be they terrorists or domestic bigots. Institutions and journals should begin planning the necessary precautions; public discussion and professional prudence are essential. One terrorist incident could shut down basic research for years.

Together, these books serve up a feast with many side-dishes. The future is not behind us. Full of danger and potential, it is still unexplored territory, waiting to be rediscovered after our early encounters in nightmares and daydreams. ■



The Terminator: should we be preparing for the perils of creating life from scratch?

RONALD GRANT ARCHIVE

Greg Bear is the author of *Quantico*, *Darwin's Radio* and *Queen of Angels*. He can be contacted at www.gregbear.com.

©2007 Nature Publishing Group

EXHIBITION

The art of arthropods

Nick Thomas

A Californian entomologist uses insects as living paintbrushes to create abstract art. After loading water-based, non-toxic paints on to the tarsi and abdomens of insects, Steven Kutcher directs his bugs to create their 'masterpieces'.

Kutcher controls the direction and movement of his arthropods — such as hissing cockroaches (pictured), darkling beetles and grasshoppers — by their response to external lighting. The result is controlled and random movements, created in a co-authorship between the artist — with predetermined ideas about colour, form, shape and creative flexibility — and his living brushes.

Kutcher's art is more than just a novelty, because it reveals the hidden world of insect footprints. "When an insect walks on your hand, you may feel the legs move but nothing visible remains, only a sensation," he says. "These works of art render the insect tracks and routes visible, producing a visually pleasing piece."

An insect-lover from childhood, Kutcher has a master's degree in entomology and has taught biological sciences at various US colleges. Since the 1970s, he has worked as a 'bug wrangler' on some 500 movies, TV shows and advertisements, where he also used hair dryers, electric tape, and chemical repellents and attractants to control insect movement. He manipulated the tiny *Steatoda grossa* spider (painted blue and red) that nipped actor Tobey Maguire in *Spiderman*.

The idea for Kutcher's bug art originated in 1985, when he was hired to create fly footprints by making a fly walk through ink for an advertisement for Steven Spielberg's television series *Amazing Stories*.

J. ALCORN

This unique artist–arthropod partnership has so far yielded over a hundred works, typically characterized by vibrant, eye-catching colours and designs, splattered with trailing dots and dashes (see www.BugArtbySteven.com). Kutcher is now gathering pieces to form a travelling exhibit for art and natural history museums throughout the United States.

"I hope people will look at these works and see the duality of art and science," he says. "Each insect is writing a page in its life, and every painting is a new discovery." ■

Nick Thomas is associate professor of chemistry at Auburn University, Montgomery, Alabama 36124, USA.

Kutcher's bug art is on display at the Entomological Society of America meeting in San Diego (9–12 December) and at the Lancaster Museum (15 December–13 January 2008), in California.

On a molecular mission

Elizabeth Blackburn and the Story of Telomeres: Deciphering the Ends of DNA
by Catherine Brady
MIT Press: 2007. 424 pp. $29.95/£19.95

Maria A. Blasco

"I want to understand how living things work," declared a young Liz Blackburn to Frank Hird, her supervisor at the University of Melbourne, when asked why she wished to pursue a scientific career. Back in the 1960s, Blackburn could not have imagined that she would later be the main player in two fundamental discoveries in biology: the molecular nature of the ends of chromosomes, or telomeres, and the identification of the enzyme telomerase.

Catherine Brady's biography is a page-turner from the first chapter, weaving together the heroine's personality with her success as a scientist. We learn about Blackburn's family and her first tentative steps in the science world that eventually led to the discovery of telomerase in the mid-1980s, and about her determination, her curiosity, her way of dealing with situations and her opinions on the peer-reviewing process.

In highlighting the factors that shaped Blackburn's career, we follow her incursions into policy-making and science ethics: first as president of the American Society of Cell Biology (ASCB) and as the chair of the Department of Microbiology and Immunology at the University of California in San Francisco (UCSF), and then as part of the Bioethics Advisory Council to President George W. Bush — from which she was dismissed for her views on stem-cell policy.

E. FALL

Arm arrangement: Liz Blackburn, discoverer of telomeres and telomerase, in her laboratory.

Blackburn has been an inspiration to those of us who started out in the field of telomeres. The book conveys a vivid impression of her that matches a personal encounter. Her equable temperament does not prevent her from having strong views, and she emerges as a valuable role model in the sometimes unsettling treatment of women in the world of science.

Born the second daughter of seven children in Tasmania, Australia, to a family of professional scientists (her parents were medical practitioners and her grandfather and great-grandfather were geologists in China). An early interest in chemistry and biochemistry propelled her to Hird's lab for her doctorate, which matured her forceful scientific mind and reaffirmed her interest in science as a *modus vivendi*. Then Blackburn went on to what at the time was the Olympus of molecular biology, the Medical Research Council (MRC) laboratory in Cambridge, a place packed with past and future Nobel laureates that would become the gold standard for today's top scientific institutions.

The MRC laboratory was hosting a revolution in molecular biology, powered by discoveries about cellular mechanisms fundamental to life. Fred Sanger's DNA-sequencing work particularly attracted Blackburn, and from him she learned her pragmatic approach to science. As Brady points out, the heady ambience of the laboratory was marred by some male chauvinism (with Watson and Perutz receiving special mention). Blackburn discovered how things

©2007 Nature Publishing Group

were in top scientific institutions: extreme dedication and long working hours, with no supporting hierarchies — what Brady calls a "rat lab". Fellow scientists became her family substitutes and friends, and there she met her future husband, John Sedat.

Blackburn's DNA-sequencing skills were for her the key to discovery, and she took them to Joe Gall's lab in Yale after a short break to climb to Mount Everest's base camp with Sedat. In Gall's lab were some of the future principals of the telomere field — Ginger Zakian, Mary-Lou Pardue and, later, Tom Cech. It was there that Blackburn discovered in 1976 that *Tetrahymena* chromosomes end in a series of repeated runs of cytosine bases that varied in length.

Although this was the first molecular insight into the structure of chromosome ends, it was not seen as important by the community, which is surprising in view of its implications for chromosome replication and transmission of genetic information. Blackburn blames the perception of *Tetrahymena* as a "freak organism". But it also fell outside what was then mainstream molecular biology. The story repeated itself when she, together with Carol Greider, discovered telomerase in 1985. By then it was clear that telomere replication was a fundamentally important process, but telomerase continued to receive scant attention until 1994–95, when it was shown to be aberrantly activated in most human cancers.

The biography succeeds in capturing Blackburn's vision, which has encouraged her to pursue unbeaten tracks to make discoveries that today hold therapeutic promise for both cancer and ageing. ■

Maria A. Blasco is head of the Telomeres and Telomerase Group in the Molecular Oncology Program at the Spanish National Cancer Centre (CNIO), Melchor Fernández Almagro 3, 28029 Madrid, Spain.

Is technology unnatural?

The Artificial and the Natural: An Evolving Polarity
edited by Bernadette Bensaude-Vincent and William R. Newman
MIT Press: 2007. 331pp. $40

Philip Ball

The topic of this book — how boundaries are drawn between the natural and the synthetic — has received too little serious attention, both in science and in society. Chemists are justifiably touchy about descriptions of commercial products as 'chemical-free', but the usual response, which is to lament media or public ignorance, fails to recognize the complex history and sociology that lies behind preconceptions about chemical artefacts. The issue is much broader, however, touching on areas ranging from stem-cell therapy and assisted conception to biomimetic engineering, synthetic biology, machine intelligence and ecosystem management.

And it is not an issue for the sciences alone. Arguably, the distinction between nature and artifice is equally fraught in what we now call the fine arts — where again it tends to be sidestepped. Some modern artists address the matter head on with their interventions in nature — for example, the artificial rainbows of Andy Goldsworthy — but much popular art criticism now imposes a contemporary view, even on the old masters. Through this lens, Renaissance writer Giorgio Vasari's astonishment that Leonardo's painted dewdrops "looked more convincing than the real thing" seems a little childish, as though he has missed the point of art. No one today believes that the artist's job is to mimic nature as accurately as possible. Perhaps with good reason, but it is left to art historians to point out that there is nothing absolute about this view.

At the heart of the matter is the fact that 'art' has not always meant what it does today. Until the late Enlightenment, it simply referred to anything human-made, whether a sculpture or an engine. The panoply of mutated creatures described in Francis Bacon's *The New Atlantis* (1627) were the products of 'art', and so were the metals generated in the alchemist's laboratory. The equivalent word in ancient Greece was *technē*, the root of 'technology' of course, but in itself a term that embraced subtle shades of meaning, examined here in ancient medicine by Heinrich von Staden and in mechanics by Francis Wolff.

The critical issue was how this 'art' was related to 'nature', roughly identified with what Aristotle called *physis*. Can art produce things identical to those in nature, or only superficial imitations of them? (The latter belief left Plato rather dismissive of the visual arts.) Does art operate using the same principles as nature, or does it violate them? Alchemy was commonly deemed to work simply by speeding up natural processes: metals ripened into gold sooner in the crucible than they did in the ground, and (al)chemical medicines accelerated natural healing. And although some considered 'artificial' things to be inferior to their 'natural' equivalents, it was also widely held that art could exceed nature, bringing objects to a greater state of perfection, as Roger Bacon believed of alchemical gold.

Oil painting by Jan van Huysum (1682–1749): does art imitate nature, or improve on it?

MUSEE FABRE, MONTPELLIER, FRANCE/GIRAUDON/THE BRIDGEMAN ART LIBRARY

The emphasis in *The Artificial and the Natural* is historical, ranging from Hippocrates to nylon. These motley essays are full of wonders and insights, but are ultimately frustrating in their microcosmic way. There is no real synthesis on offer, no vision of how attitudes have evolved and fragmented. There are too many conspicuous absences (Leonardo da Vinci for one) for the book to represent an overview.

It would have been nice to see some analysis of changing ideas about experimentation, the adoption of which was surely hindered by Aristotle's doubts that 'art' (and thus laboratory manipulation) was capable of illuminating nature. Prejudices about experiments often went further: even in the Renaissance, one was free to disregard their results if they conflicted with *a priori* 'truths' gleaned from nature, rather as Pythagoras advocated studying music by "setting aside the judgement of the ears". And it would have been fascinating to see how these issues were discussed in other cultures, particularly in technologically precocious China.

But most important, the discussion sorely lacks a contemporary perspective, except for Bernadette Bensaude-Vincent's chapter on plastics and biomimetics. This debate is no historical curiosity, but urgently needs airing today. Legislation on trans-species embryology, reproductive technology, genome engineering and environmental protection is being drawn up, based on what sometimes seems to be little more than a handful of received wisdoms (some of them scriptural) moderated by conventional risk analysis. There is, with the possible exception of discussions on biodiversity, almost no conceptual framework to act as a support and guide.

All too often, what is considered 'natural' assumes an absurdly idealized view of nature that owes more to the delusions of Rousseau's romanticism than to any historically informed perspective. By revealing how sophisticated, and yet how transitory, the distinctions have been in the past, this book is an appealingly erudite invitation to begin the conversation. ■

Philip Ball is a consultant editor for *Nature*. His most recent book is *The Devil's Doctor* (Heinemann/FSG, 2006).

©2007 Nature Publishing Group

The ethical regulation of science

Occasionally science makes procedures possible that are so radical that those at the interface between science and politics are called on to define moral standards for society.

Mary Warnock

When in 1978 the first baby was born by *in vitro* fertilization (IVF) it was inevitable that there would be calls for the procedure to be prohibited. That science develops too fast for morality had become the cliché of the twentieth century. Wisely, the UK government decided to set up a committee from which to seek advice before legislating on such a complex and emotive issue.

The Committee of Enquiry into Human Fertilisation and Embryology was founded to examine the social and ethical implications of the new techniques. Therefore the committee could not be made up entirely of physicians and scientists. With some difficulty, 16 people — including me as the chair — were gathered to look at the problem from all angles. Our areas of expertise included social work, law and theology.

We were not a group of 'moral experts', with particular moral authority derived from our expertise. Rather, our entitlement to propose legislation derived from the fact that we had been set up by government and that we had been given the time and resources to do so. The only other requirement was that we should all be capable of formulating and listening to arguments.

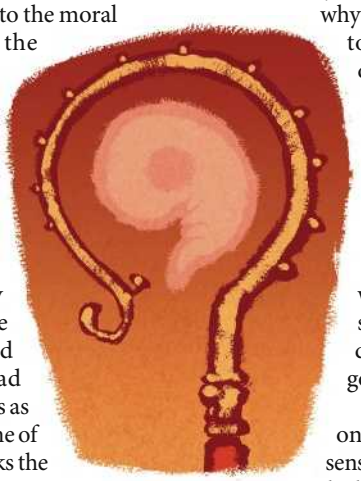
The central and most controversial issue before us was whether or not research using live embryos should be permitted. There was little possibility of a moral consensus. If research were prohibited, IVF could not continue. It would have been too risky for patients.

When legislation seemed imminent in Britain, the Catholic Church published an instruction condemning IVF and research using human embryos. The Church stated that its instruction was based on "the criteria of moral judgement as regards the application of scientific research and technology, especially in relation to human life and its beginnings".

The Church claimed a right to regulate science in this area, because of its superior knowledge of morality. In sharp contrast, the committee's entitlement to issue moral advice to ministers derived from its having been set up to do so, and from its having a wide and non-partisan membership.

Prohibition of IVF did not seem to the majority of the committee to be a serious option, given its widespread welcome as an innovative remedy for infertility. We all regarded infertility as a serious malfunction, causing much distress. Instead, we proposed a strict system of licensing, backed up by the criminal law. Regulation was not a mere sop to science-phobia. There was a real danger that women, desperate to conceive, might be exploited, taken in by unrealistic promises and charged extortionate fees for futile or dangerous treatment.

Establishing what limits should be placed on embryonic research entailed a decision by the committee as to the moral and legal status of the live human embryo *in vitro*. Those who opposed the use of embryos in research could seek to demonstrate that it was morally wrong only by answering in advance the very question we were asking. They deemed that the embryo had the same moral status as any human being. One of the most difficult tasks the committee faced was to get parliament to understand that the status of the embryo *in vitro* was a matter not of science but of moral decision. The novelty of the embryo *in vitro* meant that there could be no appeal to precedent or existing moral convention or to religious laws.



D. PARKINS

We recommended that research using early embryos might continue. But the law demands certainty. We therefore recommended that it should be a criminal offence to keep an embryo alive in the laboratory more than 14 days after fertilization. The 14-day rule was criticized by those who sought to prohibit the research altogether and by those who complained that it was arbitrary and would not be observed.

It was not, however, entirely arbitrary. Important changes occur in the embryo at about 14 days from fertilization, after which cells are differentiated, no twinning can take place and the first signs of what will be the spinal cord appear. Such considerations weighed with us. The point was to have some legislative barrier, about which there could be no dispute. As such it has survived.

The legislation that followed the committee's report included the 14-day rule, as well as the establishment of the Human Fertilisation and Embryology Authority (HFEA) that we recommended should be in charge of licensing and monitoring research and clinical procedures.

The moral decisions that such committees have to make are essentially matters of public not private morality. We had to consider our own moral or religious scruples (which would obviously influence our thinking), alongside what the consequences might be of the decisions for society as a whole. This was the reason why we could not allow ourselves to be swayed by arguments derived from a particular religious dogma. The moral standards of our society in Britain, however deeply influenced by the Judaeo-Christian and Muslim traditions, are not intrinsically religious. As a committee we were given the task of setting such a standard. This must be done by weighing up possible goods against possible harms. These harms do not include only the offending of religious sensibilities of a particular group. The legislation would govern everyone — believers and atheists — and had to take into account wider considerations, such as the relief of suffering. Nor was it to be other than permissive. No one would be compelled to seek a form of infertility treatment or engage in a form of research of which his religion taught him he should disapprove. Above all, the harm that the legislation should seek to minimize was the exploitation of the vulnerable and ignorant. This was the task of the HFEA.

One may generalize from the case of IVF to other cases where it is feared that dedicated and ambitious scientists and doctors may pursue research that some members of society find repugnant. Examples include embryonic stem-cell research, therapeutic cloning and the construction of mixed-species embryos for research purposes. But it is essential that ignorance and prejudice should not be allowed to dictate the outcome. Everyone should be educated so as to have a broad understanding of science, and an appreciation of its potential for good. Without this, we cannot responsibly erect barriers to scientific advance. ■

Mary Warnock is a philosopher and an independent crossbench member of the House of Lords, London, SW1A 0PW, UK.

For more essays and information see http://nature.com/nature/focus/scipol/index.html.

©2007 Nature Publishing Group

NEWS & VIEWS

VENUS

Express dispatches

Andrew P. Ingersoll

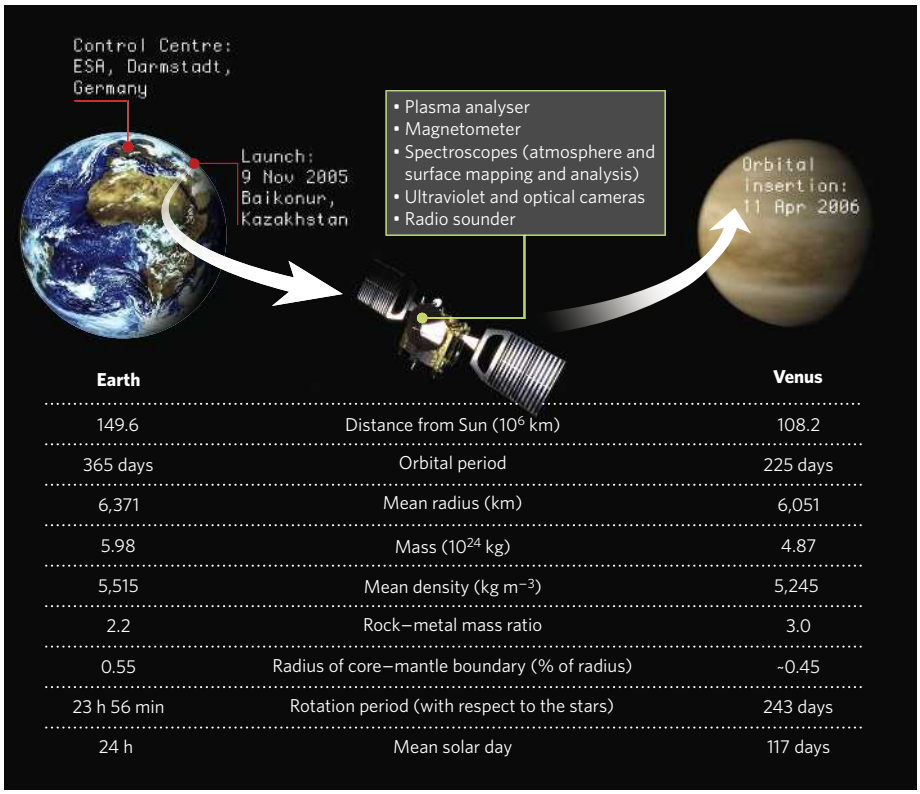
The Venus Express mission has returned its first findings on the harsh atmosphere of our sister planet. It's another step towards explaining how Venus turned out so differently from our balmy home.

Venus is the planet closest to Earth not just in distance, but also in mass, radius, density and chemical composition (Fig. 1). Yet where Earth's benign climate has fostered life, Venus is a hellish place — its surface hot and dry, its crushing atmosphere made up of carbon dioxide permeated by clouds of sulphuric acid. How did it all go wrong?

The European Space Agency's Venus Express probe was launched in November 2005 to address some big issues associated with this question. Where is Venus's water, and what is its history? What is the weather like on Venus, and is it different from that on Earth? What does the precise composition of Venus's atmosphere tell us about its evolution? The first results from Venus Express are presented in eight papers elsewhere in this issue[1–8].

We start with water. The most abundant volatile compound on Earth (where the oceans are 300 times more massive than the atmosphere), water is present only in tiny amounts on Venus, all of it in the atmosphere — about 200 parts per million, either as vapour or dissolved in its sulphuric-acid clouds. There are no oceans: Venus's surface temperature of about 730 kelvin (457 °C) means that water could not exist as a liquid there, even if it were abundant. Two other signature substances, nitrogen and CO_2, are present in about the same amount on both planets. The latter admittedly appears in different places: on Venus, it makes up 96.5% of the massive atmosphere, where it creates a surface pressure some 92 times that of Earth's; on Earth, it turns up in the extensive limestone deposits that precipitated from CO_2 dissolved in the water of the oceans.

Because Venus is close to Earth in so many ways, it seems likely that the two planets started out similarly. Venus must once have had an ocean's worth of water, but lost it somehow. Water vapour is a heat-trapping greenhouse gas, whose atmospheric abundance is controlled by evaporation from the oceans. If Earth were to move closer to the Sun, the temperature of its oceans would rise and more water would evaporate, increasing the temperature still further. At a certain point it would become a runaway greenhouse, with the oceans boiled off and all the water residing in the atmosphere[9].



Earth		Venus
149.6	Distance from Sun (10^6 km)	108.2
365 days	Orbital period	225 days
6,371	Mean radius (km)	6,051
5.98	Mass (10^{24} kg)	4.87
5,515	Mean density (kg m^{-3})	5,245
2.2	Rock–metal mass ratio	3.0
0.55	Radius of core–mantle boundary (% of radius)	~0.45
23 h 56 min	Rotation period (with respect to the stars)	243 days
24 h	Mean solar day	117 days

IMAGES: ESA

Figure 1 | Voyage to Venus — the essential facts and figures of the European Space Agency's Venus Express mission.

This could well have been what happened on Venus. The planet might have started out in its present orbit with a massive atmosphere made of water vapour, which sunlight then split into hydrogen and oxygen. Hydrogen, being lighter, escaped into space, whereas the heavier oxygen hung around and oxidized the crust to create the planet's hot, dry surface. Conventional hydrogen is lost faster than its heavier isotope deuterium, and evidence for this huge loss of water comes from the high ratio of deuterium to hydrogen on Venus, which is 100–150 times the value on Earth. The first measurement of this ratio[10] was based on droplets of sulphuric acid that clogged the inlet leak of the mass spectrometer aboard NASA's Pioneer mission to Venus, launched in 1978. Instruments on Venus Express have now confirmed and refined the high value[1].

The escape mechanism operating today probably involves the solar wind — the stream of charged particles emanating from the Sun — which strips atoms and ions out of the atmosphere. But the details are contradictory. Venus has no appreciable internal magnetic field, so the solar wind is deflected by magnetic fields induced in the planet's atmosphere. But whereas results from Venus Express's magnetometer indicate[2] that the solar wind does not enter the atmosphere, and so cannot affect its evolution, the mission's plasma analyser shows[3] that positively charged hydrogen and oxygen ions are escaping in the planet's wake at higher rates than are neutral atoms. The net result is that hydrogen and oxygen are being lost in the same 2:1 proportions as they are found in the water molecule[3]. The absolute rates still have to be worked out; as far as plotting Venus's

©2007 Nature Publishing Group

NATURE|Vol 450|29 November 2007

history is concerned, the escape mechanism could well have been different in the past, when water was much more abundant on the planet.

The extreme weather found in Venus's remaining CO_2-dominated atmosphere offers us several contrasts with Earth. The greater mass of Venus's atmosphere should store heat and distribute it more effectively than occurs on Earth, and temperature differences should therefore be much smaller. Results from the Pioneer probe[11] were consistent with this idea. Venus Express's programme of radio occultations — in which the spacecraft passes a radio signal tangentially through the atmosphere to measure its temperature — set out to test it further. The surprising result[4] is that the day–night temperature difference is large — between 30 and 40 K at the altitude range 55–60 km, within Venus's clouds (Fig. 2). Heat stored from sunlight is not enough to raise the temperature by this amount during the day. Enormous downdrafts might heat the air by compression, but they would have to be swifter than any global-scale downdrafts on Earth. A further possibility is that clouds and gas in the atmosphere partially absorbed the radio signal, inducing an error in the derived temperatures[4].

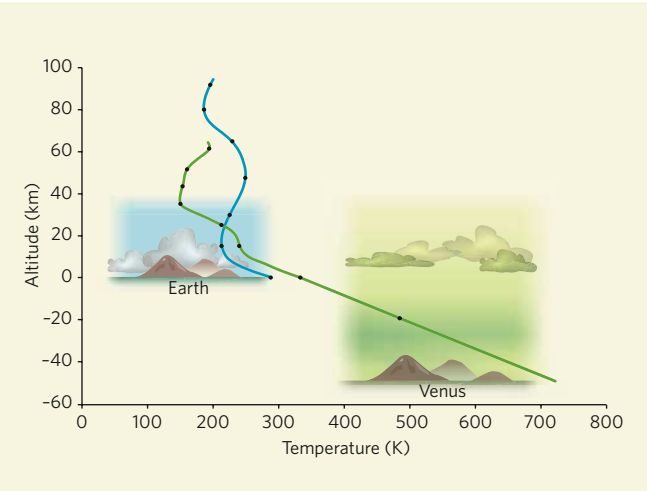


Figure 2 | Ocean of air. Venus's crushing atmosphere is composed of 96.5% carbon dioxide and 3.5% nitrogen; Earth's more clement envelope comprises roughly 79% nitrogen, 20% oxygen and 1% argon (dry air), with significant traces of other gases, chief among them carbon dioxide. (Moist air contains a few per cent of water vapour.) In this comparison of the two atmospheres' vertical structures, the zero of altitude of the temperature profiles (solid lines: blue, Earth; green, Venus) is fixed where the pressure is equal to Earth's sea-level pressure, 1 bar. Black dots on the temperature profiles indicate where the pressure has increased or decreased by a factor of 10; Venus's lower atmosphere extends almost 50 km below the 1 bar level, such that at the surface the pressure is 92 bars. The temperature there averages 730 K, compared with Earth's average surface temperature of around 288 K (15 °C). Relative to the 1-bar level, clouds in both atmospheres — of water on Earth, of sulphuric acid on Venus — occupy roughly the same altitude range, and the lower atmosphere of Venus is relatively cloud-free. The tallest mountains rise 12 km above the average surface level on Venus and 9 km above sea level on Earth.

The higher one goes in Venus's atmosphere, the faster the winds blow, and the smaller the day–night temperature differences become. Wind speeds are measured relative to the solid planet, and Venus is a slow rotator: it spins once relative to the stars every 243 Earth days, and, viewed from a point on the surface, the Sun rises, and sets, every 117 Earth days. (Venus's solar day is considerably shorter because, uniquely in the Solar System, it spins slowly backwards relative to its orbital motion. How it got into this state and how it stays there are unanswered questions[12], although tides raised by the Sun in the massive atmosphere probably have something to do with it.)

At the tops of the clouds, an altitude of nearly 70 km, the wind speeds approach 100 metres per second (360 km h^{-1}) — three times hurricane force and 2.5 times the speed of the jet streams that can help or hinder aeroplane pilots 11 km up in Earth's atmosphere. Venus Express confirmed these high winds by tracking clouds, made visible by an unknown absorber of ultraviolet light, that varied dramatically from day to day[5].

One feature that Venus shares with Earth and many other planets is the presence of large areas of circulating air at its poles, known as polar vortices. Earth's polar vortices appear in the winter of each hemisphere above a core of cold polar air. The polar vortex identified on Venus, by contrast, has a cold collar[4,6], inside which there is a curious, dipole-shaped feature about 10 K warmer than its surroundings. Again, solar heating alone cannot explain this warm dipole. The compressive effect of intense downdrafts is a possible cause: the dipole might be an extension of a 'Hadley circulation', a system consisting of hot air rising at low latitudes and cold air sinking at the poles, with a return flow at deeper levels. On Venus, the equator is warmer than the poles at altitudes above 65 km (ref. 4). What keeps a Hadley circulation going under these circumstances is not so clear[13].

Because the solid planet spins so slowly, Venus's atmosphere creates its own rotation. The direction of rotation is the same as that of the planet — backwards relative to the orbital motion. The influence of the Sun grows with increasing altitude, as the air becomes thinner, and a smaller amount of absorbed energy can produce a larger temperature change. The Sun takes over entirely at altitudes above 90 km, and the circulation shifts from rotation about the polar axis to flow from the day side to the night side. Evidence for this day–night circulation comes from compounds such as nitric oxide, carbon monoxide and oxygen, which are produced on the day side and transported to the night side, and from compounds such as hydrogen fluoride and hydrogen chloride, which are destroyed on the day side and transported as chemical by-products to the night side[1,7].

A final, intriguing point of comparison between the atmospheres of our own planet and Venus concerns lightning: there shouldn't be any on Venus, whose clouds are like terrestrial smog clouds, which do not produce lightning. And indeed, no clear-cut visible flashes were to be seen on either the day or the night side of the planet. Yet Venus Express detected whistlers[8] — low-frequency electromagnetic waves that last for a fraction of a second and are thought to originate with an electrical discharge. Cells of convecting air were seen on the day side[5], but the amount of cloud material in the cells is around 100 times less than that in a terrestrial thunderstorm. On Earth, that would not be enough material to charge the clouds.

The evidence for electrical storms on Venus is thus contradictory. Their existence would be surprising, but perhaps we have simply not thought of all the ways that electricity can be generated in a planetary atmosphere. If our intuition, based on terrestrial experience, is failing us, it wouldn't be for the first time.

Venus Express has enough fuel to last until 2013. As the mission controllers learn what the spacecraft can do, the observation sequences will become more ambitious, and the amount we know about Venus will grow. From 2011, Venus Express will be accompanied by Japan's Planet-C spacecraft, also known as Venus Climate Orbiter. Other future Venus missions might use balloons to measure the winds and sample trace gases and their isotopes, or to grab rocks from the hot surface and take them up to cooler altitudes for analysis. If one day we can follow climate processes on Venus as we follow them on Earth, then we might begin to understand weather in general. ■

Andrew P. Ingersoll is in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.
e-mail: api@gps.caltech.edu

1. Bertaux, J.-L. *et al. Nature* **450,** 646–649 (2007).
2. Zhang, T. L. *et al. Nature* **450,** 654–656 (2007).
3. Barabash, S. *et al. Nature* **450,** 650–653 (2007).
4. Pätzold, M. *et al. Nature* **450,** 657–660 (2007).
5. Markiewicz, W. J. *et al. Nature* **450,** 633–636 (2007).
6. Piccioni, G. *et al. Nature* **450,** 637–640 (2007).
7. Drossart, P. *et al. Nature* **450,** 641–645 (2007).
8. Russell, C. T. *et al. Nature* **450,** 661–662 (2007).
9. Ingersoll, A. P. *J. Atmos. Sci.* **26,** 1191–1198 (1969).
10. Donahue, T. M., Hoffman, J. H., Hodges, R. R. & Watson, A. J. *Science* **216,** 630–633 (1982).
11. Seiff, A. *et al. J. Geophys. Res.* **85,** 7903–7933 (1980).
12. Yoder, C. F. in *Venus II — Geology, Geophysics, Atmosphere, and Solar Wind Environment* (eds Bougher, S. W. *et al.*) 1087–1124 (Univ. Arizona Press, Tucson, 1997).
13. Gierasch, P. J. *et al.* in *Venus II — Geology, Geophysics, Atmosphere, and Solar Wind Environment* (eds Bougher, S. W. *et al.*) 459–500 (Univ. Arizona Press, Tucson, 1997).

©2007 Nature Publishing Group

REPRODUCTION

The unusual suspect

Colin L. Stewart

The p53 protein is widely studied for its function as a tumour suppressor, preventing cancer. It emerges that this protein also has an essential physiological role in regulating embryo implantation in mice.

The p53 protein, which has been given the accolade 'guardian of the genome', is a celebrity among cellular proteins. If you search the PubMed database for p53, you'll get more than 44,000 hits, the vast majority of which discuss the role of this protein in protecting cells from the consequences of stress-induced DNA damage. Given the essential role of p53 in preventing tumour formation, the finding that most p53-deficient mice develop normally, and only succumb to tumorigenesis later in life, was surprising[1]. But, on page 721 of this issue, Hu *et al.*[2] show that this familiar transcription factor has an unanticipated — albeit essential — role in regulating fertility in female mice.

Implantation — during which the early embryo, or blastocyst, comes into physical contact with the mother's uterus in a highly coordinated way — is an essential step in the reproduction of most mammals. It leads to the formation of the placenta, which supports the growth and development of the fetus until its birth. Before implantation, the uterus undergoes extensive rounds of cell proliferation and differentiation that make it receptive to a potential blastocyst. These events are primarily regulated by the direct effect of the ovarian steroid hormones oestrogen E2 and progesterone P4 (ref. 3), and are further augmented by the induction by these hormones of many locally produced growth factors and immune-mediator proteins (cytokines)[4].

Among these is the cytokine LIF, which was first implicated in mammalian development because of its essential function in sustaining mouse embryonic stem cells. LIF is one of many proteins that are produced and secreted by the endometrial glands of the uterus. In many different mammalian species, LIF is most highly expressed at the onset of implantation[5]. In fact, implantation cannot occur unless epithelial cells lining the uterus are exposed to LIF (ref. 6) (Fig. 1). Moreover, LIF-deficient female mice, in which blastocysts develop normally, are infertile; but a single injection of LIF into these animals, on the day in which implantation would normally occur (in mice, day 4 of pregnancy), is sufficient to initiate implantation and subsequent normal embryonic development to birth[7].

Hu *et al.*[2] now show that p53 plays a significant part in regulating LIF expression in the uterus, and so in female fecundity. These authors were intrigued by the low fertility of p53-deficient female, but not male, mice. They noticed that, although in p53-deficient female mice only low numbers of embryos could implant, the number of healthy embryos produced was not affected. The authors find that p53 regulates the transcription of the mouse *Lif* gene, and that in p53-deficient female mice, expression levels of uterine LIF are reduced about fourfold. The clincher was that the injection of mated p53-deficient females with LIF significantly increased the number of implanted embryos that successfully developed to term.

In the uterus, LIF is primarily regulated by E2. Because uterine levels of p53 seem to be constant during the pre-implantation period, and the levels of E2 in p53-deficient female mice are normal, the authors suggest that p53 probably forms a complex with the E2 receptor. They propose that this putative complex is necessary for maximal expression of LIF at day 4 of pregnancy in mice.

The function of p53 in regulating fertility is unexpected, as it is not clear why p53, which is usually activated by stress, should regulate *Lif* transcription in the uterus. At implantation, the uterine vascular system undergoes extensive re-plumbing to support the considerable increase in both uterine-cell proliferation (decidualization) and the growth of the embryo itself, and to prevent hypoxia (oxygen shortage). As hypoxia is one of the main stressors that regulate p53 activity[8], it is possible that with the evolution of the reproductive system of mammalian females, activation of p53 in response to hypoxia was hijacked to help coordinate implantation-associated changes in vascularization.

What are the implications of Hu and colleagues' findings for human fecundity? Studies on p53 and its regulators, such as Mdm2, have identified polymorphisms in amino-acid residue 72 of p53, as well as a single nucleotide polymorphism (SNP309) in the transcriptional regulatory region of *Mdm2* (ref. 9). Both of these polymorphisms are thought to affect the transcriptional activity and the levels of p53 (ref. 10). So, in light of Hu and colleagues' data, it is intriguing that the presence of a particular polymorphism (proline substituting for arginine) at residue 72 of p53 strongly correlates with women having recurrent implantation failure[11]. Genetic-variation-associated alterations in the levels of p53 may also influence the levels of this protein in the uterus, which may, in turn, affect LIF levels; reduced levels of LIF are also associated with a decrease in implantation rates in humans[12]. It would therefore be interesting to see whether the SNP309 polymorphism in *Mdm2* is linked to cases of unexplained human infertility.

Extensive efforts are under way to identify small-molecule regulators of p53, in order to suppress tumour growth. Such molecules might also be useful in enhancing implantation rates as part of fertility treatments, or even for inhibiting implantation as a means of contraception. ■

Colin L. Stewart is at the Institute of Medical Biology, 8A Biomedical Grove, 06-06 Immunos, 138648 Singapore.
e-mail: colin.stewart@imb.a-star.edu.sg

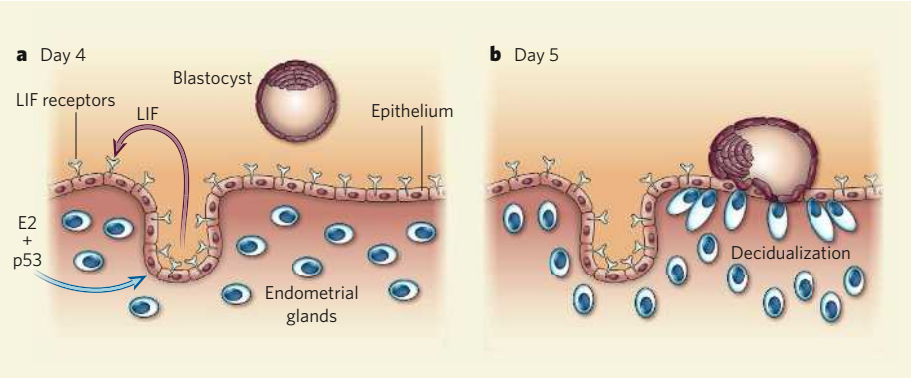


Figure 1 | Regulation of mouse-embryo implantation at a molecular level. a, At day 4 of pregnancy, oestrogen E2 induces LIF expression in the endometrial glands, leading to LIF secretion into the uterine lumen. There, LIF binds to its receptors on the surface of epithelial cells. **b**, This makes the uterus receptive to the blastocyst, which implants by day 5 of pregnancy. Hu *et al.*[2] find that LIF expression in the endometrial glands also depends on the regulatory activity of p53. In the absence of p53, insufficient LIF is produced, the uterus does not become adequately receptive, and fewer blastocysts implant.

1. Donehower, L. A. *et al. Nature* **356,** 215–221 (1992).
2. Hu, W., Feng, Z., Teresky, A. K. & Levine, A. J. *Nature* **450,** 721–724 (2007).
3. Finn, C. A. & Martin, L. *J. Reprod. Fertil.* **39,** 195–206 (1974).
4. Dimitriadis, E., White, C. A., Jones, R. L. & Salamonsen, L. A. *Hum. Reprod. Update* **11,** 613–630 (2005).
5. Cullinan, E. B. *et al. Proc. Natl Acad. Sci. USA* **93,** 3115–3120 (1996).
6. Stewart, C. L. *et al. Nature* **359,** 76–79 (1992).
7. Chen, J. R. *et al. Endocrinology* **141,** 4365–4372 (2000).
8. Graeber, T. G. *et al. Nature* **379,** 88–91 (1996).
9. Bond, G. L. *et al. Cell* **119,** 591–602 (2004).
10. Pietsch, E. C., Humbey, O. & Murphy, M. E. *Oncogene* **25,** 1602–1611 (2006).
11. Kay, C., Jeyendran, R. S. & Coulam, C. B. *Reprod. Biomed. Online* **13,** 492–496 (2006).
12. Mikolajczyk, M. *et al. Reprod. Biol.* **3,** 259–270 (2003).

©2007 Nature Publishing Group

CELL BIOLOGY

Pore puzzle

John D. Aitchison and Richard W. Wozniak

Where would you start in trying to work out the structure of a macromolecular machine consisting of 456 proteins? Taking a combined experimental and computational approach is one answer.

Consider a 1,000-piece jigsaw puzzle. There are millions of ways in which the pieces might fit together, yet there is only one solution. We solve the puzzle by considering each piece and ruling out those that don't fit physical restraints, such as colour patterns, the overall shape of the picture and the potential for interlocking. This is the basic premise for the multidisciplinary approach taken by Alber *et al.*[1,2] (pages 683 and 695 of this issue) to solve the structure of a large macromolecular machine, the yeast nuclear pore complex. But, in their case, the pieces were proteins, the various restraints were of a biochemical and morphological nature, and computers explored the placement of each protein into a single ensemble solution.

Nuclear pore complexes (NPCs) serve as regulated ports for transporting molecules into and out of the cell nucleus. Researchers have studied these complexes for decades. They have defined the overall morphology of the NPC, and have used various approaches to determine the inventory of its constituent proteins — the 30 different nucleoporins — and the abundance and rough localization of each nucleoporin. They have also begun to identify genetic and physical interactions between these proteins[3]. All NPCs — from those of yeast to human — have largely symmetrical, doughnut-shaped structures that lie within a pore that is formed by a membrane connecting the inner and outer layers of the nuclear envelope. Each NPC consists of eight morphologically similar units called spokes, which are arranged around a central channel through which transport occurs[3,4].

The previous studies were essential for characterizing NPC functions and for identifying subcomplexes within them. But because of the size and complexity of this structure, these earlier studies did not reveal the detailed architecture of the entire assembly. To tackle the NPC puzzle, Alber *et al.*[1,2] systematically generated comprehensive biochemical and morphological data of the sort collected in the past, and combined them with state-of-the-art computational integration strategies.

Using the known size and symmetry of the NPC, the team established more than 10,000 other restraints as input for their computational analysis. These included the stoichiometry (number) and localization of each NPC component, as well as interactions between different components. Their goal was to generate sufficient data so that, like the jigsaw puzzle, there could be only one possible solution.

But each data type had uncertainty associated with it. For example, because of the limitations of the imaging techniques, the localization of each nucleoporin could be narrowed down to only a relatively large volume within the NPC. Thus, each data type by itself was insufficient for determining the structure, because, by analogy, it was like a set of partly broken or faded puzzle pieces. The solution was data integration: even with uncertainty in every data type, the likelihood of satisfying all restraints with an incorrect structure becomes vanishingly small as more data are generated.

Once the experimental data had been translated into mathematical representations of restraints, the authors' procedure involved taking a random configuration of beads

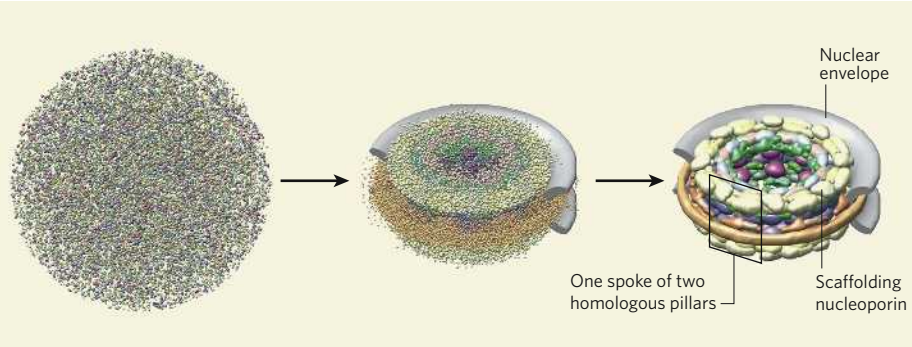


Figure 1 | Determining the structure of the nuclear pore complex. Alber *et al.*[1,2] started with 456 randomly distributed protein components of the complex (coloured beads), corresponding to several copies of each of the 30 different constituent nucleoporins. They moved the beads randomly, while computationally minimizing thousands of biochemical and morphological restraints. The final structure shows the distribution of proteins resulting from the best-scoring simulations. (Figure courtesy of F. Alber and A. Sali, University of California, San Francisco.)

50 & 100 YEARS AGO

50 YEARS AGO

The proposals of the South African Government to enforce racial segregation at the university-level and to exercise rigid control over projected colleges for non-whites has caused concern and indignation in university circles throughout the Commonwealth and beyond. Some of this concern found expression in a well-attended meeting held at the Caxton Hall in London ... The audience gave a polite but sceptical hearing to Prof. L. J. du Plessis, of the Potchefstroom University, who had come over from South Africa to present the case in favour of university apartheid ... Prof. du Plessis said :"If I were in England, I would be an integrationist too, because it is not a danger to English national character, and if I were in America I would be an integrationist too; the negroes in America are in no danger of destroying that great nation of America in its identity. But the Bantus and Indians and Coloureds of South Africa, if they are integrated more fully than they are now, would undoubtedly destroy our great South African nation".

From *Nature* 30 November 1957.

100 YEARS AGO

Since writing the notice of Mr. le Souef's book on Australian wild life ... I have been making inquiries as to the existence in collections of any examples of platypus egg definitely known to have been taken from the nest after extrusion ... In the central hall of the British Museum is shown an egg-shell of a platypus sent from Queensland by Mr. G. P. Hill in 1902, but this ... was doubtless found in its present broken condition. Such broken shells might, apparently, be extruded from the uterus with the foetus; and, so far as I can find, there still appears to be no definite evidence that the eggs are really laid entire.

From *Nature* 28 November 1907.

©2007 Nature Publishing Group

representing proteins in three-dimensional space. Then, again in a random process, the beads were iteratively moved to sample the potential space they could occupy, while simultaneously minimizing violations of restraints (Fig. 1). The configurations in each iteration were scored for their ability to satisfy the restraints. After thousands of trials, the best-scoring structures were superposed to produce a map of the localization of each component. As there was no gold-standard structure for validation, the final structure was assessed by criteria of self-consistency: all the restraints were satisfied; variability between possible solutions was small; similar structures were obtained when portions of the data were omitted; inclusion of incorrect data led to poorly resolved (frustrated) structures; and previously reported subcomplexes were rediscovered.

In biology, the motivation for structure determination involves an element of faith: that, once solved, the structure will tell you something fundamental. And this is the case here. For many years, subcomplexes of the NPC consisting of handfuls of nucleoporins have been identified and characterized in isolation. Alber and colleagues' determination of the molecular architecture of the NPC places these components in the context of the whole complex and provides researchers with a frame of reference for future studies, including devising more tightly directed approaches for testing existing and new hypotheses generated from the study. For example, the authors find that half of the nucleoporins in the yeast NPC form an interconnected network or scaffold that coats the pore membrane (Fig. 1). These proteins are composed of two types of structural domain with a similar configuration to that of the protein complexes that coat carrier vesicles involved in intracellular trafficking. This suggests that the two complexes — both of which are involved in the curvature of membranes — are evolutionarily related.

The authors also show that the core scaffold is built of repetitive structural pillars. Each of the eight symmetrical spokes is composed of paired groups of duplicated or structurally related nucleoporins, forming pairs of homologous columns (Fig. 1). This modularity has implications for ways in which the NPC structure might vary within a cell, and provides insight into how the structure could have evolved.

One class of nucleoporin is characterized by a repeated motif of phenylalanine–glycine (FG, in single-letter code) amino-acid residues, and is thought to control transport across the NPC. But how FG nucleoporins achieve this is a contentious issue[5]. As would be expected, Alber *et al.* find that these nucleoporins are exposed to the central pore, where they form a selective barrier to transport. But as the precise structure of this region of the NPC could not be resolved by this analysis, exactly how transport occurs remains unclear.

The combined experimental and computational approach that Alber and colleagues[1,2] took to determine a complex structure has yielded an excellent platform for further resolution of the NPC structure. Perhaps, more importantly, it serves as a paradigm for future data-integration strategies designed to understand the structure, dynamics and mechanistic properties of other large macromolecular machines. ■

John D. Aitchison is at the Institute for Systems Biology, 1441 North 34th Street, Seattle, Washington 98103, USA. Richard W. Wozniak is in the Department of Cell Biology, University of Alberta, Edmonton, Alberta T6G 2H7, Canada.
e-mails: jaitchison@systemsbiology.org; rick.wozniak@ualberta.ca

1. Alber, F. *et al. Nature* **450,** 683–694 (2007).
2. Alber, F. *et al. Nature* **450,** 695–701 (2007).
3. Tran, E. J. & Wente, S. R. *Cell* **125,** 1041–1053 (2006).
4. D'Angelo, M. A. & Hetzer, M. W. *Cell. Mol. Life Sci.* **63,** 316–332 (2006).
5. Stewart, M. *Nature Rev. Mol. Cell Biol.* **8,** 195–208 (2007).

MICROSCOPY

Elementary resolution

Christian Colliex

The atoms and bonds that make up complex solids can be identified chemically — a feat made possible by cleverly combining spectroscopic and structural information conveyed by electrons scattered through a thin sample.

Spying out the atomic organization of solid structures has long been a dream of materials scientists. They have built ever more refined microscopes to realize that dream, using probes of electrons, or of local currents or forces, to build up increasingly detailed pictures of solids. But identifying the chemical nature of atoms in position has been a pricklier problem.

Two techniques have reached a stage where this is becoming possible. First, there is the three-dimensional topographic atom probe, which uses mass spectroscopy to identify atoms that have been made to evaporate gradually from the surface of a specimen solid[1]. Second, there is electron energy-loss spectroscopy (EELS), which measures the energy of electrons transmitted through a thin specimen in a scanning transmission electron microscope (STEM) (Fig. 1). On page 702 of this issue[2], Kimoto *et al.* report the use of EELS for element-selective imaging of columns of atoms in a crystal. This is a further stage in a healthily competitive race in which results are appearing at an accelerated pace.

The starting signal for this race was fired almost 15 years ago. In a News & Views article published at the time[3], L. M. Brown detailed the near-simultaneous publication of three papers[4–6], all of which described the measurement, with atomic-scale resolution, of compositional changes across an interface between two grains or components in a solid. In each case, a probing electron beam was directed down a column of atoms, and the resulting EELS spectrum was used to ascertain the atomic species as well as their valence states and local coordination. The question of precedence among these three papers gained renewed piquancy last year with the appearance of a corrigendum[7] to the first of the papers[4] in which the authors admitted errors in data processing. These errors did not, they said, modify their general claim of atomic resolution.

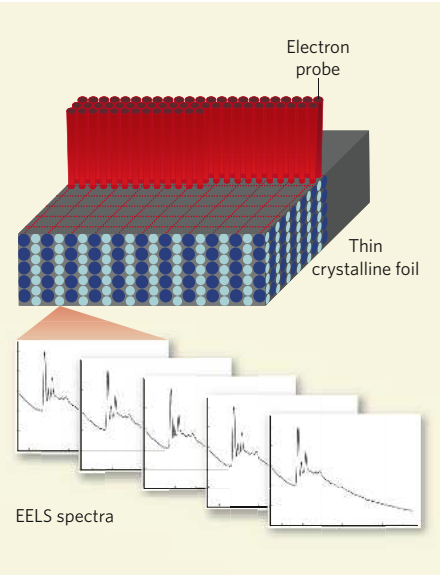


Figure 1 | EELS in action. In electron energy-loss spectroscopy (EELS), spectra are recorded successively by an ångström-size incident electron beam to build up an atomic-scale map of a thin crystalline foil. By comparing these signals with structural information from electrons scattered through large angles by interactions with the crystal atoms (the annular dark-field, or ADF, signal), Kimoto and colleagues[1] identified the chemical elements in a crystal structure.

Fortunately, advances in instrumentation and methods now allow the task to be tackled more reliably. First, the general improvement in all the electrical and mechanical components involved has reduced instability and performance degradation. Second, the so-called spectrum–image strategy[8] increases confidence in

©2007 Nature Publishing Group

the positioning of the probing electron beam, and so in the correlation of spectral and structural information. This technique involves acquiring successively, for all pixels of a two-dimensional grid, both the EELS spectrum and a second signal. This latter 'annular dark-field' (ADF) signal is formed by electrons that are scattered through large angles by interactions with the crystal atoms, and it provides supplementary information on the precise structure of the crystal.

Third and finally, improvements in focusing optics have compensated for aberrations in STEMs, allowing a reduction in probe diameter and so an increase in its spatial sensitivity[9]. This allows the crystal lattice to be directly imaged at an unprecedented sub-ångström (10^{-10} m) resolution in the ADF mode. Furthermore, by opening up the illumination angle for a given probe dimension, a significant increase in the probe current is achieved.

Kimoto *et al.*[2] have taken advantage of most of these advances. Their STEM does not, however, use an aberration corrector. The available electron intensity is therefore rather low, and the total acquisition time for an image consisting of more than 1,000 pixels — about an hour — is very long. Recent results[10] confirm that the use of an aberration corrector produces maps with a greatly improved signal-to-noise ratio within a much-reduced total recording time.

Kimoto and colleagues' most important results are given in their Figure 2 (page 704), which shows images of their chosen crystal, the layered manganite $La_{1.2}Sr_{1.8}Mn_2O_7$. White dots of variable intensity in the plane of the ADF image correspond to columns formed of different atomic species running into the page, parallel to the probing electron beam (their Fig. 2a). This pattern serves as a guide for comparison with elemental EELS maps (Fig. 2d–f). These are formed from electrons that are scattered off the sample with energies corresponding to the binding energies of certain inner electron shells of oxygen, manganese and lanthanum (Fig. 2d–f). Most of the crystal's obvious features — such as the Mn and La atomic columns — appear at positions that are in accordance with models of the manganite's crystal structure. Oxygen atoms not obvious from the ADF image can also be identified. Finally, the comparison of maps for lanthanum at electron energy losses of 99 and 832 electronvolts, corresponding to the binding energies of that element's N and M electron shells (Fig. 2c, f), clearly shows how the spatial width of the response diminishes as energy loss increases.

More refined theories will be required to model such high-resolution chemical maps fully, as their interpretation is far from intuitive. One must first take into account the effects of propagation of ångström-size, high-energy electron probes along atomic columns, such as 'channelling' and 'dechannelling'. Furthermore, non-local inelastic scattering processes can lead to spurious signals[11]. The first experimental maps are now demonstrating the importance of refining descriptions of electron–matter interactions[2], allowing guidelines to be developed to optimize experimental conditions for direct interpretation.

But what is all of this good for? Why is it so important to know the elemental distribution of the atoms in a solid with such a degree of refinement? The choice of specimens by researchers including Kimoto *et al.*[2] provides a clue. Unusual electronic features, such as the presence of two-dimensional superconducting layers[13], emerge across or near almost atomically flat interfaces within such complex oxides. Determining as best we can the nature of the atoms concerned and their bonds might help us to discern the patterns underlying such features. Preliminary results are still rather noisy, but the tremendous experimental developments of STEM machines fully targeted to this task[9,10] will ensure that the ADF–EELS technique has a bright future in this fast-expanding field of materials science. ■

Christian Colliex is in the Laboratoire de Physiques des Solides, UMR CNRS 8502, Université Paris Sud, Bâtiment 510, 91405 Orsay, France.
e-mail : colliex@lps.u-psud.fr

1. Blavette, D. *et al. Science* **286,** 2317–2319 (1999).
2. Kimoto, K. *et al. Nature* **450,** 702–704 (2007).
3. Brown, L. M. *Nature* **366,** 721 (1993).
4. Browning, N. D., Chisholm, M. F. & Pennycook, S. J. *Nature* **366,** 143–146 (1993).
5. Muller, D. A., Tzou, Y., Raj, R. & Silcox, J. *Nature* **366,** 725–727 (1993).
6. Batson, P. E. *Nature* **366,** 727–728 (1993).
7. Browning, N. D., Chisholm, M. F. & Pennycook, S. J. *Nature* **444,** 235 (2006).
8. Jeanguillaume, C. & Colliex, C. *Ultramicroscopy* **28,** 252–257 (1989).
9. Nellist, P. D. *et al. Science* **305,** 1741 (2004).
10. Muller, D. A. *et al. Microsc. Microanal. 2007 Annu. Meet.*, Ft Lauderdale, FL, Talk 718 (2007).
11. Oxley, M. P. *et al. Phys. Rev. B* **76,** 064303 (2007).
12. Bosman, M. *et al. Phys. Rev. Lett.* **99,** 086102 (2007).
13. Reyren, N. *et al. Science* **317,** 1196–1199 (2007).

NEUROSCIENCE

Sensors and synchronicity

Ruth Heidelberger

Synaptic communication is triggered by action potentials, but neurons also talk to each other in between action potentials. Specific intracellular-calcium sensors regulate these conversations.

Neurons communicate with each other through synaptic junctions. In the presynaptic neuron, voltage-gated calcium channels are activated in response to an action potential, allowing the entry of extracellular calcium. This triggers the fusion of neurotransmitter-laden synaptic vesicles with the plasma membrane and the release of neurotransmitter molecules. Released molecules bind to receptors on the postsynaptic neuron, initiating a postsynaptic response. The close proximity of release sites, calcium channels and synaptic vesicles, and the use of a calcium sensor for release with five low-affinity calcium-binding sites, lead to the generation of a burst of neurotransmitter release synchronous with the stimulus[1]. But neurons also secrete neurotransmitter through another, much less well-understood mechanism called asynchronous release. On page 676 of this issue, Sun *et al.*[2] shed light on asynchronous release, showing that it is mediated by an unidentified calcium sensor with unexpected properties.

Synchronous release predominates during low-frequency action-potential firing, whereas during high-frequency stimulation neurotransmitter release is also observed in the interval between action potentials, when calcium channels have closed and calcium entry into presynaptic neurons has ceased. It has been postulated that this asynchronous release is triggered by residual calcium and, as such, must rely on a calcium sensor with high-affinity calcium-binding sites[3]. Testing this hypothesis proved problematic, not only because most synapses are small and inaccessible, but also because it is inherently difficult to separate asynchronous from synchronous release. Consequently, the mechanisms that contribute to this form of synaptic communication have remained enigmatic.

Sun *et al.*[2] investigated asynchronous release in the calyx of Held, a large, accessible synapse in the mammalian auditory pathway. Previously, synaptotagmin-2 — the calcium sensor for fast, synchronous release in the calyx synapse[4] — has been genetically deleted in mice, allowing asynchronous release to be studied in isolation[4,5]. Sun and colleagues capitalized on the ability to manipulate and measure presynaptic calcium in the genetically modified mouse calyx of Held. By triggering release via the homogeneous, global elevation of presynaptic calcium rather than by calcium-channel activation, potential differences in release rates caused by the different positions of vesicles with respect to calcium channels were eliminated[6]. This allowed Sun and colleagues an exclusive, direct look at the calcium sensor for asynchronous release.

In agreement with previous work[1,6,7], the authors find that, in normal synapses, there is a steep relationship between the rate of neurotransmitter release and calcium

©2007 Nature Publishing Group

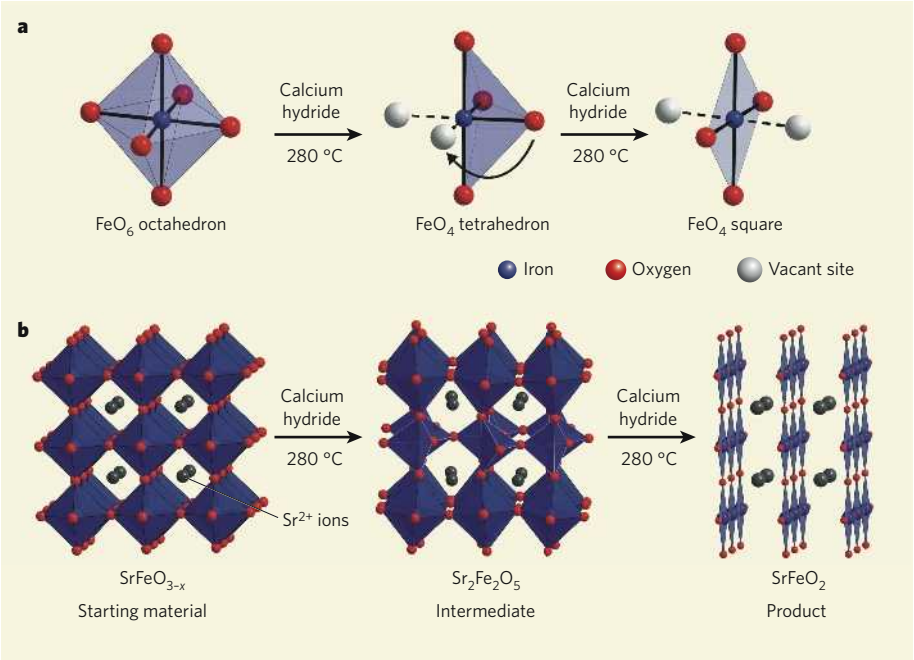


Figure 1 | Ion movement in a complex metal oxide. Tsujimoto *et al.*[1] report the preparation of $SrFeO_2$ (Sr is strontium, Fe is iron), a complex metal oxide with an unusual arrangement of ions. **a**, The starting material ($SrFeO_{3-x}$, where x is about 0.125) contains FeO_6 octahedra. On heating with calcium hydride, some of the oxygen ions (known as oxide ions) are lost, so that intermediate FeO_4 tetrahedra form. One of the remaining oxide ions then moves to a vacant site left behind by an ion (indicated by the arrow), forming an FeO_4 square. **b**, The crystal lattice of the starting material is an array of FeO_6 octahedra, in which each oxygen is shared between two iron atoms. Strontium ions (Sr^{2+}) fit in between the rows of octahedra. In the first step of the reaction, loss of some of the oxide ions leads to an intermediate compound, $Sr_2Fe_2O_5$, consisting of alternating rows of FeO_6 octahedra and FeO_4 tetrahedra. In the second step, more oxide ions are lost and some of the remaining oxide ions change position. The $SrFeO_2$ product thus forms as sheets of FeO_4 squares, interleaved with strontium ions. The iron and strontium ions retain their positions throughout the process.

consider the differences between solid-state reactions and those that occur in solution. Reactions in solution can be performed at low temperatures, because molecular diffusion occurs easily and the reacting molecules don't require much energy to mix together. Under these conditions, if a molecule can take part in several reactions, only the one with the lowest energy barrier to activation tends to occur. The product of such a 'kinetically controlled' reaction is the one that forms fastest, and is not necessarily the one that is most stable. It is therefore possible to control reactions in solution so that they occur only at specific parts of a molecule. By performing stepwise transformations on individual chemical groups, a product can be prepared that has a controlled composition and structure that are clearly related to those of the starting compound.

But nearly all complex metal oxides are prepared at high temperatures (typically greater than 1,000 °C). This is because no solvent is used to aid diffusion, yet the reacting ions must travel large distances (of the order of micrometres) to form the products. At these temperatures, enough energy is available to allow the occurrence of reactions that have high energy barriers. Given a choice of reaction pathways, the most favourable one is that which yields the most thermodynamically stable configuration of atoms. In such thermodynamically controlled systems, the product generally does not conserve any of the structural features of the reactants (unlike reactions in solution), and it is therefore much more difficult to direct the course of the reaction[5].

Tsujimoto *et al.*[1] overcome these limitations in their preparation of $SrFeO_2$. Their starting material is a complex metal oxide that contains Fe^{4+} ions ($SrFeO_{3-x}$, where x is about 0.125). The authors form their unusual product by removing an oxide ion from the starting material, a process that is coupled to a redox reaction in which Fe^{4+} ions are converted into Fe^{2+} ions.

The overall process uses a recently discovered reagent (calcium hydride) for the kinetically controlled removal of oxygen from oxides[6], and it occurs at the remarkably low temperature of 280 °C. These conditions provide insufficient thermal energy to rearrange the structure of the starting material completely — only the relatively mobile oxide ions can change position (Fig. 1). The strontium and iron ions in the product retain the positions they held in the starting material. The most thermodynamically stable products — iron metal and strontium(II) oxide — do not form, because the required long-range diffusion for the process is too slow at this temperature. The less stable $SrFeO_2$ forms instead, because this is a faster reaction.

Nevertheless, the reaction pathway that leads to $SrFeO_2$ (Fig. 1) is unexpected. An intermediate ($Sr_2Fe_2O_5$) is formed first, as oxide ions are removed from the starting material. This intermediate consists of alternating sheets of FeO_4 tetrahedra and FeO_6 octahedra. Conversion of the tetrahedra into the square planes of the final product requires that the oxide sites vacated in the formation of the intermediate be refilled with other oxide ligands. This is a crucial observation, because it demonstrates that all the oxide ions are mobile, not just those being removed from the system.

The discovery that oxide ions can be mobile at relatively low temperatures, albeit in the presence of a strong chemical driving force, opens up a host of synthetic possibilities — for example, the strong magnetic interactions seen in $SrFeO_2$ could be modified in a controlled way by making complex oxides of different transition metals. But the practical applications are just as exciting — high oxide-ion mobility is required for several emerging technologies, most notably solid-oxide fuel cells[7]. So although Tsujimoto and colleagues' discovery[1] may occur only at an atomic level, its ramifications could extend far more widely. ■

Michael A. Hayward is in the Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, South Parks Road, Oxford OX1 3QR, UK.
Matthew J. Rosseinsky is in the Department of Chemistry, University of Liverpool, Liverpool L69 7ZD, UK.
e-mails: michael.hayward@chem.ox.ac.uk; m.j.rosseinsky@liverpool.ac.uk

1. Tsujimoto, Y. *et al. Nature* **450,** 1062–1065 (2007).
2. Blundell, S. *Magnetism in Condensed Matter* (Oxford Univ. Press, 2001).
3. Rao, C. N. R. & Raveau, B. *Transition Metal Oxides: Structure, Properties, and Synthesis of Ceramic Oxides* 2nd edn (Wiley-VCH, Weinheim, 1998).
4. Cava, R. J. *J. Am. Ceramic Soc.* **83,** 5–28 (2000).
5. Stein, A., Keller, S. W. & Mallouk, T. E. *Science* **259,** 1558–1564 (1993).
6. Hayward, M. A. *et al. Science* **295,** 1882–1884 (2002).
7. Atkinson, A. *et al. Nature Mater.* **3,** 17–27 (2004).

Corrections

■ The News & Views article "Venus: Express dispatches" by Andrew P. Ingersoll (*Nature* **450,** 617–618; 2007) contained the erroneous statement that Venus's equator is warmer than the poles at altitudes above 65 km. It is colder.

■ There was an incorrect reference citation in the article "Microscopy: Elementary resolution" by Christian Colliex (*Nature* **450,** 622–623; 2007). In the statement "The first experimental maps are now demonstrating the importance of refining descriptions of electron–matter interactions[2]," the correct citation is not reference 2 but reference 12 (M. Bosman *et al. Phys. Rev. Lett.* **99,** 086102; 2007).

■ In the article "Astronomy: Sloan at five" by Robert C. Kennicutt Jr (*Nature* **450,** 488–489; 2007), we should clarify that the Sloan Digital Sky Survey was used only to select candidate stars for the spectroscopic observations that led to the discovery cited in reference 9 (W. R. Brown *et al. Astrophys. J.* **622,** L33–L36; 2005).

©2007 Nature Publishing Group

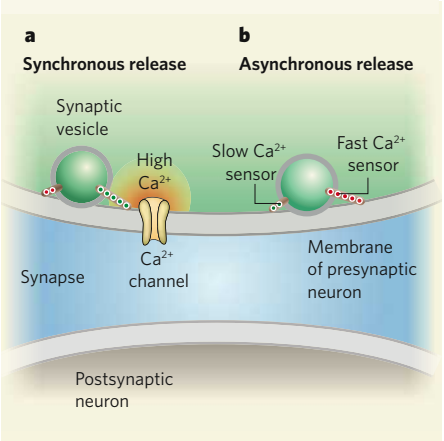


Figure 1 | Two modes of neurotransmitter release. a, Synchronous release is evoked when each of the five calcium-binding sites (green circles) on a fast calcium sensor, such as synaptotagmin-2, is occupied. This happens most readily when a synaptic vesicle is located in close proximity to an open calcium channel. **b**, Sun *et al.*[2] find that asynchronous release is evoked when two binding sites on a slow calcium sensor are occupied (green circles), whereas those of the synchronous sensor are not (red circles). This is most likely to happen in response to a smaller or more sustained increase in calcium levels. The authors propose that occupancy states of the two calcium sensors govern whether release is synchronous or asynchronous.

concentration at presynaptic calcium levels of above 1 micromolar. However, in synapses that lack synaptotagmin-2, the slope of this relationship is markedly reduced. Hints of a shallow relationship for asynchronous release are also apparent in normal synapses[6,7]. Below about 1 micromolar, the curves from normal and synaptotagmin-2-deficient synapses are superimposable.

These fascinating results lead to three important conclusions. First, the calcium sensor for asynchronous release is not only distinct from synaptotagmin-2, but, in the range of calcium concentrations studied, also requires fewer bound calcium ions to trigger release. Second, because release is slow and protracted despite rapid, homogeneous calcium increase, properties intrinsic to this calcium sensor must contribute to asynchronous release. Third, even at normal synapses, release can occur through the asynchronous pathway at low calcium concentrations.

Next, the authors used a computational approach to quantitatively analyse the calcium dependence of the rate of neurotransmitter release. The best fit to the experimental data came from a mathematical model that presupposed a single population of releasable synaptic vesicles containing one calcium sensor for synchronous release and another for asynchronous release (Fig. 1). This assumption is plausible because, according to earlier work[8], the releasable pool of vesicles in the calyx of Held comprises both rapidly and slowly releasing vesicles, corresponding to synchronous and asynchronous modes of secretion. Moreover, Sun and colleagues themselves found that the entire pool of releasable vesicles could be released through the asynchronous pathway in the absence of synaptotagmin-2. The authors' model also presupposes that release could occur spontaneously or be triggered by the binding of calcium ions to either the five low-affinity binding sites on the sensor for synchronous release[1] or the two sites on the sensor for asynchronous release.

The remarkable outcome is that asynchronous release can be triggered by the binding of just two calcium ions to a sensor — rather than the traditional five proposed for synchronous release — and that the calcium affinity of the sensor that mediates this release process is almost identical to that for synchronous release. These findings do not support earlier postulations that asynchronous release must be governed by a high-affinity sensor. Rather, it is the relative speeds with which calcium binds and unbinds the sensors that are important; the simulations suggest that these differ by at least one order of magnitude.

So the properties of the faster, synchronous sensor, synaptotagmin-2, are better suited for triggering release in response to large, rapid increases in calcium concentrations that allow simultaneous occupancy of all five binding sites, whereas the slower, unknown asynchronous sensor described by Sun *et al.*[2] is better suited for triggering release in response to slower, more sustained changes in calcium. These results also indicate that the magnitude and duration of the calcium signal determine which sensor drives release, adding a new dimension to the multiple functions of presynaptic calcium.

In light of the exciting and unexpected findings of Sun *et al.* several questions must be addressed. These include identifying the mysterious calcium sensor that mediates asynchronous release; determining whether this sensor has additional calcium-binding sites, even if not all need to be calcium-bound to trigger fusion[9]; gaining a better understanding of presynaptic calcium dynamics, particularly in the vicinity of the releasable pool of vesicles; and identifying mechanisms that regulate the recruitment and/or coupling of releasable vesicles to presynaptic calcium channels. Such knowledge is vital for a full understanding of the factors that determine the tenor of neurotransmitter release and synaptic signalling[6,10]. ■

Ruth Heidelberger is in the Department of Neurobiology and Anatomy, University of Texas Medical School at Houston, 6431 Fannin Street, Houston, Texas 77025, USA.
e-mail: ruth.heidelberger@uth.tmc.edu

1. Heidelberger, R. & Innocenti, B. *Cell Sci. Rev.* **3,** No. 2, 151–183 (2006).
2. Sun, J. *et al. Nature* **450,** 676–682 (2007).
3. Geppert, M. *et al. Cell* **79,** 717–727 (1994).
4. Pang, Z. P., Sun, J., Rizo, J., Maximov, A. & Südhof, T. C. *EMBO J.* **25,** 2039–2050 (2006).
5. Südhof, T. C. *Nature* **375,** 645–653 (1995).
6. Wadel, K., Neher, E. & Sakaba, T. *Neuron* **53,** 563–575 (2007).
7. Wölfel, M., Lou, X. & Schneggenburger, R. *J. Neurosci.* **27,** 3198–3210 (2007).
8. Sakaba, T. *J. Neurosci.* **26,** 5863–5871 (2006).
9. Lou, X., Scheuss, V. & Schneggenburger, R. *Nature* **435,** 497–501 (2005).
10. Bollmann, J. H. & Sakmann, B. *Nature Neurosci.* **8,** 426–434 (2005).

MOLECULAR MOTORS

A step dissected

Zeynep Ökten and Manfred Schliwa

The motor protein kinesin 'walks' by alternately advancing its two motor structural domains. A cutting-edge, single-molecule fluorescence technique reveals further details of this stepping mechanism.

Suppose that walking required energy input in the form of, say, one Gummi bear for every step. In what position would you pop the next Gummi bear into your mouth? When one foot is firmly planted on the ground while the other is lifted up and poised next to it to be thrust forward, or when both feet are on the ground spaced apart at step size? Mori *et al.*[1] (page 750 of this issue) have addressed these questions, not in a person, but in a miniature, biological walking machine — the molecular motor protein kinesin.

The human genome encodes up to 45 kinesins[2], each of which has specialized cellular functions. Common to all kinesins are an evolutionarily conserved structural domain called the head, which acts as a motor, and a variable tail region. One class of kinesin motor, kinesin-1, consists of two identical polypeptide chains linked by a neck region to form a dimer. Kinesin-1 molecules can move great distances on cellular tracks called microtubules (hollow, cylindrical polymers of the protein tubulin), which form an organized web of roadways in the cell. The head region has one binding site for the track and another for kinesin's equivalent of a Gummi bear — an ATP molecule.

Previous work has revealed amazing details about the walking mechanism of kinesin-1 and other motor proteins. Processive (long-distance) movement of kinesin requires the hydrolysis of one ATP molecule per step[3]. The step size is 8 nanometres, which corresponds to the spacing of tubulin subunits[4]. Intriguingly,

©2007 Nature Publishing Group

NATURE|Vol 450|29 November 2007

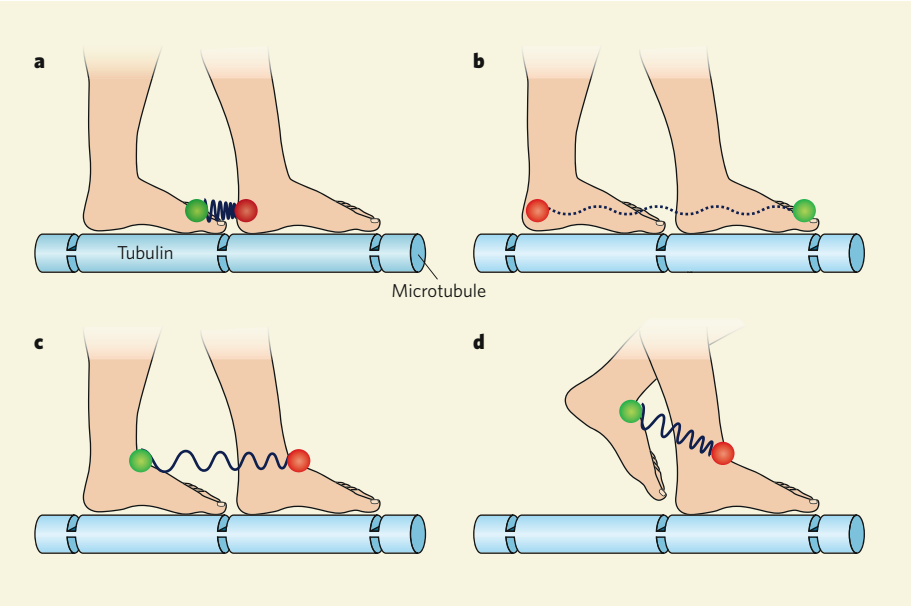


Figure 1 | Positioning of fluorescent sensors on a walking kinesin molecule. If dimeric kinesin-1 is likened to the feet of a walking person, the probes used by Mori *et al.*[1] (green and red balls) for analysis by fluorescence resonance energy transfer (FRET) would be positioned at the toes, heel and near the ankle. Intensity of FRET interaction is indicated by blue waves. The toe–heel pair of sensors yields (**a**) high or (**b**) negligible signals depending on which head leads. The ankle–ankle pair yields (**c**) low signals when both feet are attached and (**d**) a stronger signal when one foot is lifted and poised next to the other. The authors used these position-dependent variations in fluorescence to study the movement of kinesin-1 along microtubules.

kinesin stepping resembles human walking: one head remains firmly bound to the track while the other head moves forward; the two heads alternate in taking steps (the asymmetric hand-over-hand stepping model[5]). During 'walking', the hydrolysis cycles of the ATP molecules on the two heads are kept out of phase[6]. Otherwise, if both heads transit to a detached state synchronously, the motor would lose contact with the microtubule.

Several studies[7–10] also used sophisticated molecular and technical wizardry to answer another question — how the heads of the kinesin-1 dimer are coordinated during processive movement. That is, does the free head pass the microtubule-bound head in one smooth sweep, or is there a characteristic position in which kinesin-1 waits for the next supply of energy? But addressing this question is difficult. Whereas watching a person take a step is easy when you stand nearby, trying to 'see' kinesin move is equivalent to watching a person walk when standing a distance of a kilometre or so away from them.

So how can the gait of a distant walker be figured out? One way would be to attach differently coloured floodlights to the two feet and observe stepping in the dark. From the way the lights add up, mix or become superimposed, one could then draw conclusions about the person's stride.

In a sense, this is the principle Mori *et al.*[1] applied to watch kinesin-1 walk. The first trick they used was to attach molecular floodlights to the two heads of a dimeric kinesin-1. This was done by introducing in desired positions cysteine amino acids to which fluorescent molecules could be attached. If we compare kinesin-1 heads to human feet, informative positions for the floodlights are the toes, heel and ankle. The authors' second trick was to use floodlights of different colours, which could then interact — the third trick — in a process called fluorescence resonance energy transfer (FRET) (Fig. 1).

The FRET technique is based on energy transfer from a short-wavelength donor fluorescent molecule to a longer-wavelength acceptor fluorescent molecule when the two molecules are close together (1–10 nm). This interaction is then detected by monitoring the quenching of the donor molecules' fluorescence or a rise in the fluorescence of the acceptor molecule.

Mori *et al.* applied this technique to kinesin-1 in which the fluorescent molecules were attached to the 'toe' of the rear head region and the 'heel' of the front head; the signal generated from energy transfer between these two molecules would be different if the positions of the two heads were reversed. In the first case (Fig. 1a), the FRET signal would be high because the distance between the fluorescent molecules is only 3 nm; but in the second case, the signal would be negligible because the distance between the donor and acceptor molecules is 13 nm (Fig. 1b). Similarly, when the two fluorescent molecules are attached at positions analogous to the ankle — that is, at neck linkers, one of which emanates from the middle of each of the two kinesin-1 heads — FRET efficiency would be low when the two heads are bound to microtubules (8 nm apart) (Fig. 1c), but higher when one head is detached, hovering near the bound head (Fig. 1d).

By designing experiments with the two fluorescent molecules at different positions on a kinesin-1 dimer, the authors find that when a step is taken in the presence of low levels of ATP, the motor spends most of its time in a one-head-bound state, while the free head lingers next to the bound head. In this case, the free head transits through a short-lived state, when both heads are attached to the next waiting position. But at high ATP levels, an average FRET signal is observed, suggesting that kinesin-1 spends most of its time with both heads bound to the microtubule. These results suggest that, at low ATP levels, the rate-limiting factor for kinesin-1 to take a step is exit from the one-head-bound state. Are these results indicative of two fundamentally different stepping modes of kinesin? Probably not. The basic stepping mechanism is still the same; only some phases may be longer lived than others.

A more precise analysis of the transitions at high ATP levels was impossible because current instrumentation lacks the necessary temporal resolution. This is not surprising. Mori and colleagues' experiments[1] might seem easy on paper, but in practice they are very tricky. The outcome depends not only on the positions of the fluorescent molecules, but also on their relative orientation and the mobility of the head regions. Making these observations at a single-molecule level adds to the technical challenge, and invokes extra caution when interpreting the results. Considering these difficulties, the authors have managed to extract remarkably clear information on the positioning of the two kinesin-1 heads under well-defined experimental conditions. And the use of several engineered kinesins helped them to exclude potential sources of uncertainty.

So despite the big step taken by Mori *et al.*[1], the kinesin walk has not yet given up all of its secrets, as possible transitions and intermediate positions in the fast-stepping mode still remain to be elucidated. Who would have thought that taking a step could be so complex and its study so exciting? ■

Zeynep Ökten and Manfred Schliwa are at the Institute for Cell Biology and the Center for Integrated Protein Science, University of Munich, Schillerstraße 42, 80336 Munich, Germany.
e-mail: schliwa@med.uni-muenchen.de

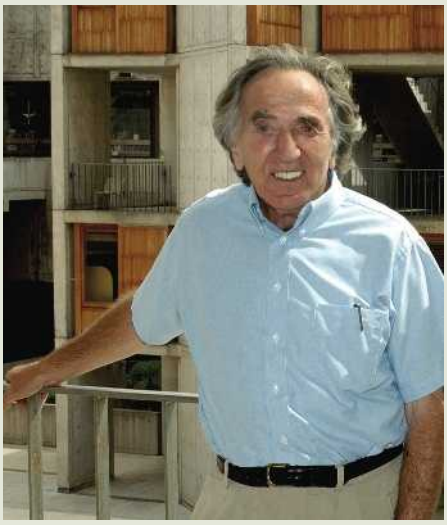
1. Mori, T., Vale, R. D. & Tomishige, M. *Nature* **450**, 750–754 (2007).
2. Richardson, D. N., Simmons, M. P. & Reddy, S. N. *BMC Genomics* **7,** 18 (2006).
3. Schnitzer, M. J. & Block, S. M. *Nature* **388,** 386–390 (1997).
4. Svoboda, K. *et al. Nature* **365,** 721–727 (1993).
5. Block, S.M. *Biophys. J.* **92,** 2986–2995 (2007).
6. Hackney, D. D. *Proc. Natl Acad. Sci. USA* **91,** 6865–6869 (1994).
7. Asenjo, A. B., Weinberg, Y. & Sosa, H. *Nature Struct. Mol. Biol.* **13,** 648–654 (2006).
8. Yildiz, A. *et al. Science* **303,** 676–678 (2004).
9. Alonso, M. C. *et al. Science* **316,** 120–123 (2007).
10. Kawaguchi, K. & Ishiwata, S. *Science* **291,** 667–669 (2001).

©2007 Nature Publishing Group

OBITUARY

Leslie Orgel (1927–2007)

Chemist, molecular biologist and pioneer in the study of the origins of life.



SALK INST.

Leslie Orgel was a brilliant man who excelled in two very different scientific areas. So disparate were these fields that most would think his accomplishments were those of two different individuals. Chemists know of his classic work on devising 'ligand-field theory', which provides an understanding of the structural and physical properties of transition-metal complexes. 'Orgel diagrams' have become a standard way to depict the various energy states of these complexes, and Orgel was recognized for his achievements in this area by his election as a fellow of the Royal Society at the unusually young age of 35. But a broader group of people know that Orgel set the agenda for the experimental study of the origins of life. He pursued that agenda with vigour and intellectual clarity right up until his death on 27 October.

Born in London, Orgel began his career as a chemist at the University of Oxford. But it wasn't long before his second love for biology became apparent. Orgel was one of the small circle of scientists who were present at the birth of molecular biology. Between 1954 and 1955 he worked at the California Institute of Technology with Linus Pauling, purportedly engaged in theoretical chemistry. But Orgel spent most of his time that year talking with the founding fathers of molecular biology — such as Alex Rich and James Watson — and learning from the likes of George Beadle and Max Delbrück, who later won Nobel prizes for their work on genetics.

From 1955 to 1963, Orgel was a faculty member in theoretical chemistry at the University of Cambridge, where he wrote his seminal book on ligand-field theory. It was here that his enthusiasm for biology grew through extensive conversations with Sydney Brenner and Francis Crick.

Orgel was a member of the RNA Tie Club, an informal association of academics who aimed to solve the DNA coding problem — the relationship between the sequence of a gene and that of its corresponding protein. Each member was assigned a different amino acid (Orgel was threonine) or nucleotide base, and given a woollen tie decorated with a stylized double helix. Club members exchanged ideas freely, and played a large part in getting Orgel hooked on molecular biology. By 1964, when he moved to the Salk Institute in La Jolla, California, his second scientific career had begun.

The seeds of Orgel's arguably greatest contributions to science were planted in 1953, when Stanley Miller performed an influential experiment in what would become the field of prebiotic chemistry. Miller demonstrated the abiotic synthesis of amino acids in a mixture of gases that was thought to simulate the atmosphere of the prebiotic Earth. This experiment was highly significant to Orgel, and when he went to La Jolla he sought to address the problems of life's origins by integrating prebiotic chemistry and molecular biology. Accordingly, his focus was not on the machinery of contemporary metabolism (amino acids that form proteins), but on the machinery of a primitive genetic system (nucleotides that form RNA).

Orgel addressed many aspects of the origins of life on Earth, including the prebiotic synthesis of the building-blocks of nucleic acids, the joining together of nucleotides to form polynucleotides, and the non-enzymatic copying of RNA. He was the first to demonstrate that information can be transferred from a preformed RNA template to a newly synthesized RNA molecule in a purely chemical system. The concept of information transfer was critical for Orgel because his ultimate goal was to demonstrate a chemical system that is capable of undergoing darwinian evolution. "After that," he would say, "the rest is just history." The origin of life, in his view, is defined as the origin of evolutionary history.

In the late 1960s, Orgel — along with Crick and the microbiologist Carl Woese — proposed that early life on Earth was based on RNA genes and RNA enzymes, rather than DNA genes and protein enzymes as it is today. Fifteen years later, following the Nobel-prizewinning discovery by Thomas Cech and Sidney Altman of RNA enzymes in contemporary biology, Orgel's proposal came to be known as the 'RNA world' hypothesis.

I was a graduate student in the Orgel laboratory when word came that Cech had discovered self-splicing RNA enzymes. No work was done at the bench that day; everyone was at the blackboard, as Orgel led a rapid-fire discussion of how an RNA enzyme might bring about the replication of RNA, and of how such an enzyme might arise. At afternoon tea — a tradition in the Orgel laboratory — Crick joined the conversation and the pace quickened. There was no crowing from Orgel that he had predicted catalytic RNA, and no complaining that he had overlooked the possibility that RNA enzymes might still exist today — only the joy of fresh scientific ideas triggered by new experimental facts.

Although Orgel was a theoretician, he always demanded that theory be subject to rigorous experimental validation. This, he felt, was especially true in the field of the origins of life, where "theories are a dime a dozen and facts are in short supply". He took great pleasure in a positive result, to the point of rooting for the pen on a graph-plotter during chromatography experiments. But he also delighted in negative results, because they pushed him to devise new hypotheses. This, of course, is the way scientists are supposed to behave, but Orgel was one of the few who actually did so.

Following the discovery of catalytic RNA, Orgel continued to pursue the RNA-world hypothesis as both a strong proponent and a tough critic. He pointed out that the notion of an RNA world hardly solves the problem of the origins of life, and suggested that RNA was preceded by some other genetic material, just as DNA and protein were preceded by RNA. Many of his later publications concerned experimental studies of possible pre-RNA-world molecules. His theories brought him into conflict with creationists, who sometimes quoted Orgel out of context, pointing to his admitted uncertainty about life's origins as if this were a failing of the scientific approach. It was, of course, typical of Orgel and of the best practice of science. He had no time for proponents of 'intelligent design', and avoided those prone to magical thinking.

When engaged in scientific discourse, Orgel was intellectually demanding, but never unkind. As his student — and over the decades since then — if I were to advance a dunderheaded hypothesis, he would frown slightly and begin his corrective response with the words: "With respect…". Like all those who knew him, I had the deepest respect for Leslie as a scientist, colleague and dear friend.

Gerald F. Joyce
Gerald F. Joyce is in the Departments of Chemistry and Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.
e-mail: gjoyce@scripps.edu

©2007 Nature Publishing Group

PROGRESS

Venus as a more Earth-like planet

Håkan Svedhem[1], Dmitry V. Titov[2,3], Fredric W. Taylor[4] & Olivier Witasse[1]

Venus is Earth's near twin in mass and radius, and our nearest planetary neighbour, yet conditions there are very different in many respects. Its atmosphere, mostly composed of carbon dioxide, has a surface temperature and pressure far higher than those of Earth. Only traces of water are found, although it is likely that there was much more present in the past, possibly forming Earth-like oceans. Here we discuss how the first year of observations by Venus Express brings into focus the evolutionary paths by which the climates of two similar planets diverged from common beginnings to such extremes. These include a CO_2-driven greenhouse effect, erosion of the atmosphere by solar particles and radiation, surface–atmosphere interactions, and atmospheric circulation regimes defined by differing planetary rotation rates.

Venus, Earth and Mars—the three terrestrial planets with atmospheres, grouped close together in the inner solar system—have many features in common. Earth and Venus, in particular, are nearly the same size and seem to have been quite similar in the epoch when they formed and cooled, probably with large inventories of CO_2 in Earth's atmosphere and liquid water oceans on the surfaces of Venus (and Mars). Today they have very different conditions on their surface as a result of evolutionary processes that we try to understand by measuring and modelling the common processes, aided by data from space missions designed to probe the planets and their environments. Earth and Venus have roughly the same amount of CO_2; on Earth it is bound in carbonates in the crust, whereas on Venus it exists mostly as gas. The extreme climate at the surface of Venus, driven by this excess of CO_2 in the atmosphere, reminds us of pressing problems caused by similar physics on Earth.

More than 30 spacecraft have made the trip to Venus since the Americans sent Mariner 2 in 1962—the first successful man-made mission to another planet. The Soviet Venera and Vega and the American Pioneer Venus missions in 1967–92 were particularly influential in establishing a basic description of the physical and chemical conditions prevailing in the atmosphere. They showed the venusian atmosphere to be filled with corrosive gases and thick clouds, extraordinarily active, with high winds and complex cloud formations sculpted by meteorological systems that seemed to defy categorization by terrestrial analogy, and a vast double-eyed vortex over each pole. Now the European Space Agency has sent its first mission to our nearest planetary neighbour, to investigate how the global atmospheric circulation, the cloud chemistry, surface–atmosphere physical and chemical interactions including volcanism, atmospheric escape processes and the global energy balance and the 'greenhouse' effect at the surface all act together to produce a climate apparently defiantly different from Earth's[1].

Venus Express

The Venus Express design is based on the successful Mars Express spacecraft—a 600-kg, three-axis-stabilized platform with a body-fixed communications antenna[2]. It was launched by the Russian Soyuz-Fregat launcher from Baikonur, Kazakhstan, on 9 November 2005 and is the first mission dedicated to atmospheric and plasma investigations of Venus since NASA launched its Pioneer Venus orbiter and probes more than a quarter of a century ago. It arrived at Venus on 11 April 2006 and became fully operational in June of that year, deploying a new generation of instrumentation[2] and using new modes of observation[3]. The core of the payload is composed of optical instruments including spectrometers and spectro-imagers (Fig. 1 and Table 1), which make the first systematic use of the spectral windows between 1 and 3 μm for three-dimensional imaging of the atmosphere all the way down to the surface[4]. Solar, stellar and Earth radio occultation is used for vertical profiling of atmospheric properties. The highly elliptical polar orbit combines global nadir observations of extended duration of the southern hemisphere with close-up snapshots of the equatorial and northern latitudes[3].

Middle and lower atmosphere

These first observations covering a range of depths in Venus's atmosphere at high spatial resolution have revealed, in addition to

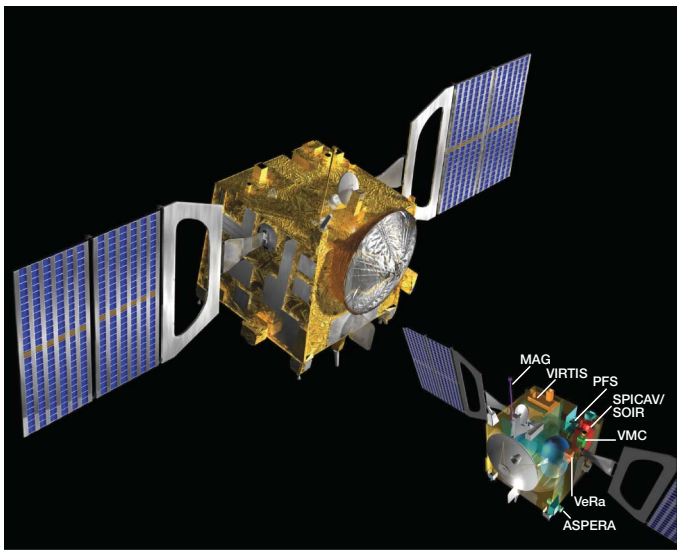


Figure 1 | The Venus Express spacecraft. The inset shows the positions of the seven scientific instruments in a semi-transparent view. The optical instruments for remote sensing are mounted below the upper platform; all apertures are aligned with the $+z$ axis (pointing towards the top of the image). The magnetometer (MAG) on its 1-m-long boom can be seen on the upper platform, while the two ASPERA sensors are mounted on the bottom platform (only one can be seen in this view). PFS, Planetary Fourier Spectrometer; VeRa, Venus Radio Science Experiment; VMC, Venus Monitoring Camera.

[1]ESA/ESTEC, PB 299, 2200AG Noordwijk, The Netherlands. [2]Max Planck Institute for Solar System Research, Max-Planck-Strasse 2, 37191 Katlenburg-Lindau, Germany. [3]Space Research Institute (IKI), Profsojuznaja ul. 84/32, 117997 Moscow, Russia. [4]Department of Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford OX1 3PU, UK.

©2007 Nature Publishing Group

Table 1 | The scientific payload of Venus Express

Name (acronym)	Description	Measured parameters
ASPERA-4	Detection and characterization of neutral and charged particles	Electrons 1 eV–20 keV; ions 0.01–36 keV/*q*; neutral particles 0.1–60 keV
MAG	Dual sensor fluxgate magnetometer, one sensor on a 1-m-long boom	**B** field 8 pT–262 nT at 128 Hz
PFS	Planetary Fourier Spectrometer (currently not operating)	Wavelength 0.9–45 μm; spectral resolving power about 1,200
SPICAV/SOIR	Ultraviolet and infrared spectrometer for stellar and solar occultation measurements and nadir observations	Wavelengths 110–320 nm, 0.7–1.65 μm and 2.2–4.4 μm; spectral resolving power up to 20,000
VeRa	Radio Science investigation for radio-occultation and bi-static radar measurements	X- and S-band Doppler shift, polarization and amplitude variations
VIRTIS	Ultraviolet–visible–infrared imaging spectrometer and high-resolution infrared spectrometer	Wavelength 0.25–5 μm for the imaging spectrometer and 2–5 μm for the high-resolution channel; resolving power about 2,000
VMC	Venus Monitoring Camera for wide-field imaging	Four parallel channels at 365, 513, 965 and 1010 nm

These instruments are expected to produce more than 2 terabits of data during the design lifetime of four Venus sidereal days (about 1,000 Earth days). Venus Express is operating in an elliptical polar orbit with a period of 24 h and an apocentre altitude of 66,000 km. The pericentre altitude is maintained between 250 and 400 km approximately over the north pole. *q* is elementary charge.

localized 'weather' phenomena, the overall organization of the atmospheric circulation. Three broad regimes are clearly present in the middle and lower atmosphere, with convective and wave-dominated meteorology in the lower latitudes and an abrupt transition to smoother, banded flow at middle to high latitudes[5]. The latter terminates at about 30° from the pole, where the cold polar collar discovered by earlier missions lies. This encloses a vast vortex-type structure several thousand kilometres across with a complex double 'eye' that rotates every 2.5–2.8 Earth days. Simultaneous observations in the ultraviolet and thermal infrared spectral ranges show correlated patterns, indicating that the contrasts at both wavelengths, although representing different atmospheric levels, are driven by the same circumpolar dynamical regime[5,6]. Spectroscopic observations indicate marked changes in the temperature and cloud structure in the vortex, with the cloud top in the polar collar located at an altitude of 70–72 km, about 5 km or one scale height higher than in the eye. Night-side observations in the transparent spectral windows showed that the vortex structure and circulation exist at as least as great a depth as the lower cloud deck at 50–55 km, although its 'dipole' appearance seems to be confined to the cloud-top region[6]. The edge of the polar collar at 50–60° latitude apparently marks the poleward limit of the Hadley circulation, the planet-wide overturning of the atmosphere in response to the concentration of solar heating in the equatorial zones (Fig. 2a). Indirect evidence of such meridional circulation is provided by monitoring of the latitude distribution of minor constituents, especially carbon monoxide, as dynamical tracers in the lower atmosphere.

The mesopause on Venus at 100–120 km altitude marks another transition between different global circulation regimes, this time in the vertical. The predominance of zonal super-rotation in the lower atmosphere below the mesopause is replaced by solar to antisolar flow in the thermosphere above, as revealed by non-LTE (non-local thermodynamic equilibrium) emission in the spectral band of O_2 at 1.27 μm that originates from the recombination of oxygen atoms in descending flow on the night side (Fig. 2b). The observed emission patterns are highly variable, with the maximum at about the anti-solar point and the peak altitude at about the mesopause[7]. A mesospheric temperature maximum is observed on the night side[8], produced by adiabatic heating in the subsiding branch of the thermospheric solar to anti-solar circulation.

Sequences of ultraviolet and infrared images have been used to measure the wind speeds at different altitudes by tracking the motions of contrast features in the clouds. Zonal winds at the cloud tops (~70 km) derived from the ultraviolet imaging are in the range $100 \pm 10\ \mathrm{m\,s^{-1}}$ at latitudes below 50° (ref. 5), in good agreement with the earlier observations[9,10]. The new data, which penetrate the bright upper haze obscuring the main cloud at middle latitudes, find that the cloud-top winds quickly decline poleward of 50°. The infrared observations[6] sound the dynamics in the main cloud deck at ~50 km altitude on the night side, finding strong vertical wind shear of about $3\ \mathrm{m\,s^{-1}\,km^{-1}}$ below 50°, and no shear poleward of this latitude, when compared with the higher-altitude ultraviolet-derived winds. The wind velocity profiles on Venus are found to be roughly, although not exactly, in agreement with those predicted by the cyclostrophic

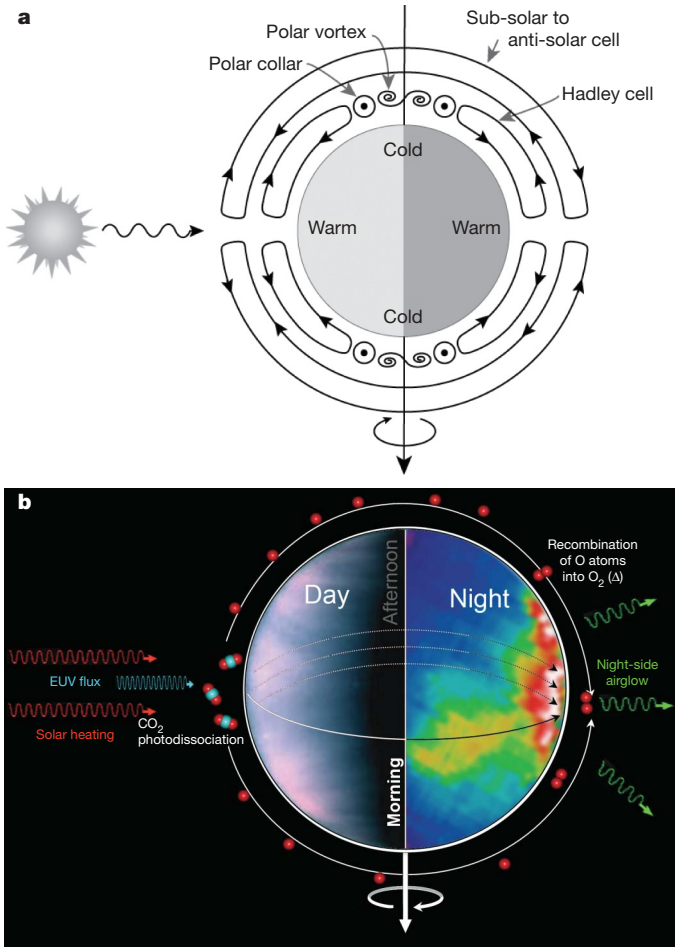


Figure 2 | Schematic view of the general circulation of Venus's atmosphere. **a**, The main feature is a convectively driven Hadley cell, which extends from the equatorial region up to about 60° of latitude in each hemisphere. The trend is polewards at all levels that can be observed by tracking the winds (at about 50–65 km altitude above the surface), so the return branch of the cell must be in the atmosphere below the clouds. A cold 'polar collar' is found around each pole at about 70° latitude; the Hadley circulation evidently feeds a mid-latitude jet at its poleward extreme, inside which there is a circumpolar belt characterized by remarkably low temperatures and dense, high clouds. Inside the collar a thinning of the upper cloud layer forms a complex and highly variable feature, called the 'polar dipole' in earlier literature describing poorly resolved observations, which appears bright in the thermal infrared[6]. Because in general terms thinner-than-average or lower-than-average cloud is often associated with a descending air mass, and vice versa, the vortex may represent a second, high-latitude circulation cell, resembling winter hemisphere behaviour on Earth. **b**, Above about 100 km altitude the circulation regime on Venus changes completely to a sub-solar to anti-solar pattern. Oxygen airglow emission at 1.27 μm reveals the recombination of oxygen atoms into molecular oxygen while descending to lower altitudes in the anti-solar region. Additional evidence of this circulation is given by the upper-atmosphere temperature profiles, which show a pronounced temperature maximum on the night side that is due to compressional heating in the downward branch of the circulation cell[8].

©2007 Nature Publishing Group

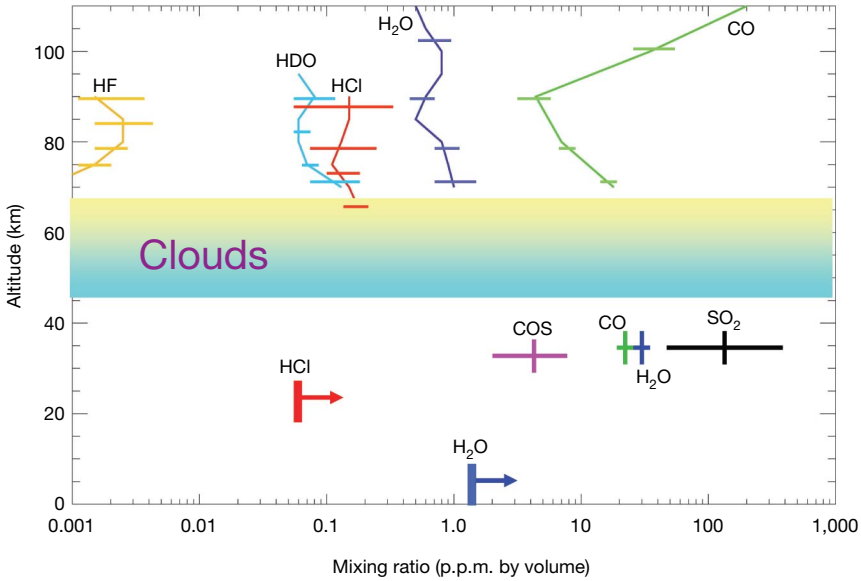


Figure 3 | Atmospheric composition from the Venus Express observations. The colours mark different trace gases. The vertical profiles of H_2O, HDO, CO, HCl and HF above the clouds have been derived from SPICAV/SOIR solar occultation measurements[8]; the abundances of H_2O, CO, SO_2 and COS below the clouds are derived from VIRTIS spectra. The error bars in mixing ratios indicate the minimum and maximum detections over all latitudes, and the error bars in altitudes for the lower atmosphere indicate the width of the weighting functions used for deriving the altitude. The bars with arrows in the lower atmosphere show the expected sensitivity of the Venus Express measurements for which data analysis is still in progress.

approximation[9], which postulates a balance of pressure gradient and centrifugal force in a purely zonal flow. This is, as expected, in contrast with Earth, where so-called geostrophic balance generally applies, dominated by the Coriolis forces produced by the relatively rapid rotation of Earth.

Figure 3 summarizes the results so far of composition measurements by Venus Express[8]. In the deep atmosphere, the most remarkable result is the distribution of CO, which shows a larger systematic latitudinal variability than the other minor species observed, including water vapour. With a large source high in the atmosphere from the photolysis of CO_2 and sinks in the clouds and near the surface, and a lifetime measured in weeks, CO turns out to be an excellent tracer of the general circulation. VIRTIS (Visible and Infrared Thermal Imaging Spectrometer) global maps show a well-defined maximum in the abundance of CO at a latitude of about 60°, near the outer edge of the cold collar, which probably marks the poleward extent of the Hadley cell (Fig. 2a). The other species seem to be much more uniformly distributed over the globe in the deep atmosphere but may be variable at the ≤10% level. The clouds show tremendous variability, with a variety of meteorological systems in several different layers, and distinct regions in which different mean particle sizes predominate. There are, for instance, distinctly larger particle sizes in the clouds in the polar region, although no definite departure from the composition of sulphuric acid and water has been detected.

Upper atmosphere and plasma environment

The absence of an internal magnetic field for Venus means that the solar wind interacts directly with the upper atmosphere, leading to a different distribution of the energies and densities of electrons, ions and energetic neutral atoms from that around Earth[11,12] (Fig. 4). Venus Express measurements are taken at solar minimum, thus complementing the Pioneer Venus plasma studies that were acquired during solar maximum. Photoelectrons with a typical energy of 22–28 eV are measured *in situ* when the satellite dips into the ionosphere while passing the pericentre at 250–350 km. Below this altitude the vertical distribution of electron density is sounded by radio occultation[13], suggesting a stable bottom of the ionosphere at 120 km. The electron density peaks at about $4 \times 10^5\ cm^{-3}$ at about 140 km altitude, and a very dynamic topside ionosphere is observed.

Simultaneous measurements of the vertical profiles of hydrogen-bearing species in the upper atmosphere and plasma *in situ* monitoring have begun to characterize the escape processes that have been responsible for the depletion of water on Venus over the planet's history. Earlier measurements established a D/H ratio ~150 times the terrestrial value in the lower atmosphere[14], which is consistent with the long-term loss of much larger amounts of hydrogen—presumably from water—from Venus compared with Earth. Still higher values of D/H, up to twofold higher, are now being found above the clouds by SPICAV/SOIR (Spectroscopy for Investigation of Characteristics of the Atmosphere of Venus/Solar Occultation at Infrared), which has also uncovered strong variability in both H_2O and HDO content[8]. This unexpected behaviour has been

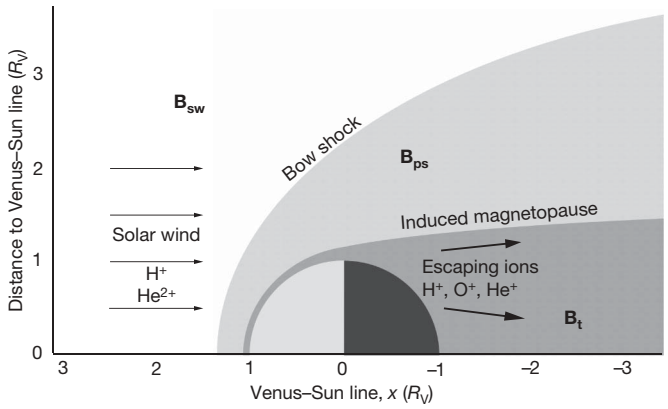


Figure 4 | The plasma environment of Venus as determined by Venus Express. All parameters noted in the figure are measured on a regular basis by the magnetometer and the ASPERA instrument, in three distinctly different regions: the unperturbed solar wind (sw), the plasma sheath (ps) and the induced magnetosphere/tail (t). The boundaries determined by the two instruments are shown approximately to scale. Oxygen ions are observed at high concentrations around the terminator and at lower concentrations well into the tail, indicating escape from a specific source region. He^+ shows similar behaviour, whereas H^+ is observed much more evenly distributed around the planet[11]. The figure shows cylindrical coordinates; the x axis is aligned with the Venus–Sun line. R_V, radius of Venus.

©2007 Nature Publishing Group

tentatively explained by a combination of fractionation in the condensation of ice particles and atmospheric transport, which would imply that species bearing hydrogen and deuterium are already fractionated in the mesosphere right beneath the region from which escape occurs.

ASPERA-4 (Analyser of Space Plasmas and Energetic Atoms 4) has established for the first time the composition of the escaping planetary ions, finding that, after H^+, the main escaping ion is O^+. This is in contrast with Mars, where the escaping plasma consists of approximately equal amounts of O^+, O_2^+ and CO_2^+, and it results from the higher gravitational acceleration at Venus, which tends to retain heavier components such as CO_2^+. High fluxes of escaping He^+ are also detected[11]. Oxygen and hydrogen ions are formed by the dissociation of neutral atmospheric species, including water, by solar ultraviolet radiation, which are then blown away towards the outer reaches of the Solar System (Fig. 4). This happens at a faster rate on Venus than on Earth, not just because Venus is closer to the Sun but also because it lacks the magnetic field that protects Earth from the flux of energetic charged particles from the Sun. These loss processes must have removed large amounts of water from Venus during the first billion years or so after the formation of the Solar System. A detailed quantification of the loss rates enables a more accurate estimate of how much water Venus has lost over its entire history, and by the end of the mission we should know better whether the planet once had an ocean as extensive and deep as Earth's.

Lightning

For a long time the existence of lightning on Venus has been controversial. Whistler-mode waves, which can be considered reliable evidence of lightning, were detected by the Venus Express magnetometer during more than 10% of the pericentre passes[15]. This corresponds to a lightning rate at least half that of Earth. Frequent lightning represents a significant energy input that has important implications for the chemistry in the lower and middle atmosphere on Earth, and this now seems likely to be true for Venus also.

Venus is more Earth-like

The overall sense of the results from the first year of operation of Venus Express is that the differences, particularly in climate, between Venus and Earth are much less mysterious than previously thought after the early phase of spacecraft exploration. They are consistent with theoretical ideas and interpretations suggesting that the two planets had similar surface environments in the past and that they evolved differently, with Earth's oceans converting most of its atmospheric CO_2 to carbonate rocks, and Venus losing most of its water to space. Both processes can now be seen to be still going on. The high zonal winds and near-equatorial turbulence on Venus, as well of course as the high surface temperatures, result from the depth of the atmosphere and huge inventory of greenhouse gas retained by Venus. The slow rotation of Venus, as well as possibly being responsible for the lack of magnetic field that makes erosion of the atmosphere by the solar wind so effective, permits the Earth-like Hadley cell component of the atmospheric circulation to extend closer to the poles, where it breaks down in spectacular fashion to form mid-latitude jets and polar vortices that are larger and more energetic than Earth's but are in many respects quite similar.

1. Taylor, F. W. Venus before Venus Express. *Planet. Space Sci.* **54,** 1249–1262 (2006).
2. Svedhem, H. *et al.* Venus Express—The first European mission to Venus. *Planet. Space Sci.* **55,** 1636–1652 (2007).
3. Titov, D. V. *et al.* Venus Express science planning. *Planet. Space Sci.* **54,** 1279–1297 (2006).
4. Baines, K. H. *et al.* To the depths of Venus: Exploring the deep atmosphere and surface of our sister world with Venus Express. *Planet. Space Sci.* **54,** 1263–1278 (2006).
5. Markiewicz, W. J. *et al.* Morphology and dynamics of the upper cloud layer of Venus. *Nature* doi:10.1038/nature06320 (this issue).
6. Piccioni, G. *et al.* South polar features on Venus similar to those near the north pole. *Nature* doi:10.1038/nature06209 (this issue).
7. Drossart, P. *et al.* A dynamic upper atmosphere of Venus as revealed by VIRTIS on Venus Express. *Nature* doi:10.1038/nature06140 (this issue).
8. Bertaux, J.-L. *et al.* A warm layer in Venus' cryosphere and high-altitude measurements of HF, HCl, H_2O and HDO. *Nature* doi:10.1038/nature05974 (this issue).
9. Gierasch, P. *et al.* in *Venus-II* (eds Bougher, S. W., Hunten, D. M. & Phillips, R. J.) 459–500 (Univ. of Arizona Press, Tucson, 1997).
10. Peralta, J., Hueso, R. & Sánchez-Lavega, A. A reanalysis of Venus winds at two cloud levels from Galileo SSI images. *Icarus* **190,** 469–477 (2007).
11. Barabash, S. *et al.* The loss of ions from Venus through the plasma wake. *Nature* doi:10.1038/nature06434 (this issue).
12. Zhang, T. *et al.* Little or no solar wind enters Venus' atmosphere at solar minimum. *Nature* doi:10.1038/nature06026 (this issue).
13. Pätzold, M. *et al.* The structure of Venus' middle atmosphere and ionosphere. *Nature* doi:10.1038/nature06239 (this issue).
14. Donahue, T. M., Grinspoon, D. H., Hartle, R. E. & Hodges, R. R. (eds Bougher, S. W., Hunten, D. M. & Phillips, R. J.) 385–414 (Univ. of Arizona Press, Tucson, 1997).
15. Russell, C. T. *et al.* Lightning on Venus inferred from whistler-mode waves in the ionosphere. *Nature* doi:10.1038/nature05930 (this issue).

Acknowledgements We thank R. Hueso and J. Bailey for the provision of graphics and data for Fig. 2b, and E. Marcq, C. Tsang, P. Drossart and J.-L. Bertaux for providing data for Fig. 3.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence should be addressed to H.S. (h.svedhem@esa.int).

©2007 Nature Publishing Group

LETTERS

Morphology and dynamics of the upper cloud layer of Venus

W. J. Markiewicz[1], D. V. Titov[1,2], S. S. Limaye[3], H. U. Keller[1], N. Ignatiev[2], R. Jaumann[4], N. Thomas[5], H. Michalik[6], R. Moissl[1] & P. Russo[1]

Venus is completely covered by a thick cloud layer, of which the upper part is composed of sulphuric acid and some unknown aerosols[1]. The cloud tops are in fast retrograde rotation (super-rotation), but the factors responsible for this super-rotation are unknown[2]. Here we report observations of Venus with the Venus Monitoring Camera[3] on board the Venus Express spacecraft. We investigate both global and small-scale properties of the clouds, their temporal and latitudinal variations, and derive wind velocities. The southern polar region is highly variable and can change dramatically on timescales as short as one day, perhaps arising from the injection of SO_2 into the mesosphere. The convective cells in the vicinity of the subsolar point are much smaller than previously inferred[4–6], which we interpret as indicating that they are confined to the upper cloud layer, contrary to previous conclusions[7,8], but consistent with more recent study[9].

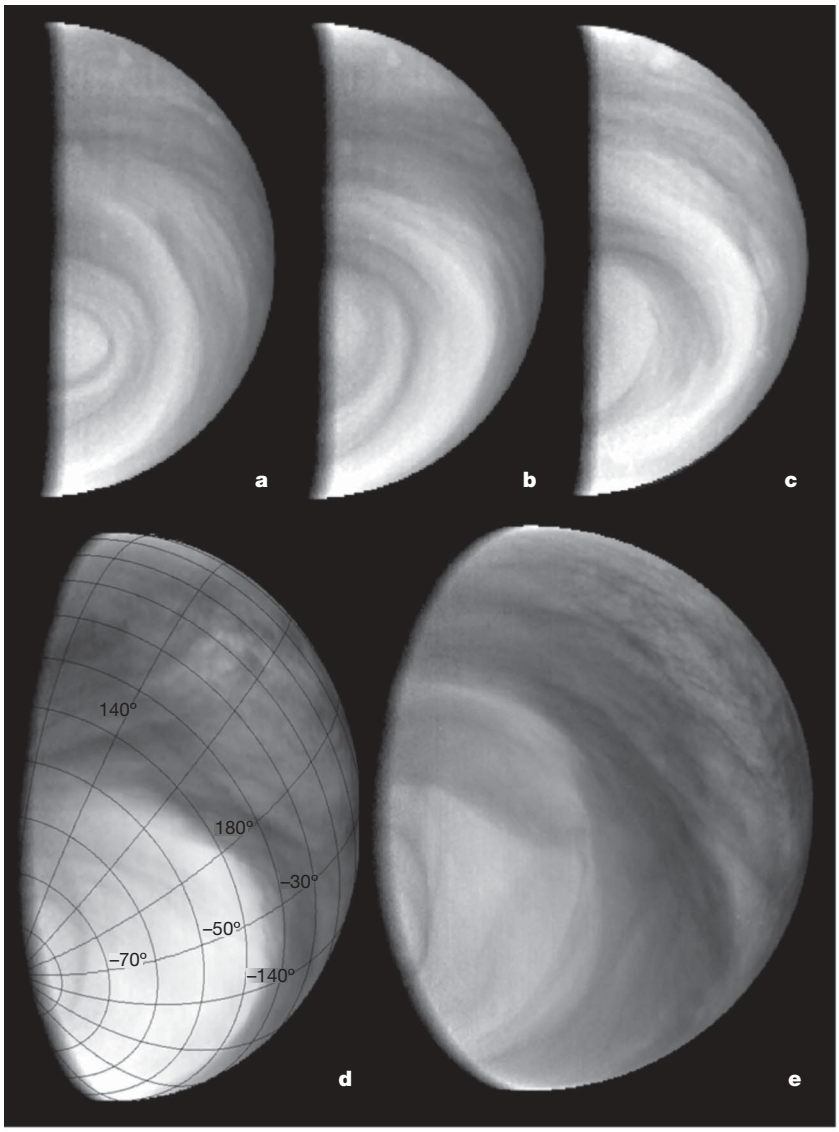


Figure 1 | Variability of the clouds and hazes in the south polar region. On 11 April 2006, the Venus Express spacecraft was inserted into an elliptical polar orbit around Venus with a 24-hour period. The mission aims at a global investigation of the Venus atmosphere by a powerful suite of remote sensing instruments[6], including the VMC, which is a wide-angle digital camera, to investigate the cloud morphology and dynamics of the visible cloud tops. The VMC takes images in four narrow spectral bands—in ultraviolet (365 nm), visible (513 nm) and near-infrared (965 and 1,010 nm)—with spatial resolution from 50 km per pixel down to 200 m per pixel. The ultraviolet channel is centred at the characteristic wavelength of the mysterious ultraviolet absorber, for which the observed contrasts are maximal. This figure shows images of Venus captured by the VMC in the ultraviolet channel. The south pole is at the terminator in the bottom left of the images. Super-rotation is in the anticlockwise direction. All five images are corrected for the solar incidence angle to enhance the near-terminator region. **a**–**c**, These images are from orbits 67 (**a**), 68 (**b**) and 70 (**c**). The period of super-rotation varies between three and five days, depending on the latitude (compare Fig. 2). Strong variability of the near-polar features from day to day is obvious. An extreme case of such variability is shown in **d**: the bright haze located above the absorbing cloud layer expanded to ~35° S, and also became denser, masking all the features except the oval. Temporal variations in the polar regions of Venus were previously observed by the Pioneer Venus Orbiter Cloud Photopolarimeter (OCPP) experiment at various timescales[16]. The timescale of the observed changes over about one day in the appearance of polar region is very short. Panel **e** was obtained from the ascending branch of the orbit, and shows the full view of the southern hemisphere from equator (right) to the pole.

[1]Max-Planck-Institut für Sonnensystemforschung, D-37191 Katlenburg-Lindau, Germany. [2]Space Research Institute (IKI), 117997 Moscow, Russia. [3]Space Science and Engineering Center, University of Wisconsin, Madison, Wisconsin 53706, USA. [4]Institut für Planetenforschung, DLR, 12489 Berlin, Germany. [5]Physikalisches Institut, CH-3012 Bern, Switzerland. [6]IDA, TU Braunschweig, D-38106 Braunschweig, Germany.

©2007 Nature Publishing Group

The Venus Express orbit provides an opportunity both to image the South pole in nadir geometry and to obtain high-resolution close-up views of the Northern hemisphere. Figure 1 shows global views of the Southern hemisphere taken by the ultraviolet channel of the Venus Monitoring Camera (VMC) from a distance of 60,000–40,000 km. A bright mid-latitude band separates darker low latitudes from the polar regions, in which global streaks indicate atmospheric parcels spiralling towards the pole. The polar vortex, originally discovered at thermal infrared wavelengths[10], is now observed by the VMC as the dark oval ultraviolet feature (Fig. 1a). The cloud pattern at high latitudes very much resembles the structure of terrestrial hurricanes, but the Venus polar vortex is three to four times larger. In the lower latitudes the cloud morphology is mottled, indicating vigorous convection (see below).

We find that the south-polar region is highly variable on short timescales. Figure 1a–c shows images taken in three nearly consecutive orbits. They show significant changes in the polar cloud pattern on timescales of one day or less. The observed variations are manifestations of global atmospheric motions and wave activity or changes in the properties of the bright upper haze. On 13 January 2007 (orbit 267), the VMC observed a dramatic change of this type (Fig. 1d). Over several days, from 9 to 13 January, the brightness of the southern polar region increased by about 30%. The bright polar haze, the boundary of which is usually located at ~55° S, expanded to ~35° S. Dark streaks commonly seen in the polar region almost disappeared, indicating that the total opacity of the haze veil exceeded unity. The polar contrast features began reappearing again one day later.

The processes of diffusion growth of sulphuric acid particles, eddy mixing and sedimentation have timescales of a few months at these altitudes and thus do not seem to be relevant[1]. Homogeneous nucleation can lead to quick formation of large amounts of new particles in the submicrometre size range ($r \approx 0.1$ μm), providing the system reaches high values of supersaturation (~10–100). In the case of the Venus upper atmosphere, this implies either abrupt cooling or injection of large amounts of SO_2 molecules in this region. Both mechanisms imply some changes in the dynamical regime of high and mid-latitudes. Subsequent clearing can be explained by the quick decrease of haze opacity due to coagulation, a process that effectively decreases the number of particles and opacity. The coagulation timescale is $t_{coag} \approx 2/Kn$, where n is the number density and K is the coagulation kernel. For the brownian coagulation[1], $K \approx 10^{-9}\,cm^3\,s^{-1}$. VMC observations of clearing of the polar regions during one day ($\sim 10^5$ s) indicate that the aerosol number density in the upper haze is $n \approx 10^4\,cm^{-3}$. VMC observations show that in the peak of brightening on 13 January the upper haze obscured the cloud top, which implies an opacity of $\tau \approx 1$. We can also estimate the geometrical thickness H of the upper haze layer. The opacity is $\tau \approx \pi r^2 QnH$,

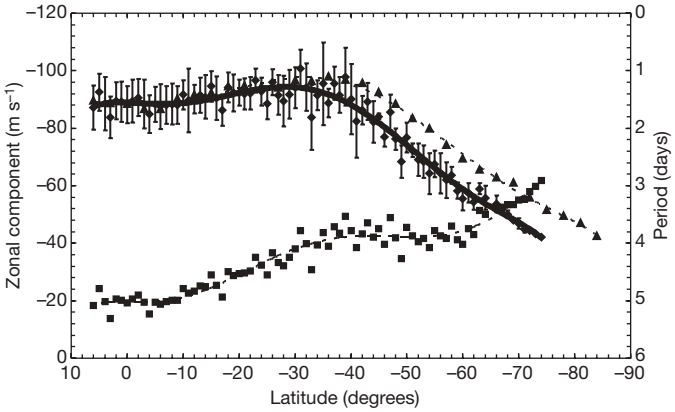


Figure 2 | Wind speeds from tracking cloud features. Latitudinal profile of the east–west component (the negative sign implies the same sense as the rotation of the solid planet) of the cloud motions obtained from digital tracking of features observed by the VMC from orbits 263–267 (solid and dashed curves). The solid curve shows the profile when only the vectors between 11:00 and 13:00 Local Solar Time are averaged, whereas the dashed curve includes all local times measured (~9:00 to 15:00) The rotation period is also shown (right-hand scale, lower, dot-dashed curve). The data points represent averages in one-degree-wide latitude bins and the curves show sixth-degree polynomial fits to the data. The fit equations for the zonal component U (in m s^{-1}) and their regression coefficients R are: $U(11{:}00\text{–}13{:}00\,\text{LST}) = 10^{-8}x^6 + 2\times10^{-6}x^5 + 0.0002x^4 + 0.0041x^3 + 0.022x^2 - 0.0758x - 88.976$; $R^2 = 0.9662$ (diamond symbols; solid line). $U(\text{all LST}) = 5\times10^{-9}x^6 + 10^{-6}x^5 + 7\times10^{-5}x^4 + 0.0015x^3 - 0.0121x^2 - 0.1183x - 88.432$; $R^2 = 0.9949$ (triangle symbols; dashed line). The fit equation for the period (in days) and its R value are: $P = -5\times10^{-10}x^6 - 10^{-7}x^5 - 7\times10^{-6}x^4 - 0.0002x^3 - 2\times10^{-6}x^2 + 0.0027x - 5.0148$; $R^2 = 0.9395$ (square symbols; dot-dashed line) Two independent visual tracking estimates were obtained that provide confidence in the wind profile at latitudes between ~10° N and ~−45° S. At higher latitudes the contrast in the ultraviolet images is much reduced and yields far fewer targets for visual tracking. The root-mean-square (r.m.s.) deviation of digital tracking results is typically ~20 m s^{-1}. The vectors were first averaged over each orbit and then averaged together over orbits 263–267. The error bars represent 1 r.m.s. deviation values of the average of the zonal component to each orbit (which in turn is an average of all the points in one-degree-wide latitude bins).The latitude wind profiles show almost constant velocity of 90–100 m s^{-1} at low latitudes in agreement with the earlier measurements. However, above −40° latitude the wind velocity quickly decreases towards the pole. Cyclostrophic balance from thermal structure observations from the Pioneer Venus and Venera missions yields strong mid-latitude jets with wind speeds up to 140 m s^{-1} at ~−55° (refs 14, 17). However, the difference may not be significant, given the different time periods of the observations and the insufficient thermal structure data.

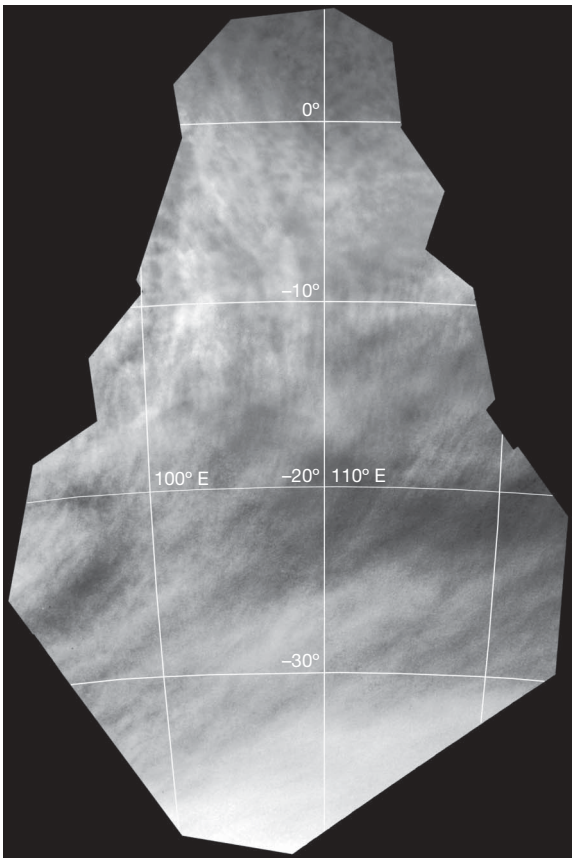


Figure 3 | A mosaic of VMC ultraviolet images showing streaks, wave trains and convection cells (orbit 116). The elongated orbit of Venus Express allows us to zoom into the cloud features while the spacecraft approaches the planet. This mosaic shows that mottled and chaotic cloud patterns at low latitudes give way to approximately zonally oriented streaks at about −15° latitude. This indicates a transition from a dynamical regime that is dominated by local convection at the subsolar point to a quasi-laminar flow.

©2007 Nature Publishing Group

where $Q \approx 2$ is the extinction efficiency. Assuming other values are as derived above, we obtain $H \approx 2$ km, which seems to be a reasonable value.

Tracking the motions of ultraviolet markings was routinely used in the earlier observations to determine wind velocities at the cloud tops[2,5,11–14] (~70 km). The VMC observations have significant advantages in comparison to the earlier missions. First, the imaging sequences have significant duration (~8 h) and good temporal resolution. Second, the Venus Express orbit allows the VMC to extend wind tracking to middle and high latitudes that were hardly accessible by the earlier missions. Figure 2 shows the preliminary results of zonal winds derived from observations in orbits 263–267.

Because the Venus Express pericentre is kept in the altitude range 250–350 km, the highest spatial resolution is less than 200 m per pixel. Figures 4 and 5 show more detailed images of various regions in Fig. 3. Wave trains, located perpendicular to the direction of cloud streaks, are ubiquitously present at the equatorial edge of the bright polar band and indicate vivid wave activity in this transition region.

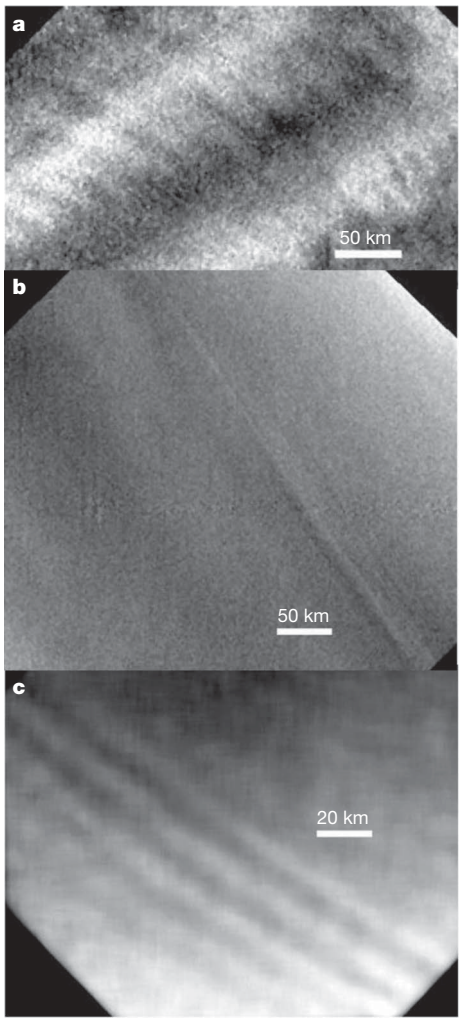


Figure 4 | Three VMC views of Venus clouds with wave-like structures. **a**, An image of a wave train in mid-latitudes at 52° N, −166° E. The typical wavelength of these waves is a few tens of kilometres. **b**, Long linear features such as this one at 54° N, −62° E are occasionally observed, the nature of which is not clear. **c**, A wave train in the polar region (77° N, −51° E). Images **a** and **b** are taken in the ultraviolet channel, and image **c** in the near-infrared. We note that the correlation between the ultraviolet and near-infrared cloud markings that was first evident in Galileo observations[12] frequently appears in the VMC close-up images. This correlation implies that similar mechanisms are responsible for both ultraviolet and near-infrared contrasts and that the unknown ultraviolet absorber absorbs in the near-infrared as well.

Figure 5 shows details of the equatorial region near the subsolar point (local noon). Figure 5a shows what looks like a wave undergoing instability into a more turbulent convection. Figure 5b and c shows more vigorous activity. The morphology of the clouds here is dominated by small-scale cells, but many larger-scale linear features can also be seen. The cellular patterns are believed to be convection cells. They are most prominent in the early afternoon but can also be found near the subsolar point on the morning side. Further from the subsolar point, the patterns return to wavy and streaky morphology even in the equatorial region. The horizontal scale of the convection cells is about 20 km. Because the convection is probably in the upper cloud deck, extending from ~57 to 67 km in altitude, the aspect ratio (diameter/depth) of the cells is about 2. This is five times smaller than the lower limit of aspect ratios reported previously[4–8]. These earlier

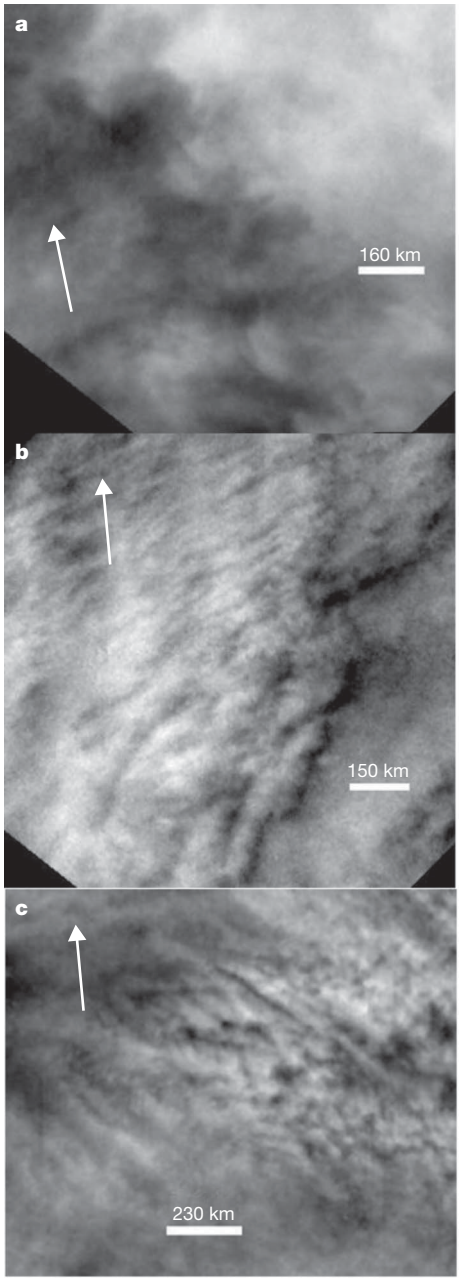


Figure 5 | Three VMC ultraviolet images of the upper cloud deck near the subsolar point. The arrows point to the north. **a**, Wave-like structures transforming into a convective mottled morphology (17° N, −143° E), LST = 14:32. **b**, Streaks and convective cells on the equator (0° N, −150° E), LST = 14:41. **c**, Small convection cells downstream of the subsolar point (10° N, 125° E), LST = 13:16.

©2007 Nature Publishing Group

estimates were based on data with lower temporal and concurrent spatial resolution than is possible with the VMC. We believe that their inferred convection cells were actually clusters of cells that are now individually resolved—as seen in Fig. 5b, c. Studies of penetrative convection[9,15] have investigated the 40–60 km cloud layer, deeper than the VMC can image, and have relied on the existence of the neutral and stable layers seen in the thermal structure revealed from Pioneer probes and confirmed by Soviet landers. These studies showed that convection may penetrate downward into the stable layer from the neutral layer with aspect ratios of ~100 in the subsolar region. Our observations suggest that the convection in the subsolar region occurs on smaller scales and takes place in the upper reaches of the cloud layer, consistent with more recent study[9]. The subsolar region is where Venus' atmosphere absorbs most of the solar radiation. This energy somehow has to be distributed throughout the planet to drive super-rotation. With deep convection, the transport of energy could be efficiently accomplished dynamically. Our present conclusion—that the convection is shallow—requires a reassessment of radiative energy transport from the subsolar region to the rest of the atmosphere. Further analysis should help in understanding the vertical coupling in the Venus atmosphere and the mechanism of super-rotation.

Received 8 May; accepted 20 September 2007.

1. Esposito, L. W. *et al.* in *Venus* (eds Hunten, D. M., Colin, L., Donahue, T. M. & Moroz, V. I.) 484–564 (Univ. Arizona Press, Tucson, 1983).
2. Limaye, S. S. Venus atmospheric circulation: known and unknown. *J. Geophys. Res.* **112,** E04S09, doi:10.1029/2006JE002814 (2007).
3. Markiewicz, W. J. *et al.* Venus monitoring camera for Venus Express. *Planet. Space Sci.* **55,** 1701–1711 (2007).
4. Murray, B. C. *et al.* Venus: Atmospheric motion and structure from Mariner 10 pictures. *Science* **183,** 1307–1315 (1974).
5. Rossow, W. B. *et al.* Cloud morphology and motions from Pioneer Venus images. *J. Geophys. Res.* **85,** 8107–8128 (1980).
6. Covey, C. C. & Schubert, G. Mesoscale convection in the clouds of Venus. *Nature* **290,** 17–20 (1981).
7. Baker, R. D. II & Schubert, G. Cellular convection in the atmosphere of Venus. *Nature* **355,** 710–712 (1992).
8. Belton, M. J. S. *et al.* Cloud patterns, waves, and convection in the Venus atmosphere. *J. Atmos. Sci.* **33,** 1394–1417 (1976).
9. Baker, R. D., Schubert, G. & Jones, P. W. High Rayleigh number compressible convection in Venus' atmosphere: Penetration, entrainment and turbulence. *J. Geophys. Res.* **104,** 3815–3832 (1999).
10. Taylor, F. W. *et al.* Structure and meteorology of the middle atmosphere of Venus: infrared remote sensing from the Pioneer Orbiter. *J. Geophys. Res.* **85,** 7963–8006 (1980).
11. Rossow, W. B., Del Genio, A. D. & Eichler, T. Cloud tracked winds from Pioneer Venus OCPP images. *J. Atmos. Sci.* **47,** 2053–2084 (1990).
12. Belton, M. J. S. *et al.* Images from Galileo of the Venus cloud deck. *Science* **253,** 1531–1536 (1991).
13. Toigo, A., Gierasch, P. J. & Smith, M. D. High resolution cloud feature tracking on Venus by Galileo. *Icarus* **109,** 318–336 (1994).
14. Newman, M. G., Schubert, G., Kliore, A. J. & Patel, I. R. Zonal winds in the middle atmosphere of Venus from Pioneer Venus radio occultation data. *J. Atmos. Sci.* **41,** 1901–1913 (1984).
15. Baker, R. D., Schubert, G. & Jones, P. W. Cloud-level penetrative compressible convection in the Venus atmosphere. *J. Atmos. Sci.* **55,** 3–18 (1998).
16. Sato, M., Travis, L. D. & Kawabata, K. Photopolarimetry analysis of the Venus atmosphere in the polar regions. *Icarus* **124,** 596–585 (1996).
17. Zasova, L. V. Ignatiev, N. Khatuntsev, I. & Linkin, V. Structure of the Venus atmosphere. *Planet. Space Sci.* **55,** 1712–1728 (2007).

Acknowledgements S.S.L. was supported by a NASA grant.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to W.J.M. (markiewicz@mps.mpg.de).

©2007 Nature Publishing Group

LETTERS

South-polar features on Venus similar to those near the north pole

G. Piccioni[1], P. Drossart[4], A. Sanchez-Lavega[5], R. Hueso[5], F. W. Taylor[6], C. F. Wilson[6], D. Grassi[4], L. Zasova[7], M. Moriconi[2], A. Adriani[3], S. Lebonnois[8], A. Coradini[3], B. Bézard[4], F. Angrilli[9], G. Arnold[10], K. H. Baines[11], G. Bellucci[3], J. Benkhoff[10], J. P. Bibring[12], A. Blanco[13], M. I. Blecka[14], R. W. Carlson[11], A. Di Lellis[15], T. Encrenaz[4], S. Erard[4], S. Fonti[13], V. Formisano[3], T. Fouchet[4], R. Garcia[16], R. Haus[10], J. Helbert[10], N. I. Ignatiev[7], P. G. J. Irwin[6], Y. Langevin[12], M. A. Lopez-Valverde[17], D. Luz[4,19], L. Marinangeli[18], V. Orofino[13], A. V. Rodin[7], M. C. Roos-Serote[19], B. Saggin[20], D. M. Stam[21], D. Titov[22], G. Visconti[23], M. Zambelli[1] & the VIRTIS-Venus Express Technical Team*

Venus has no seasons, slow rotation and a very massive atmosphere, which is mainly carbon dioxide with clouds primarily of sulphuric acid droplets. Infrared observations by previous missions to Venus revealed a bright 'dipole' feature surrounded by a cold 'collar' at its north pole[1–4]. The polar dipole is a 'double-eye' feature at the centre of a vast vortex that rotates around the pole, and is possibly associated with rapid downwelling. The polar cold collar is a wide, shallow river of cold air that circulates around the polar vortex. One outstanding question has been whether the global circulation was symmetric, such that a dipole feature existed at the south pole. Here we report observations of Venus' south-polar region, where we have seen clouds with morphology much like those around the north pole, but rotating somewhat faster than the northern dipole. The vortex may extend down to the lower cloud layers that lie at about 50 km height and perhaps deeper. The spectroscopic properties of the clouds around the south pole are compatible with a sulphuric acid composition.

We used the Visible and Infrared Thermal Imaging Spectrometer (VIRTIS) instrument[5,6] on the Venus Express spacecraft to observe the south polar region of Venus. VIRTIS measures radiation intensity at wavelengths between 0.3 and 5 μm. The wavelengths between 3 and 5 μm are sensitive to both temperature and cloud opacity variations. Despite the uncertainties that this introduces, extensive radiative transfer modelling indicates that the real temperature profile in the mesosphere (60–100 km altitude) can be retrieved from VIRTIS data with errors of a few kelvin.

The observations in Fig. 1 show that the atmospheric structure at the south pole as seen in the thermal region of the spectrum at about 5 μm exhibits an inverse 'S' shape, indicating a polar dipole having two centres of rotation in an anticlockwise direction (in agreement with the direction of the atmospheric superrotation), surrounded by a cold collar very much like that observed in the north from past missions. Both the dipole and the collar are thus similar to those previously observed at the north pole by Pioneer Venus[4] but in mirror image (owing to the clockwise rotation of the atmosphere at the north pole). The horizontal gradient of temperature is greatest at an altitude of about 60 km (the lowest layer effectively probed in the thermal region at about 5 μm), with a brightness temperature as high as 250 K observed inside the dipole in contrast with a temperature as low as 210 K observed in the cold collar region. This is shown in Fig. 2, where the real retrieved temperatures are also reported. The detailed, double-eye structure seen in Figs 1 and 2 is not always observed; an oval-shaped dipole with less contrast and variable morphology is more frequently observed in the VIRTIS data from the first few months of Venus Express operations.

Images of the dipole taken during the first 45 days of the spacecraft at Venus indicate that it rotates with a period of -2.48 ± 0.05 days. The negative sign indicates that the dipole rotates in the direction of the solid body of the planet and the superrotating atmosphere, which is retrograde with respect to the rotation of the Earth and most other planets in the Solar System. This rotation is somewhat faster than the rotation of the northern dipole observed from Pioneer Venus in 1979, which exhibited a rotation period varying from -2.79 days to -3.21 days over the 72 days of observation[7]. It could be that the difference is in part due to hemispherical asymmetry, but temporal variability also seems to have an important role. The variability may be caused by external sources—for example, solar influences—or may be indicative of dynamical instability in the atmospheric circulation. The VIRTIS observations reported here were taken sporadically during spacecraft commissioning, so we cannot yet quantify the variability of the dipole rotation rate. However, further observations will help clarify this question.

The analysis of three-dimensional atmospheric thermal fields retrieved from the measured radiances indicates that regions of cold

[1]INAF-IASF (Istituto di Astrofisica Spaziale e Fisica Cosmica), [2]CNR-ISAC (Istituto di Scienze dell'Atmosfera e del Clima), [3]INAF-IFSI (Istituto di Fisica dello Spazio Interplanetario), via del fosso del Cavaliere 100, 00133 Rome, Italy. [4]LESIA, Observatoire de Paris, CNRS, UPMC, Université Paris-Diderot, 5 place Jules Janssen, 92195 Meudon, France. [5]Departamento de Física Aplicada I, Escuela Superior de Ingenieros Universidad del Pais Vasco Alda, Urquijo s/n 48013, Bilbao, Spain. [6]Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK. [7]Space Research Institute of Russian Academy of Sciences (IKI), Profsojuznaja 84/32, 117997 Moscow, Russia. [8]Laboratoire de Meteorologie Dynamique, Jussieu, Box 99, 75252 Paris cedex 05, France. [9]CISAS Università di Padova, via Venezia 1, 35131 Padova, Italy. [10]German Aerospace Center (DLR), Institute of Planetary Exploration, Planetary Physics Section, Berlin-Adlershof Rutherfordstrasse 2, 12489 Berlin, Germany. [11]Jet Propulsion Laboratory, MS 183-601, Pasadena, California 91011, USA. [12]Institut d'Astrophysique Spatiale, Bâtiment 120, Université Paris-Sud, 91405 Orsay Cedex, France. [13]Università degli Studi di Lecce, Dipartimento di Fisica, Via Arnesano, 73100 Lecce, Italy. [14]Space Research Centre of Polish Academy of Science, Bartycka 18A, 00-716 Warszawa, Poland. [15]AMDLSPACE s.r.l., Via Giovanni Angelini 33, 00149 Rome, Italy. [16]Département des Études Spatiales, Institut de Physique du Globe de Paris 4, Avenue de Neptune, F-94107 Saint Maur des Fossés cedex, France. [17]Instituto de Astrofísica de Andalucía (CSIC), Camino Bajo de Huétor, 24 Apartado 3004, 18080 Granada, Spain. [18]International Research School of Planetary Sciences, Dipartimento di Scienze, Universita' d'Annunzio, Viale Pindaro 42, 65127 Pescara, Italy. [19]Observatorio Astronomico de Lisboa, Centro de Astronomia e Astrofisica da Universidade de Lisboa, Tapada da Ajuda 1349-018, Lisboa, Portugal. [20]Politecnico di Milano, Polo di Lecco, Via Marco D'Oggiono 18/A, 23900 Lecco, Italy. [21]Astronomical Institute "Anton Pannekoek", University of Amsterdam, Kruislaan 403, 1098 SJ, Amsterdam, The Netherlands. [22]Max Planck Institute for Aeronomy, Max Planck Str. 2, 37191 Katlenburg-Lindau, Germany. [23]Dipartimento di Fisica, Università de L'Aquila, via Vetoio Loc. Coppito, 67010 Coppito, L'Aquila, Italy.
*A list of participants and affiliations of the VIRTIS-Venus Express Technical Team appears at the end of the paper.

©2007 Nature Publishing Group

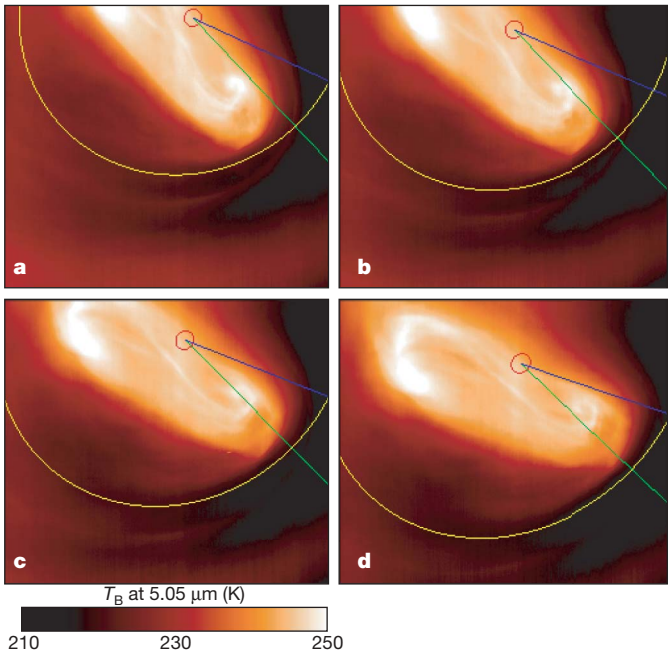


Figure 1 | The Venus south-polar dipole. a–d, A sequence of four false-colour images at a wavelength of 5.05 μm acquired during orbit 38 on 28 May 2006, time 17:41:30, starting from **a** with a time interval of 1 hour; the brightness temperature, T_B, is colour coded as shown. The dipole was clearly seen in these observations with unprecedented detail from a distance of about 60,000 km as it rotates around the pole. The images provided by VIRTIS, when used in high spatial resolution mode such as in this figure, have a size of 256 × 256 pixels and each pixel has an instantaneous field of view of 250 μrad. Major and minor axis dimensions of the dipole are about 2,700 km and 900 km, respectively. Radiative transfer calculation sets the main source of the radiance at this wavelength as coming from a layer at about 60 km altitude where the maximum contrast is seen for the detailed thermal structure of the dipole. The blue and the green curves in the four panels are the meridians at 330° and 350° of longitude, respectively. The red circle indicates the south pole. The yellow curve is the parallel at −70° of latitude. The cold collar is just beyond this latitude and its temperature is coldest on the right side, where the morning terminator resides. The centre of the dipole in temperature is offset by 4° with respect to the south pole in this image. However we expect some variability, which can be studied over a long term period during the course of the mission.

atmosphere (210 K) are associated with the cold collar at the 100 mbar level (about 65 km altitude). This spatial structure vanishes at higher altitudes (>75 km), where the layers become more spatially isothermal.

Atmospheric vertical temperature gradients between 100 to 50 mbar (about 68 km altitude) appear moderate inside the 'S' shape, in the range 0 to $-3\,\mathrm{K\,km^{-1}}$, whereas strong inversions are observed above the cold collar with values up to $6\,\mathrm{K\,km^{-1}}$. Another weak

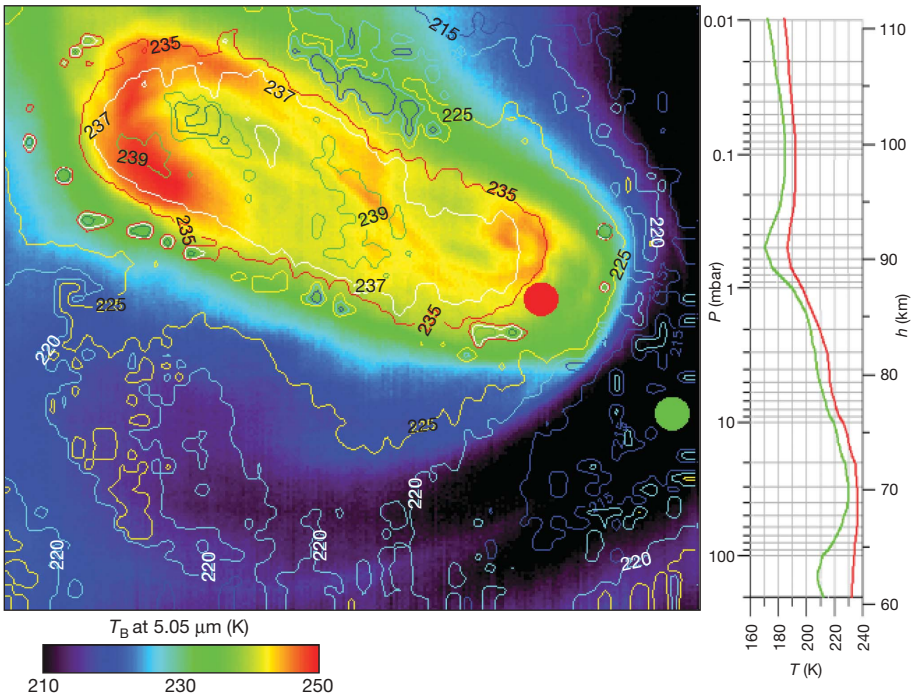


Figure 2 | The dipole in its highest-thermal-contrast atmospheric layer. Left, the colour-coded brightness temperature, T_B, at 5.05 μm of the dipole and the cold collar region. The level curves in the image are the real atmospheric temperatures retrieved from the VIRTIS measured radiances corresponding to a pressure of 200 mbar (about 60 km) and they correlate quite well with the T_B values at this wavelength. The contrast in T_B is maximum at this layer, and the warmer region corresponds to the ellipse containing the dipole where a value of T_B as high as 250 K is observed in the centre of rotation on the left side (real retrieved atmospheric temperature being only a few kelvin less). This part was on the dayside at this time and thus it may be warmed by the solar flux. The overall contrast inside the ellipse is not more than 8 K at 5.05 μm. The spatial thermal gradient is steepest from the cold collar to the inner part on the right side of the right-hand pole of the dipole where a temperature difference of 15 K is observed over a distance of the order of 200 km, just outside the edge of the ellipse. The coldest region of the polar collar with temperatures of about 210 K is registered on the right part, on the morning terminator side. Right, the atmospheric vertical temperature profile in two different regions. The green curve is retrieved in the green spot on the left image, inside the polar collar region. This is also the region where the more prominent thermal inversion is observed in the range from 60 to 70 km altitude. The red curve is from the red spot inside the dipole where the thermal inversion is much more limited. The image also reveals filaments in the circular shape following the rotation of the dipole in the outer part of the ellipse, and also a fine detailed structure in the inner part where the two bright features seem to be connected together.

©2007 Nature Publishing Group

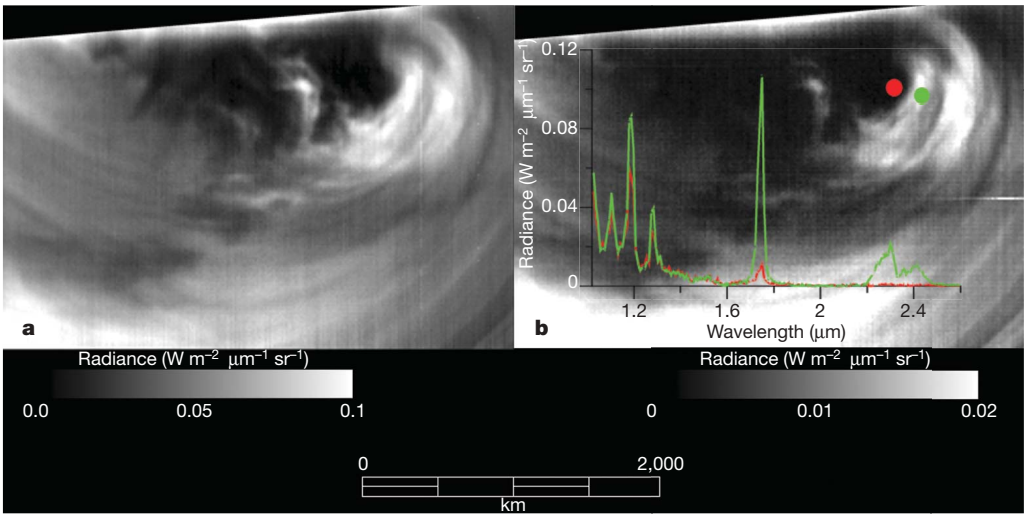


Figure 3 | The deepest view of the dipole seen so far. a, b, Nightside images at 1.74 μm (**a**) and 2.3 μm (**b**) selected from spectral image cube number 4 of orbit 38; radiance is coded according to the grey scale shown. The thermal radiation at these wavelengths mainly originates below the cloud layers, so the structure is produced by spatial variations in the opacity of the clouds. The morphology of the vortex here is remarkable and presents many details. The circular features falling in the region of the cold collar are probably due to clouds elongated by the strong winds coming from the superrotation. The same cloud structures are seen in both spectral regions, with differences in the contrast of the features. The reason for the lower signal at 2.3 μm can be attributed to the different number density or size of the cloud particles[17,18]. The single scattering albedo for sulphuric acid mode 2 particles is $\omega_0 = 0.99978$ and 0.988, respectively, for these spectral windows. At 1.74 μm wavelength the signal decreases and at 2.3 μm it becomes dramatically low, sometimes disappearing altogether when the cloud optical depth increases. This is shown in two spectra in **b** related to two different regions, bright (green circle and green spectrum) and dark (red circle and red spectrum), see Supplementary Fig. 1 for further information. The rising of the continuum radiance on the left side of the spectrum (shorter wavelengths) is due to contamination by scattered light coming from the dayside. The behaviour of the windows corresponds to the variation of the cloud optical depth: the intensity does not change significantly in the window with conservative scattering (for wavelengths lower than 1.27 μm, the absorption optical depth of the cloud deck is less than unity) but changes strongly in the 2.3 μm window. The whole of the dipole is not visible at these wavelengths, because the strong reflected solar light from the higher levels on the dayside does not allow the detection of the deeper levels of the structure beyond the terminator. The three-dimensional view of the vortex seen in these images is not typical, and it may have been enhanced by the relatively high emergence angle of these observations, about 40° in the centre and ranging from 20° to 60° from the top left to the bottom right corner.

maximum in vertical atmospheric temperature profiles is observed around 0.2 mbar (about 95 km altitude), with a peak temperature ranging from 185 to 195 K and apparently no clear spatial organization. The coldest side of the cold collar is identified on the morning terminator side, similarly to the one observed on the north pole. The centre of the dipole from a thermal point of view, as shown in Fig. 1, is 4° offset with respect to the geometric south pole, although this displacement evidently varies with time. The structure of the dipole at different cloud heights and wavelengths is shown in Supplementary Fig. 1.

Figure 3 shows an image of the deep atmospheric structure of the vortex in two of the spectral 'window' regions, at wavelengths near 1.74 and 2.3 μm. Radiative transfer models indicate that the radiation at these wavelengths originates about 10–20 and 20–30 km above the surface, respectively[8,9]. The spatial contrast structure is produced as this radiation passes through the clouds, and is attenuated by cloud layers with differing optical depths[10–12]. It was found earlier from Venera 15 infrared spectrometry that sulphuric acid is the main component of the clouds in the north-polar region[13]. From Mie theory calculations, at wavelengths $\lambda < 1.27$ μm the single scattering albedo of sulphuric acid particles is $\omega_0 > 0.9999$, compared to $\omega_0 = 0.988$ at $\lambda = 2.35$ μm. The relative variation of radiance in the near-infrared spectral windows (as shown in Fig. 3 and Supplementary Fig. 1) is compatible with a sulphuric acid composition of the clouds. The increasing thickness of the clouds does not greatly influence the conservative scattering part of the spectrum ($\lambda < 1.27$ μm), as may be seen in Fig. 3 and Supplementary Fig. 1, panel i. This can be considered as the first indication of the sulphuric acid composition of the clouds in the south polar region.

The polar vortex region is typically very cloudy, even though there are occasionally 'holes' where it is possible to probe very deeply at a wavelength of 1.74 μm. The outline of the dipole seen in the thermal region corresponding to the upper cloud layer overlaps with the pattern seen in the near infrared, which is largely modulated by the opacity of the lower and middle cloud layers, but the structure is somewhat different (Fig. 3). The images probing the deep atmosphere show that the vortex may extend as far as the base of the cloud layer at about 50 km height, and perhaps deeper.

VIRTIS observations usually show a strong correlation between the details observed at about 5.0 μm and 3.8 μm, suggesting that the radiance at these wavelengths may depend on the thermal emission from clouds at about the same level. In contrast, simultaneous observations at ultraviolet and infrared wavelengths taken on orbit 29 (19 May 2006; Supplementary Fig. 1) show a strong anticorrelation of the dark ultraviolet cloud features and bright emission regions in the infrared in both the dipole and the spiral arm around it. A possible explanation is that the dark ultraviolet features absorb more solar radiation and are actually warmer than the ultraviolet bright features, but heating by solar radiation is less effective at high latitudes in the polar region. Alternatively, bright regions in the 400 nm images of the high-latitude dayside may also be interpreted as being due to higher clouds located below the thermal inversion layer. Such clouds would have lower temperatures and appear dark in the thermal infrared, and the effective level from which radiation comes in the ultraviolet spectral range would be in this case significantly lower than the previously estimated value around 70 km height[14–16].

Small-scale features in the images taken at 1.74 μm and 5.05 μm during orbit 38 were tracked to derive winds at different vertical depths. The wind velocity measured at 1.74 μm (assumed to originate at 45–50 km altitude) decreases from $50 \pm 3\ \mathrm{m\,s^{-1}}$ at latitude 75° S to $0\ \mathrm{m\,s^{-1}}$ at the pole, implying a meridional wind shear of $(3.2 \pm 0.2) \times 10^{-5}\ \mathrm{s^{-1}}$. This matches the rotation rate of the large-scale dipole feature at this wavelength. Small-scale features observed in the dipole at 5.05 μm (altitude about 60 km) move with velocities

©2007 Nature Publishing Group

that decrease from $35 \pm 5\,\mathrm{m\,s^{-1}}$ at 75° S to $0\,\mathrm{m\,s^{-1}}$ at the pole, implying a meridional wind shear of $(2.2 \pm 0.3) \times 10^{-5}\,\mathrm{s^{-1}}$. Both the upper and lower structures move at the same velocity at latitudes poleward of −80°. The mean vertical wind shear at the dipole external limit at −75° is $(1.5 \pm 0.5) \times 10^{-3}\,\mathrm{s^{-1}}$, assuming that the images taken at 1.74 μm and 5.05 μm are sensing at respective altitudes of 50 km and 60 km. The interesting temporal and structural variability of the dipole will be further analysed over the course of the mission.

Received 8 April; accepted 27 August 2007.

1. Taylor, F. W., McCleese, D. J. & Diner, D. J. Polar clearing in the Venus clouds observed from the Pioneer Venus Orbiter. *Nature* **279,** 613–614 (1979).
2. Elson, L. S. Wave instability in the polar region of Venus. *J. Atmos. Sci.* **39,** 2356–2362 (1982).
3. Taylor, F. W. in *Middle Atmosphere of Venus* (eds Schaeffer, K. & Spankuch, D.) 93–97 (Veroffentlichungen des Forchungsberichs Geo- und Kosmoswischenschaften, Vol. 18, Academie-Verlag Berlin, 1990).
4. Taylor, F. W. *et al.* Temperature, cloud structure, and dynamics of Venus middle atmosphere by infrared remote-sensing from Pioneer Orbiter. *Science* **205,** 65–67 (1979).
5. Drossart, P. *et al.* Scientific goals for the observation of Venus by VIRTIS on ESA/Venus Express mission. *Planet. Space Sci.* (in the press) ⟨http://dx.doi.org/10.1016/j.pss.2007.01.003⟩.
6. Piccioni, G. *et al. VIRTIS: the Visible and Infrared Thermal Imaging Spectrometer* (ESA-SP-1295, ESA Publications Division, Noordwijk, The Netherlands, in the press).
7. Schofield, J. T. & Diner, D. J. Rotation of Venus's polar dipole. *Nature* **305,** 116–119 (1983).
8. Bezard, B. *et al.* The deep atmosphere of Venus revealed by high-resolution nightside spectra. *Nature* **345,** 508–511 (1990).
9. Pollack, J. B. *et al.* Near infrared light from Venus' nightside: A spectroscopic analysis. *Icarus* **103,** 1–42 (1993).
10. Allen, D. A. & Crawford, J. W. Cloud structure on the dark side of Venus. *Nature* **307,** 222–224 (1984).
11. Kamp, L. W., Taylor, F. W. & Calcutt, S. B. Structure of the Venus atmosphere for modelling of night side infrared spectra. *Nature* **336,** 360–362 (1988).
12. Crisp, D. *et al.* The nature of the near-infrared features on the Venus night side. *Science* **246,** 506–509 (1989).
13. Zasova, L. V. *et al.* Structure of the Venus atmosphere. *Planet. Space Sci.* **55,** 1712–1729 (2007).
14. Esposito, L. W. *et al.* in *VENUS II: Geology, Geophysics, Atmosphere and Solar Wind Environment* (eds Bougher, S. W., Hunten, D. M. & Phillips, R. J.) 415–458 (Univ. Arizona Press, Tucson, 1997).
15. Murray, B. C. *et al.* Venus: Atmospheric motion and structure from Mariner 10 pictures. *Science* **183,** 1307–1315 (1974).
16. Suomi, V. E. & Limaye, S. S. Venus — further evidence of vortex circulation. *Science* **201,** 1009–1011 (1978).
17. Grinspoon, D. H. *et al.* Probing Venus's cloud structure with Galileo NIMS. *Planet. Space Sci.* **41,** 515–542 (1993).
18. Carlson, R. W. *et al.* Variations in Venus cloud particle properties — A new view of Venus's cloud morphology as observed by the Galileo Near-Infrared Mapping Spectrometer. *Planet. Space Sci.* **41,** 477–485 (1993).

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

Acknowledgements We acknowledge the work of the entire Venus Express team, which allowed these data to be obtained. We thank ASI, CNES and the other national space agencies that have supported this research.

Author Contributions G.P. and P.D. coordinated the work as Principal Investigators of VIRTIS. A.S.-L. and R.H. contributed to the dynamics calculation and to Supplementary Fig. 1. F.W.T., C.F.W. and L.Z. contributed to the comparison with earlier data and to the finalization of the paper. D.G. contributed to the temperature retrieval model and calculation, M.M. and A.A. contributed to the thermal structure, and S.L. contributed to the coordination of the dynamics study. All authors contributed equally to the planning of the work and data analysis.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.P. (Giuseppe.piccioni@iasf-roma.inaf.it).

***The VIRTIS-Venus Express Technical Team** Eleonora Ammannito[24], Alessandra Barbis[25], Rainer Berlin[26], Carlo Bettanini[27], Angelo Boccaccini[24], Guillaume Bonnello[28], Marc Bouye[29], Fabrizio Capaccioni[30], Alejandro Cardesin Moinelo[30], Francesco Carraro[31], Giovanni Cherubini[25], Massimo Cosi[25], Michele Dami[25], Maurizio De Nino[32], Davide Del Vento[30], Marco Di Giampietro[25], Alessandro Donati[25], Olivier Dupuis[29], Sylvie Espinasse[31], Anna Fabbri[25], Agnes Fave[29], Iacopo Ficai Veltroni [25], Gianrico Filacchione[30], Katia Garceran[29], Yamina Ghomchi[29], Maurizio Giustini[25], Brigitte Gondet[28], Yann Hello[29], Florence Henry[29], Stefan Hofer[33], Gerard Huntzinger[29], Juergen Kachlicki[26], René Knoll[29], Kouach Driss[29], Alessandro Mazzoni[25], Riccardo Melchiorri[29], Giuseppe Mondello[25], Francesco Monti[32], Christian Neumann[33], Fabrizio Nuccilli[24], Jerome Parisot[29], Claudio Pasqui[25], Stefano Perferi[25], Gisbert Peter [26], Alain Piacentino[29], Carlo Pompei[25], Jean-Michel Reess[29], Jean-Pierre Rivet[29], Antonio Romano[25], Natalie Russ[26], Massimo Santoni[25], Adelmo Scarpelli[25], Alain Semery[29], Alain Soufflot[28], Douchane Stefanovitch[29], Enrico Suetta[25], Fabio Tarchi[25], Nazzareno Tonetti[25], Federico Tosi[24] & Bernd Ulmer[26]

Affiliations for participants: [24]INAF-IFSI, 00133 Rome, Italy. [25]Galileo Avionica, Florence, 50013, Italy. [26]DLR, 12489 Berlin, Germany. [27]Università di Padova, 35131, Italy. [28]IAS Orsay, 91405, France. [29]LESIA, Obs. de Paris, 92195 Meudon, France. [30]INAF-IASF, 00133 Rome, Italy. [31]ASI, Rome, 00198, Italy. [32]Techno System Developments, Naples, 80078, Italy. [33]Kayser Threde, Munich, 81379, Germany.

©2007 Nature Publishing Group

LETTERS

A dynamic upper atmosphere of Venus as revealed by VIRTIS on Venus Express

P. Drossart[1], G. Piccioni[2], J. C. Gérard[4], M. A. Lopez-Valverde[5], A. Sanchez-Lavega[6], L. Zasova[7], R. Hueso[6], F. W. Taylor[8], B. Bézard[1], A. Adriani[3], F. Angrilli[9], G. Arnold[10], K. H. Baines[11], G. Bellucci[3], J. Benkhoff[10], J. P. Bibring[12], A. Blanco[13], M. I. Blecka[14], R. W. Carlson[11], A. Coradini[3], A. Di Lellis[15], T. Encrenaz[1], S. Erard[1], S. Fonti[13], V. Formisano[3], T. Fouchet[1], R. Garcia[16], R. Haus[10], J. Helbert[10], N. I. Ignatiev[7], P. Irwin[8], Y. Langevin[12], S. Lebonnois[17], D. Luz[1,19], L. Marinangeli[18], V. Orofino[13], A. V. Rodin[7], M. C. Roos-Serote[19], B. Saggin[20], D. M. Stam[21], D. Titov[22], G. Visconti[23], M. Zambelli[2], C. Tsang[8] & the VIRTIS-Venus Express Technical Team*

The upper atmosphere of a planet is a transition region in which energy is transferred between the deeper atmosphere and outer space. Molecular emissions from the upper atmosphere (90–120 km altitude) of Venus can be used to investigate the energetics and to trace the circulation of this hitherto little-studied region. Previous spacecraft[1] and ground-based[2–4] observations of infrared emission from CO_2, O_2 and NO have established that photochemical and dynamic activity controls the structure of the upper atmosphere of Venus. These data, however, have left unresolved the precise altitude of the emission[1] owing to a lack of data and of an adequate observing geometry[5,6]. Here we report measurements of day-side CO_2 non-local thermodynamic equilibrium emission at 4.3 μm, extending from 90 to 120 km altitude, and of night-side O_2 emission extending from 95 to 100 km. The CO_2 emission peak occurs at ~115 km and varies with solar zenith angle over a range of ~10 km. This confirms previous modelling[7], and permits the beginning of a systematic study of the variability of the emission. The O_2 peak emission happens at 96 km ± 1 km, which is consistent with three-body recombination of oxygen atoms transported from the day side by a global thermospheric sub-solar to anti-solar circulation, as previously predicted[8].

Observations with the Visible and Infrared Thermal Imaging Spectrometer (VIRTIS) on Venus Express[5,9] are obtained at wavelengths in the infrared between 1 and 5 μm in the night side. These data cover the lower atmosphere up to the altitude of the upper cloud top at about 65 km and temperature profiles from the cloud tops to ~90 km by the thermal inversion of the CO_2 4.3 μm band. Also seen in the near-infrared range are night-side airglow emission from O_2 (at 1.27 μm), and the non-local thermodynamic equilibrium (non-LTE) day-side emission by CO_2 (4.3 μm) and CO (4.6 μm). This emission originates in the upper layers of the atmosphere between 90 and 140 km, where pressures and collision rates are low, so that molecules in excited energy states can emit a photon before collisions bring them into thermal equilibrium. These observations were obtained both in limb geometry, where vertical variations are best investigated, and nadir geometry, where horizontal variations can be mapped in two dimensions. The highly eccentric 24 h orbit of Venus Express[10] permits regular mapping of the southern hemisphere, on both the day and the night sides. The VIRTIS pixel size of 0.25 mrad gives a spatial resolution of 15 km on Venus from apocentre, at a distance of about 66,000 km.

The differences in CO_2 day-side emissions in limb and nadir configurations are detailed in Fig. 1, showing the strong enhancement of the 4.3 μm band at the limb, well above the cloud level, first observed by Galileo/Near Infrared Mapping Spectrometer[6]. This is a feature common to the three telluric planets with atmospheres, because Mars, Earth and Venus all exhibit non-LTE CO_2 emission[11–13]. The altitude of peak emission is defined by a competition between the exponentially increasing density of the atmosphere with depth (increasing the population of molecules that can emit) and the increasing collision frequency, which brings the molecules into LTE, quenching this emission. The altitude and the intensity of the peak emission are therefore related to the density structure of the atmosphere, and their observed variations to atmospheric behaviour, opening a new window on atmospheric processes at these altitudes.

A complete description of a non-LTE radiative transfer model that was originally developed for the Martian atmosphere[14] but is relevant to these new observations of Venus can be found in recent updates[6,7]. The model simulates solar absorption and atmospheric thermal emission by CO_2 molecules, including a large number of CO_2 vibrational energy levels. It also includes a microphysical calculation of their state populations, with a detailed treatment of collisional processes. The probability that solar photons absorbed by a CO_2 molecule and re-emitted resonantly will be reabsorbed at another level is included by using the Curtis matrix method. The standard profile of atmospheric temperature and pressure versus height is taken from the Venus International Reference Atmosphere (VIRA)[15]. As shown in Fig. 1, the model predicts the shape and peak altitude of the intensity quite well, and the fit is improved by suitable changes in the temperature–pressure profile, which affect not only the peak

[1]LESIA, Observatoire de Paris, CNRS, UPMC, Université Paris-Diderot, 5 place Jules Janssen, 92195 Meudon, France. [2]INAF-IASF, [3]INAF-IFSI, 100 Via del Fosso del Cavaliere, 00133 Rome, Italy. [4]LPAP-Université de Liège, 5 Avenue de Cointe, B-4000 Liège, Belgium. [5]Instituto de Astrofisica de Andalucia (CSIC), 50 Camino Bajo de Huetor, 18008 Granada, Spain. [6]Universidad del Pais Vasco, Escuela Superior de Ingenieros, Física Aplicada I, C/Alda. Urquijo s/n 48013, Bilbao, Spain. [7]Space Research Institute (IKI), Profsouznaya 84/32, Moscow, Russia. [8]Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK. [9]CISAS Università di Padova, 1 via Venezia, I-35131 Padova, Italy. [10]German Aerospace Center (DLR), 2 Rutherfordstraße, 12489 Berlin-Adlershof, Germany. [11]Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA. [12]Institut d'Astrophysique Spatiale, batiment 121, Université de Paris-Sud, 91405 Orsay, France. [13]Dipartimento di Fisica, Università degli Studi di Lecce, Via Arnesano, 73100 Lecce, Italy. [14]Centrum Badan Kosmicznych Pan (CBK), 00716 Warszawa, Bartycka 18A, Poland. [15]AMDLSPACE, 33 Via Giovanni Angelini, 00149 Rome, Italy. [16]Dynamique Terrestre et Planétaire, CNRS UMR5562, Observatoire Midi Pyrénées, 31400 Toulouse, France. [17]Laboratoire de Météeorologie Dynamique, UPMC, 4 place Jussieu, 75252 Paris cedex 05, France. [18]Universita d'Annunzio, 42 Viale Pindaro, 65127 Pescara, Italy. [19]Lisbon Astronomical Observatory, Tapada da Ajuda, 1349-018 Lisbon, Portugal. [20]Politecnico di Milano, 26 Piazza Leonardo Da Vinci, 20133 Milano, Italy. [21]University of Amsterdam, 403 Kruislaan, 1098 SJ Amsterdam, The Netherlands. [22]Max-Planck Institute for Aeronomy, 2 Max-Planck-Strasse, 37191 Katlenburg-Lindau, Germany. [23]University of l'Aquila, Dipartimento di Fisica, 10 Via Vetoio, 67010 Coppito l'Aquila, Italy.
*A list of participants and affiliations of the VIRTIS-Venus Express Technical Team appears at the end of the paper.

©2007 Nature Publishing Group

altitude and intensity, but also the spectral shape of the observed emission. The fit of the model calculations to the observations is within a margin consistent with the estimated uncertainties in the model due to spectroscopy, input geophysical parameters and numerical approximations. Although a non-LTE retrieval from the observations is still beyond the scope of the present calculations, its future implementation will permit the atmospheric structure to be determined from limb observations.

In addition to the strong limb emissions, the planetary disk can be observed in nadir mode (Fig. 1a). With the radiative transfer model calibrated against the high-accuracy limb observations, future work will be devoted to the interpretation of the nadir emission. In particular, as the emission is sensitive to atmospheric density variations, it will be possible to search globally for density fluctuations at the altitude of peak emission, including those related to wave activity. Limb viewing geometry will still provide the best chance for the local detection of such waves, thanks to the strong enhancement of the emission due to the slant geometry.

Night-side O_2 emissions at 1.27 μm on Venus have been observed and mapped from the Earth[8,16] since their first detection[2]. VIRTIS night-side observations complement these at much higher spatial resolution because the view from the Venus Express polar orbit covers the southern hemisphere from equator to pole. In addition, limb observations from an orbit around Venus give unprecedented access to the vertical distribution of the airglow layer, providing key constraints on the models. Figure 2 shows examples of the spectral and spatial distribution of O_2 emission.

The O_2 airglow in the near-infrared is produced by three-body recombination of oxygen atoms formed on the day side by photodissociation of CO_2 and CO. These oxygen atoms are transported to the night side by the global circulation, where they recombine according to the reaction:

$$O + O + M \rightarrow O_2^* + M \qquad (1)$$

where O_2^* indicates one of the excited states of the O_2 molecule and M is any neutral constituent. A fraction of the O_2 molecules, estimated to be ~7%, is formed directly in the $^1\Delta_g$ metastable state. A substantial fraction of the upper states cascades into the $^1\Delta_g$ state, so that the net efficiency ε of the production of this state in the three-body recombination may be considerably larger. We adopt here a value[8] of $\varepsilon = 0.75$. Quenching of O_2 $^1\Delta_g$ molecules, mostly by collisions with CO_2, can cause non-radiative transitions to the O_2 ground state.

In Fig. 2c, the vertical distribution of the O_2 ($^1\Delta g$) emission rate calculated with a one-dimensional photochemical-diffusive transport model is compared with the limb profile observed on orbit 76 at 35° N. The observations integrate all the O_2 emission along the line of sight, which is accounted for in the model. A downward flux of oxygen atoms is allowed through the upper boundary, and the oxygen atom density decreases with depth below the peak according to reaction (1). The altitude of the peak of the 1.27 μm emission is thus controlled by the competition between vertical transport and recombination, whereas the brightness is proportional to the downward

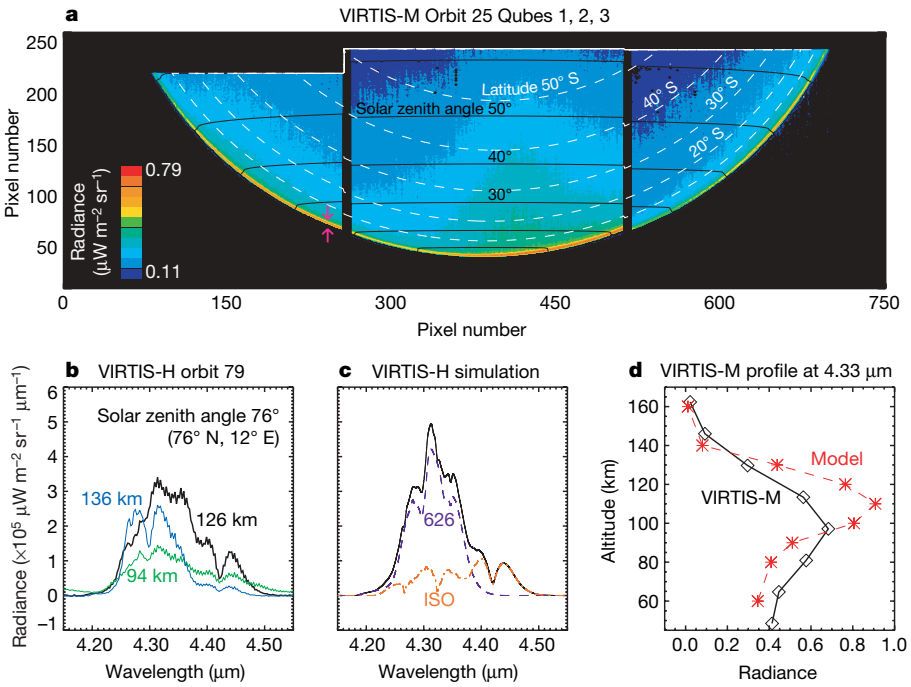


Figure 1 | CO_2 non-LTE emission on the day side of Venus. Example of CO_2 observations around 4.3 μm from VIRTIS. **a**, A VIRTIS-M (mapping channel) image of Venus at 4.33 μm for Orbit 25 composed of three consecutive images (named Qubes 1–3), each with partial disk coverage, showing the CO_2 emission enhanced from nadir to the limb. The variation of the peak emission along the limb is partly due to the solar zenith angle change, but additional variability is also present, which may be due to real atmospheric variability. Radiance (shown by colour scale) is integrated over the VIRTIS-M spectral resolution, which is ~400; the noise level is around 0.06 radiance units. The colour scale is linear between the extreme values, as indicated. **b**, Spectrum of VIRTIS-H (spectroscopic channel) during orbit 79 in a limb geometry at tangent heights and with a spectral resolution of ~1,800, with several emission peaks corresponding to different vibrational–rotational bands of CO_2, whose relative weights depend on the saturation of each band, and therefore on the slant altitude of the observation. The noise radiance is $<1.5 \times 10^4$ units. **c**, Simulations of CO_2 non-LTE emission at 126 km tangent height using a non-LTE radiative transfer model in standard atmospheric condition, adapted to the same illumination conditions and spectral resolution. Contributions from the main CO_2 isotope (abbreviated 626) and from all other minor isotopes (abbreviated ISO) are indicated. Most of the features observed are captured by the model, but it seems to overestimate the peak radiance and to underestimate the relative contribution of the minor isotopes. **d**, An example of a limb emission profile, from the Venus disk at the location shown by the arrows in **a**. The vertical scale is the tangent altitude of the observation, and the highest emission corresponds to ~100 km. Simulation with the non-LTE radiative transfer model explains the limb brightening, but seems to overestimate the peak emission and its altitude.

©2007 Nature Publishing Group

flux of oxygen atoms. The temperature and CO_2 vertical distributions are again taken from the VIRA model[15]. Vertical transport is by molecular diffusion at high altitude and parameterized below the homopause by an eddy diffusion coefficient in the form $K = A/\sqrt{n}\ \mathrm{cm^2\,s^{-1}}$, where A is an empirically estimated parameter and n is the total number density. The best fit to the profile corresponds to $A = 2 \times 10^{12}$ and a downward flux of oxygen atoms of $F_0 = 5 \times 10^{11}\ \mathrm{atoms\,cm^{-2}\,s^{-1}}$. The derived maximum oxygen atom number density, reached at 100 km, is $1.3 \times 10^{11}\ \mathrm{cm^{-3}}$. This quantity has not been measured previously by other techniques.

The derived value of F_0 can be compared with an average production of about $8 \times 10^{12}\ \mathrm{atoms\,cm^{-2}\,s^{-1}}$ on the day side, such that if all atoms were carried to the night side without chemical loss, the average vertical emission rate would be 3 megaRayleigh (one megaRayleigh corresponds to the brightness of an extended source emitting $10^{12}\ \mathrm{photons\,cm^{-2}\,s^{-1}}$ in 4π steradians), that is, about seven times the value deduced from this limb profile. The geometry of the limb emission allows us to distinguish the flux from different layers, so the contribution of the reflection of the O_2 emission by the cloud layer below can be neglected, unlike observations in nadir geometry where a factor of 1.75 enhancement has been derived[8]. The non-homogeneous, time-dependent distribution of the O_2 $^1\Delta_g$ nightglow indicates that the local downward flow of oxygen may differ substantially from the mean value, in response to variations in factors such as the efficiency of the global day-to-night transport, the focusing effect of the night-side subsidence, changing zonal wind speeds, eddy transport efficiency, and gravity wave breaking.

Observations by VIRTIS in the nadir mode (Fig. 3) can be used to construct extensive maps of the Venus atmosphere in the O_2 emission band. In nadir viewing, the contamination of the O_2 emission by the thermal emission of the deeper atmosphere has been subtracted to present pure O_2 airglow images. The O_2 night glow exhibits a large spatial and temporal variability, but the mean value, integrated over the night side of the southern hemisphere, is typically in the range of 0.8 megaRayleigh, which is in agreement with the early ground-based observations[2], which found 1.2 megaRayleigh for the night side. Recent ground-based observers have also seen variability as large as 20% over periods of hours[16].

A new feature of these observations from Venus Express is the ability to observe these non-LTE and airglow phenomena from a vantage point above the South Pole, at an apocentre altitude of 66,000 km. This unique perspective reveals the night-side emission over most of the southern hemisphere, from which the latitudinal distribution, and especially the distribution around local midnight, is obtained. Despite difficulties in interpretation of O_2 variability in terms of bulk velocities, a search to derive the velocity fields from the large body of statistics in the full mission will be undertaken. The current observations are interpreted as showing signatures of the solar to antisolar circulation in the mesosphere. Although observations in the emission bands of O_2 and CO_2 sample similar altitude levels of the atmosphere, we emphasize that the differences in the processes involved give different insights into atmospheric models. The CO_2 emission has a small dependence on atomic oxygen amount and is thought to be mostly regulated by a geographically simple solar illumination parameter, modulated by density variations like those expected from upward-propagating waves. A further dependence on the solar cycle is expected via the O/CO_2 ratio, and possibly via the thermal and density structure of the lower mesosphere. In contrast, the O_2 emission is very sensitive to atomic oxygen amounts, and is driven by the highly variable and poorly understood dynamics of the global thermospheric sub-solar to anti-solar circulation

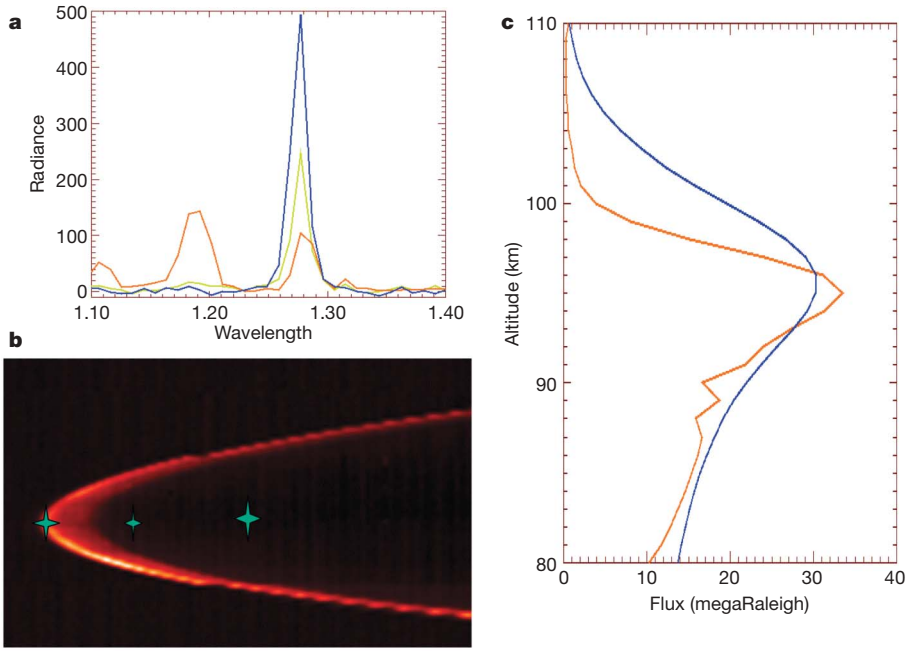


Figure 2 | O_2 emission in the night side upper atmosphere of Venus. **a**, Spectra of O_2 emission at 1.27 μm; the three spectra show regions with dominant O_2 emission (blue), intermediate O_2 emission (green), and dominant thermal emission (red). The O_2 emission is dominant at 1.27 μm, whereas thermal emission makes the major contribution in the adjacent spectral elements. Because the relative influence of the two contributions is very different in adjacent wavelength channels, the lower atmosphere contribution can be subtracted by linear interpolation to get the vertical profile of the O_2 emission to first order. **b**, Image of the limb at 1.27 μm in orbit 76 by VIRTIS-M. In this reconstructed image, the horizontal dimension is the scan direction of VIRTIS-M. The limb was approached in the vertical dimension down to a minimum altitude of ~80 km, then the spacecraft motion was inverted, giving a second limb sounding of the atmosphere. The three stars give the path used for the retrieval of the vertical profiles of **c**. **c**, Limb vertical profile in O_2 emission at a latitude of 35° N, retrieved from the image, comparing the observed emission profile (red), with simulations from a one-dimensional model (blue). Peak altitude and intensities are well reproduced, but a smaller scale height is observed, compared to the model.

©2007 Nature Publishing Group

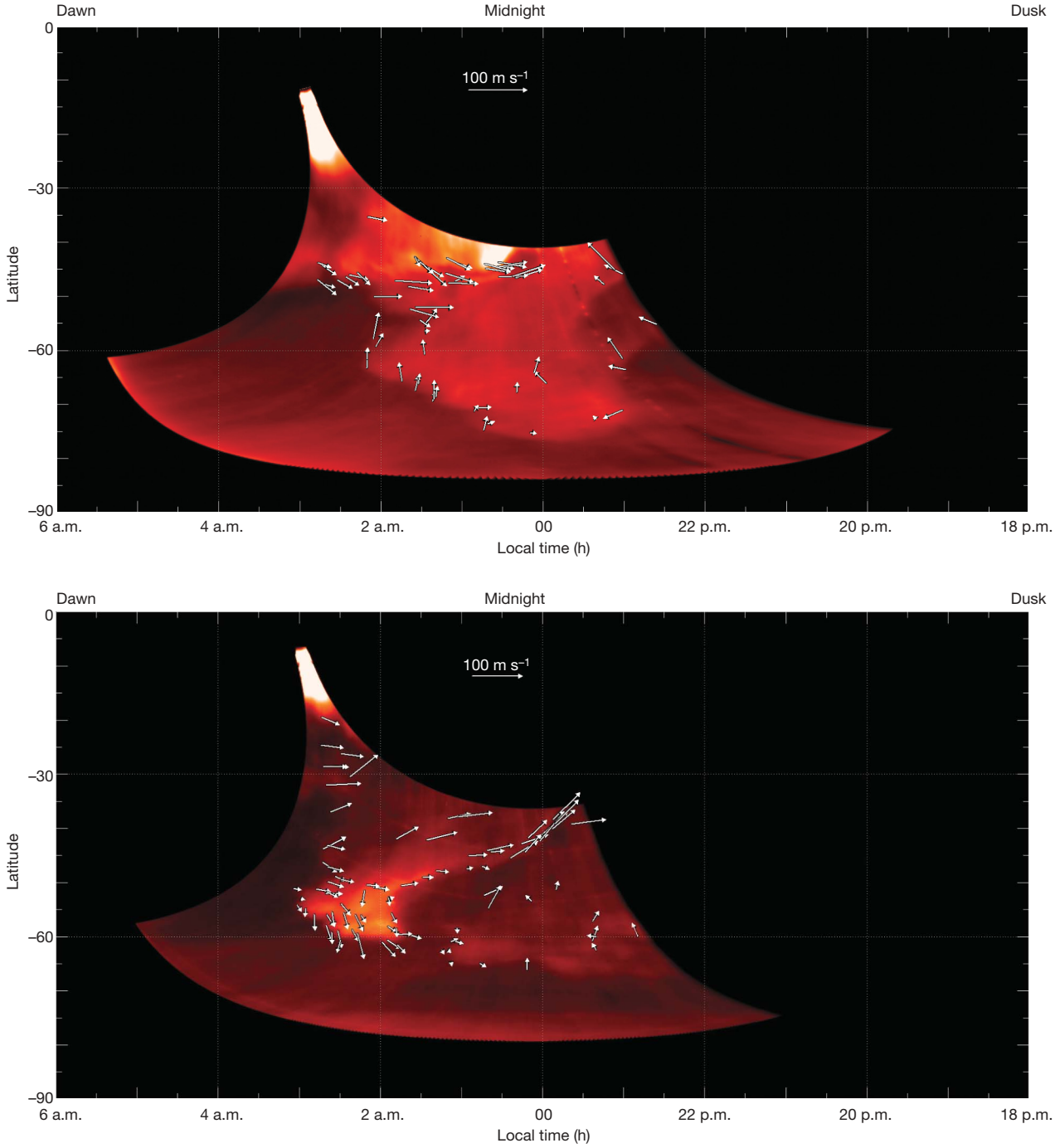


Figure 3 | Apparent motions in the O_2 emission in nadir geometry during orbits 84 (above) and 96 (below). The images have been corrected for thermal emission from the deeper atmosphere, and projected in terms of the local time at each longitude. The maps exhibit almost pure O_2 emission extending in the southern hemisphere from the equatorial zone to high latitudes. For each orbit a sequence of two images separated by ~1 h shows the displacement of the emitted structures. The airglow apparent motions have velocities of the order of 60 m s^{-1}. The flow is dominated by meridional motions from polar latitudes equatorward and by a zonal flow from dawn to midnight. Both are signatures of the solar to antisolar circulation of the mesosphere. Repetition of such images from orbit to orbit shows a high degree of variability and only a statistically large set of observations will provide a global view of the general characteristics of the O_2 emission over the full Venus Express mission.

(Fig. 3). The systematic observations now underway should reveal the detailed characteristics of the latter and its dependence on solar activity.

Received 9 April; accepted 27 July 2007.

1. Bougher, S. W., Alexander, M. J. & Mayr, H. G. Upper atmosphere dynamics: global circulation and gravity waves. In *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment* (eds Bougher, S. W. Hunten, D. M. & Philips R. J.) 259–292 (Univ. Arizona Press, Tucson, 1997).
2. Connes, P. Noxon, J. F., Traub, W. A. & Carleton, N. O_2 $^1\Delta$ emission in the day and night airglow of Venus. *Astrophys. J. Lett.* **233,** L29–L32 (1979).
3. Crovisier, J. *et al.* Carbon monoxide emissions at 4.7μm from Venus' atmosphere. *Planet. Space Sci.* **54,** 1398–1414 (2006).
4. Lellouch, E. *et al.* Monitoring of mesospheric structure and dynamics. In *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment* (eds Bougher, S. W. Hunten, D. M. & Philips R. J.) 295–324 (Univ. Arizona Press, Tucson, 1997).
5. Drossart, P. *et al.* Scientific goals for the observation of Venus by VIRTIS on ESA/Venus Express mission. *Planet. Space Sci.* doi:10.1016/j.pss.2007.01.03 (in the press).

©2007 Nature Publishing Group

6. Lopez-Valverde, M. A., Drossart, P., Carlson, R., Mehlman, R. & Roos-Serote, M. Non-LTE infrared observations at Venus: From NIMS/Galileo to VIRTIS/Venus Express. *Planet. Space Sci.* doi:10.1016/j.pss.2007.01.008 (in the press) (2007).
7. Roldan, C., Lopez-Valverde, M. A., Lopez-Puertas, M. & Edwards, D. P. Non-LTE infrared emissions of CO_2 in the atmosphere of Venus. *Icarus* **147,** 11–25 (2000).
8. Crisp, D. *et al.* Ground-based near-infrared observations of the Venus nightside: 1.27-μm $O_2(^1\Delta_g)$ airglow from the upper atmosphere. *J. Geophys. Res.* **101,** 4577–4594 (1996).
9. Piccioni, G. *et al.* South-polar features on Venus similar to those near the north pole. *Nature* doi:10.1038/nature06209 (this issue).
10. Titov, D. V. *et al.* Venus Express science planning. *Planet. Space Sci.* **54,** 1279–1297 (2006).
11. Lellouch, E. *et al.* The 2.4–45μm spectrum of Mars observed with the infrared space observatory. *Planet. Space Sci.* **48,** 1393–1405 (2000).
12. Picard, R. H., Wintersteiner, P. P., Hegblom, E. R. & Richards, E. Remote sensing of discrete stratospheric gravity-wave structure at 4.3-μm from the MSX satellite. *Geophys. Res. Lett.* **25,** 2809–2812 (1998).
13. Formisano, V., Maturilli, A., Giuranna, M., D'Aversa, E. & Lopez-Valverde, M. A. Observations of non-LTE emission at 4–5 microns with the planetary Fourier spectrometer aboard the Mars Express mission. *Icarus* **182,** 51–67 (2006).
14. Lopez-Valverde, M. A. & Lopez-Puertas, M. A non-local thermodynamic equilibrium radiative transfer model for infrared emissions in the atmosphere of Mars. 1. Theoretical basis and nighttime populations of vibrational states. *J. Geophys. Res.* **99,** 13093–13115 (1994).
15. Keating, G. M. *et al.* Models of Venus neutral upper atmosphere: Structure and composition. *Adv. Space Res.* **5,** 117–171 (1985).
16. Ohtsuki, S. *et al.* Ground-based observation of the Venus 1.27-μm O_2 airglow. *Adv. Space Res.* **36,** 2038–2042 (2005).

Acknowledgements We acknowledge the work of the entire Venus Express team of ESA and Astrium, who allowed these data to be obtained. This work is supported by the national space agencies CNES and ASI, and by SFTC in the UK. J.C.G. thanks the FNRS and the PRODEX-ESA program for funding.

Author Contributions P.D. and G.P. have coordinated the work as Principal Investigators of VIRTIS. J.C.G. contributed to the O_2 model, L.Z. to the O_2 data selection, and M.A.L.-V. to the CO_2 model. A.S.-L. and R.H. contributed to the dynamics calculation for Fig. 3. All authors contributed equally to the planning of the work, instrumental calibration and data calibration. F.W.T. and B.B. helped to finalize 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 P.D. (pierre.drossart@obspm.fr).

***The VIRTIS-Venus Express Technical Team** Eleonora Ammannito[24], Alessandra Barbis[25], Rainer Berlin[26], Carlo Bettanini[27], Angelo Boccaccini[24], Guillaume Bonnello[29], Marc Bouyé[30], Fabrizio Capaccioni[31], Alejandro Cardesin[31], Francesco Carraro[32], Giovanni Cherubini[25], Massimo Cosi[25], Michele Dami[25], Maurizio De Nino[32], Davide Del Vento[24], Marco Di Giampietro[25], Alessandro Donati[25], Olivier Dupuis[30], Sylvie Espinasse[31], Anna Fabbri[25], Agnès Fave[30], Iacopo Ficai Veltroni[25], Gianrico Filacchione[31], Katia Garceran[30], Yamina Ghomchi[30], Maurizio Giustizi[25], Brigitte Gondet[29], Yann Hello[30], Florence Henry[30], Stefan Hofer[33], Gerard Huntzinger[30], Juergen Kachlicki[26], René Knoll[30], Driss Kouach[30], Alessandro Mazzoni[25], Riccardo Melchiorri[30], Giuseppe Mondello[25], Francesco Monti[32], Christian Neumann[33], Fabrizio Nuccilli[28], Jérôme Parisot[30], Claudio Pasqui[25], Stefano Perferi[25], Gisbert Peter[26], Alain Piacentino[30], Carlo Pompei[25], Jean-Michel Réess[30], Jean-Pierre Rivet[30], Antonio Romano[25], Natalie Russ[26], Massimo Santoni[25], Adelmo Scarpelli[25], Alain Sémery[30], Alain Soufflot[29], Douchane Stefanovitch[30], Enrico Suetta[25], Fabio Tarchi[25], Nazzareno Tonetti[25], Federico Tosi[28] & Bernd Ulmer[26]

Affiliations for participants: [24]INAF-IFSI, 00133 Rome, Italy. [25]Galileo Avionica, Florence, 50013, Italy. [26]DLR, 12489 Berlin, Germany. [27]Università di Padova, 35131, Italy. [28]INAF-IASF, 00133 Rome, Italy. [29]IAS Orsay, 91405, France. [30]LESIA, Observatoire de Paris, France. [31]ASI Rome, 00198, Italy. [32]Techno System Developments, Naples, 80078, Italy. [33]Kayser Threde, Munich, 81379, Germany.

©2007 Nature Publishing Group

LETTERS

A warm layer in Venus' cryosphere and high-altitude measurements of HF, HCl, H_2O and HDO

Jean-Loup Bertaux[1,2,5], Ann-Carine Vandaele[3], Oleg Korablev[4], E. Villard[1,2,5], A. Fedorova[4], D. Fussen[3], E. Quémerais[1,2,5], D. Belyaev[4], A. Mahieux[3], F. Montmessin[1,2,5], C. Muller[3], E. Neefs[3], D. Nevejans[3], V. Wilquet[3], J. P. Dubois[1,2,5], A. Hauchecorne[1,2,5], A. Stepanov[4,6], I. Vinogradov[4], A. Rodin[4,7] & the SPICAV/SOIR team*

Venus has thick clouds of H_2SO_4 aerosol particles extending from altitudes of 40 to 60 km. The 60–100 km region (the mesosphere) is a transition region between the 4 day retrograde superrotation at the top of the thick clouds and the solar–antisolar circulation in the thermosphere (above 100 km), which has upwelling over the subsolar point and transport to the nightside[1,2]. The mesosphere has a light haze of variable optical thickness, with CO, SO_2, HCl, HF, H_2O and HDO as the most important minor gaseous constituents, but the vertical distribution of the haze and molecules is poorly known because previous descent probes began their measurements at or below 60 km. Here we report the detection of an extensive layer of warm air at altitudes 90–120 km on the night side that we interpret as the result of adiabatic heating during air subsidence. Such a strong temperature inversion was not expected, because the night side of Venus was otherwise so cold that it was named the 'cryosphere' above 100 km. We also measured the mesospheric distributions of HF, HCl, H_2O and HDO. HCl is less abundant than reported 40 years ago[3]. HDO/H_2O is enhanced by a factor of ~2.5 with respect to the lower atmosphere, and there is a general depletion of H_2O around 80–90 km for which we have no explanation.

The use of solar (at the terminator: the circle on the planet that separates the day side from the night side) and stellar occultation technique (at night), applied for the first time to the atmosphere of Venus with the SPICAV/SOIR spectrometers on board the Venus Express spacecraft, allows us to measure the atmospheric transmission and to derive information about the vertical structure and composition of the 60–140 km region. In this region, many processes (transport, chemistry, temperature, aerosols condensation and evaporation) govern the three-dimensional distribution of haze and chemical species. In the ultraviolet range (110–310 nm) of the SPICAV ultraviolet spectrometer (a copy of the SPICAM ultraviolet instrument in orbit around Mars[4]), the most important absorbers are CO_2 (the main Venus atmospheric constituent) at $\lambda < 200$ nm, and aerosol particles of the haze layer at all wavelengths. The CO_2 local density and temperature profiles are derived by assuming that the atmosphere is in hydrostatic equilibrium[5] (Supplementary Information).

On Fig. 1 are represented the temperature profiles obtained during six stellar occultations performed on the night side. They are compared to previous and scarce measurements[6–9]. All our new profiles show a large temperature excess (30–70 K) with respect to previous measurements, peaked around 100 km. The altitude range 100–150 km has largely been unexplored up to now—the upper limit for infrared soundings and radio occultation is ~100 km, descent probes have so far measured below this range, and drag measurements are made above 150 km. Occultation measurements can probe this region efficiently, allowing the discovery of this hot atmospheric layer. Orbits 95, 96 and 98 have similar profiles, while orbits 102, 103 and 104 have the same kind of profiles, but with an even higher maximum temperature. These three orbits are much closer to the anti-solar point (solar zenith angle, SZA = 167–170°).

We interpret this newly found temperature peak to be caused by adiabatic (or diabatic) heating during air subsidence near the anti-solar point, as the end result of the solar–antisolar circulation pattern suspected to exist in the thermosphere (100–200 km) from the day side to the night side. Such heating was not well predicted by current circulation models, although there was evidence of day-to-night transport and downward vertical transport on the night side: the emission of NO ultraviolet delta and gamma bands already observed by Pioneer Venus[10,11], and O_2 emission at 1.27 μm discovered from the ground[12] (also well detected with VIRTIS/VEX[13]). These emissions occur when O and N atoms (produced by solar extreme-ultraviolet photo-dissociation of N_2 and CO_2 on the day side at 100–120 km) recombine in the night side. However, modelling of NO ultraviolet emission[11,14] described N and O atoms as being transported vertically downwards through the CO_2 background gas by eddy diffusion (turbulence).

In such a description, the CO_2 background gas does not move vertically, and therefore experiences no adiabatic heating. Our observations indicate rather that N and O atoms are advected downwards with the CO_2 background gas during its descent. Such a vertical atmospheric motion is an essential ingredient of the solar–antisolar circulation pattern, which may influence the behaviour of the whole mesosphere, because it implies a compensating upwelling on the day side. The descent velocity may be estimated, given that at temperature $T = 165$ K the night-time infrared cooling rate is about 100 K per day (ref. 2). Assuming a T^4 dependence, it would amount to 377 K per day for a temperature of 230 K as measured here near the anti-solar point. This cooling rate can be compensated by an equal adiabatic heating rate, corresponding to a descent velocity $V = 0.43\ \mathrm{m\,s^{-1}}$, using the relation $\mathrm{d}T/\mathrm{d}t = -\Gamma V$ where t is time and Γ is the adiabatic lapse rate (about $10\ \mathrm{K\,km^{-1}}$; ref. 15). The negative temperature gradient in the layer at 100–120 km is on average equal to $-5\ \mathrm{K\,km^{-1}}$ for orbit profiles 102 to 104 and locally reaches values near the adiabatic lapse rate. This suggests that the layer at 100–120 km is dynamically nearly unstable and that turbulence may occur, inducing a downward heat flux by

[1]Service d'Aéronomie du CNRS/IPSL, Verrières-le-Buisson 91371, France. [2]Université Pierre et Marie Curie, 75252, Paris, France. [3]Belgian Institute for Space Aeronomy, 3 avenue Circulaire, B-1180 Brussels, Belgium. [4]Space Research Institute (IKI), 84/32 Profsoyuznaya, 117810 Moscow, Russia. [5]Institut Pierre Simon Laplace, Université de Versailles-Saint-Quentin, 78 Saint-Quentin en Yvelines, 78280 Guyancourt , France. [6]Faculty of Physics, Moscow State University, GSP-2 119992 Moscow, Russia. [7]Moscow Institute of Physics and Technology, Institutsky dr. 141700 Dolgoprudny, Russia.
*Lists of participants and affiliations appear at the end of the paper.

©2007 Nature Publishing Group

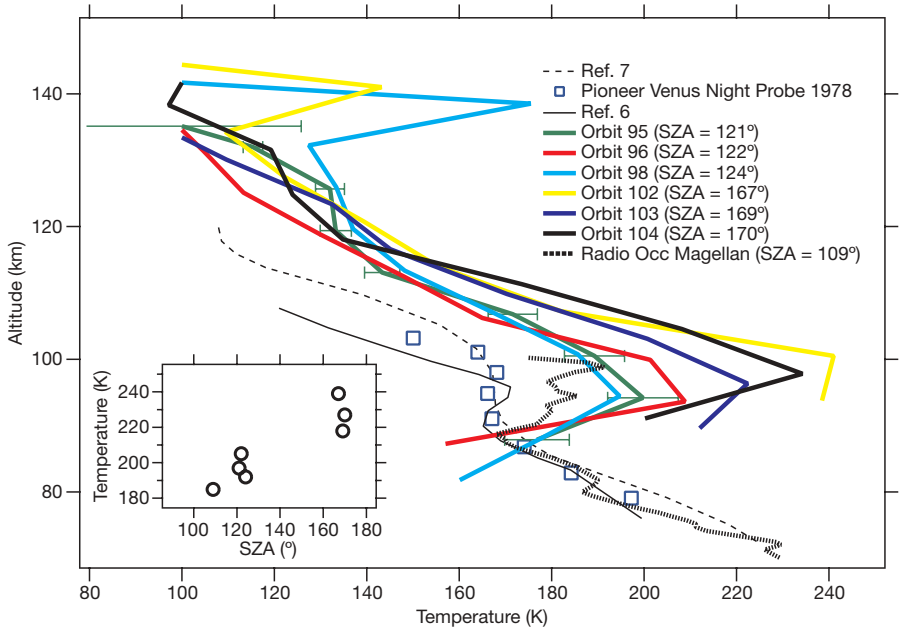


Figure 1 | Night-side temperature of Venus atmosphere. The mesopause may be defined at 90 km of altitude, separating the mesosphere (<90 km) from the lower thermosphere (>90 km). Solid thick curves, obtained by stellar occultations on the night side from Venus Express (this work), are compared to previous measurements. The blue squares were obtained during the descent of the Pioneer Venus night-side probe. The thin solid line shows sub-millimetre observations from a ground-based radio-telescope, Venus disk integrated[6]. The dashed line shows millimetre observations from a ground-based radio-telescope, Venus disk integrated[7]. The thick dotted line was obtained from a radio-occultation of the Magellan orbiter[9], at a SZA of 109°. This single measurement shows an increase of temperature at altitudes above 85 km, similar to the ones seen in this work, but less pronounced. The SPICAV error bars (1σ) are indicated for orbit 95 (green) and are typical of the errors for the other curves. The two groups of measurements were taken at latitude 39° N (local time ~21:00 h, SZA = 121–124°) for orbits 95, 96, 98 and latitude 4° S (local time ~23:20 h, SZA = 167–170°) for orbits 103, 103 and104. The orbit is 24 h long. There is a clear pattern for the value of the maximum temperature (90–100 km), increasing with the SZA, as emphasized in the bottom left plot (the point at SZA = 109° is from Magellan).

mixing of potential temperature and a reinforcement of the peak temperature.

This is the first time that a temperature inversion is so clearly identified at this altitude in the night-side atmosphere of Venus. However, there is a hint of temperature inversion in previous data at their maximum altitude soundings: a 2–4 K increase from 95 to 100 km in infrared spectra[8], and a 10 K increase from 90 to 100 km in the radio-occultation profile of Magellan[9] taken at a SZA of 109°. We note that, in our profiles, the higher (the nearer to the anti-solar point) is the SZA, the higher is the temperature maximum, this trend being confirmed on the low side of the SZA by Magellan (which has an excellent vertical resolution and sampling). Whether our new hot profiles are sporadic and patchy, or are permanent and large-scale features (having escaped detection before) will be clarified by future Venus Express measurements.

Hydrogen-bearing molecules (HCl, HF, H_2O and HDO) are broken by solar ultraviolet in the altitude range 80–120 km, and their abundance in the mesosphere is relevant to the ultimate escape rate of H, and possible evolution of the Venus atmosphere. HCl and HF were discovered in 1967 (ref. 3). More recently, the abundance of HDO (ref. 16) and H_2O (ref. 17) above cloud top were found to experience large time variations (by a factor of 30), even when the measurements were averaged over the whole disk of Venus, with no explanation as yet. Almost nothing is known about the vertical profile of these species, although it is essential for modelling of H escape. A new compact high-spectral-resolution infrared instrument, called Solar Occultation in the InfraRed (SOIR)[18] was implemented on Venus Express as an extension of SPICAV, to measure the vertical distribution of hydrogen-bearing species in the range 60–110 km by the technique of solar occultation at terminator (see the Supplementary Information). SOIR is a new type of spectrometer, with an echelle spectrometer associated to an acousto-optical tunable filter for wavelength domain selection, measuring the solar spectrum and the atmospheric transmittance (Fig. 2) in the infrared region

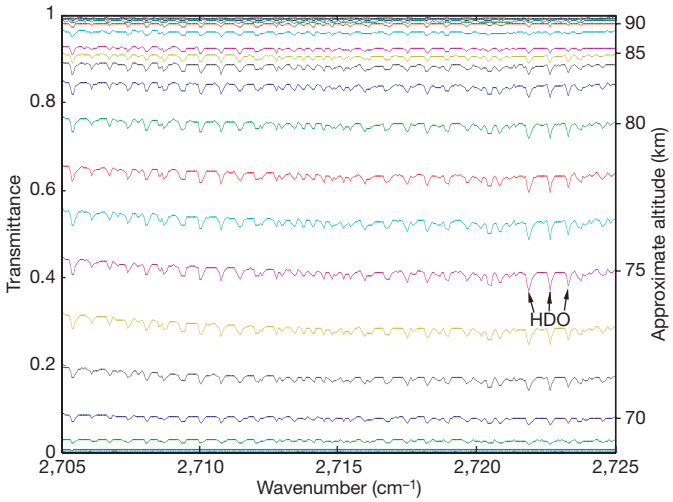


Figure 2 | Typical evolution of atmospheric spectral transmittances through one solar occultation observed by SOIR spectrometer. It is obtained by determining the ratio of the solar spectrum seen through the Venus atmosphere to the unattenuated solar spectrum measured above the atmosphere, at high spectral resolution and a 3 to 6 km altitude sampling. At the beginning of the series, the light path does not cross the atmosphere. No absorption signatures are present and transmittances are equal to unity. As the sun sets, the light path goes deeper and deeper into the atmosphere, and two absorption processes take place: the overall signal decreases owing to extinction by aerosols, and gaseous absorption signatures appear. At the end, the light path crosses the cloud layer located at an altitude around 60 km above the Venus surface (at 6,051.5 km radius) and light is no longer transmitted. This solar occultation was collected on 26 November 2006 during a sunset. The selection of a spectral interval is achieved through the acousto-optical tunable filter, tuned in this case to the 2,703.5–2,727.5 cm^{-1} spectral range. In this particular range, the main absorption lines are from HDO (a trio of lines indicated by arrows), and other features are from weak CO_2 spectral lines.

©2007 Nature Publishing Group

(2.2–4.3 μm) at a resolution of 0.15 cm^{-1}. At present, using it, absorption structures of CO_2, H_2O, HDO, CO, HCL and HF have been observed with certainty.

Retrieval of the vertical distribution of the various species is done by comparing the observed transmittance spectra to a full forward model simulation[19]. Some results of this best-fit exercise are shown in Fig. 3 for HF and HCl. The abundance of HF measured below the clouds[20] at 0.005 ± 0.002 p.p.m. agrees with other measurements[21] corresponding to altitudes above 72 km (0.0065 ± 0.0003 p.p.m.). Our SOIR-measured values (0.001 to 0.003 p.p.m. at 75–85 km) are significantly lower (by a factor of 3).

At altitudes between 15 and 30 km, lower than those probed by the SOIR instrument, an HCl mixing ratio of 0.5 ± 0.15 p.p.m. was obtained in 1989 (ref. 20) while a similar value of 0.6 ± 0.12 p.p.m. was found[3] in 1967 above the clouds (60 km). The abundances of HCl obtained from the SOIR measurements during the occultations of orbits 136 and 247, shown in Fig. 3a, are slightly different from each other (by a factor of ~1.5 at 70–75 km), and are lower than values reported in the literature (factor of 4 to 6): 0.1 ± 0.03 p.p.m. for orbit 136 and 0.17 ± 0.03 p.p.m. for orbit 247, both at 70–75 km. Current models of photochemistry of the atmosphere of Venus above the clouds[22] are assuming a value of 1 p.p.m.v. of HCl for this very active component and should be revised accordingly. Another possible explanation of our low values is that the solar occultation probed a high latitude, where thermospheric air depleted in HF and HCl (because of photo-dissociation at high altitude) may be spiralling down the polar vortex, along the descending branch of the Hadley's cell whose up-leg is at the subsolar point (Y. L. Yung & R. T. Clancy, personal communication).

Figure 4a shows retrievals of H_2O and HDO mixing ratios obtained simultaneously with the acousto-optical tunable filter sequence alternating between the H_2O spectral window and the HDO window, and with a window also dedicated to CO_2 retrieval. The three profiles were obtained over 18 days, at a high north latitude (80–75°) terminator. H_2O is at 0.3–1 p.p.m.v, an intermediate value within the very large range of 0.1 to 4 p.p.m.v reported from previous measurements by different techniques[16,17] at various times. At variance, our three vertical profiles show little time variability in this limited sample. There is a marked depletion of H_2O in the range 80–90 km, for which we have no explanation yet, other than noting that this altitude range coincides with the mesospheric minimum temperature and the top of the haze layer. For H_2O, there is no abrupt decreasing at high altitudes (up to 110 km), which would be a sign of local condensation.

The HDO/H_2O ratio profiles (Fig. 4b) are quite similar. At the lower boundary of our measurements (70 km), the HDO/H_2O ratio is ~0.1 with an error bar (1σ) exceeding 50%, which encompasses the value of 0.05 measured in the bulk atmosphere at lower altitudes[20,23–25] (to be compared to the Earth's value of 1/3,000). Although we could have expected a decrease of this ratio with increasing altitude because of preferential condensation of HDO[26], the trend is instead an increase in the HDO/H_2O ratio, up to a value of 0.12, about 2.5 times the bulk atmosphere value. Hence, there is no effective cold trap preventing the photo-dissociation of HDO in the upper atmosphere. The observed bulge of HDO at 90–95 km, above the haze and free from extinction of solar flux, might be due to a lower photo-dissociation rate of HDO (versus H_2O), as has been suggested for Mars from laboratory cross-section measurements[27].

Alternatively, it could be a sign that H atoms are escaping to outer space faster than D atoms (because of their larger thermal velocity): D atoms left behind in excess will recombine with OH radicals, generating a downward flow of HDO. If this interpretation is correct, it would be the first indication of this expected differential escape of H versus D, acting at present, that could explain the high D/H ratio in the present atmosphere of Venus. However, while the present 3 cm

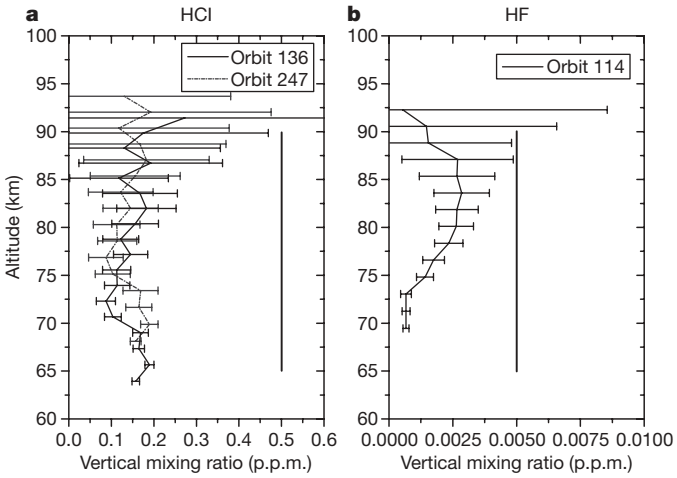


Figure 3 | HF and HCl mixing ratio vertical profiles retrieved from SOIR occultations. The mixing ratios are computed from the density retrievals of HF and HCl, divided by the CO_2 density retrieved from CO_2 absorption lines. The number of the orbit analysed is indicated on each graph. **a**, Two HCl profiles are compared, taken at orbit 136 (4 September 2006) and 247 (24 December 2006). They are somewhat different in the range 70–75 km of altitude, and the mixing ratio seems to increase slightly with altitude in both profiles. Error bars are 1σ. The amount of HCl is much less than is assumed in the photochemistry models (1 p.p.m.), and less than 0.6 ± 0.12 p.p.m. reported above the clouds[3], even after revision of this measurement to 0.42 ± 0.07 p.p.m. (ref. 28). The vertical bar represents a constant volume mixing ratio of 0.5 p.p.m., as an average of the two derivations[3,28] from 1967 observations. **b**, The HF (hydrofluoric acid) volume mixing ratio SOIR profile at orbit 114 (13 August 2006) is somewhat lower by a factor of ~3 than a previous measurement corresponding to altitudes above 72 km (vertical bar).

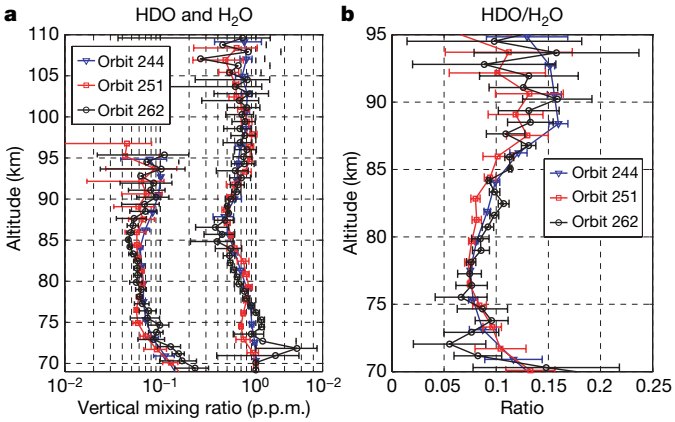


Figure 4 | HDO and H_2O mixing ratio, HDO/ H_2O vertical profiles. Both H_2O and HDO were measured simultaneously with SOIR during solar occultation at orbits 244, 251 and 262 between 70 and 110 km respectively at polar latitudes +85,83,73° (December 2006 to 8 January 2007). Error bars are 1σ. **a**, HDO slant densities were deduced using absorption structures located around 2,722.5 cm^{-1} (Fig. 2) and those of H_2O from structures in the 3,832.0–3,852.0 cm^{-1} spectral interval. After vertical inversion to obtain local densities, they were divided by the CO_2 density retrieved simultaneously from CO_2 absorption lines to get the volume mixing ratios. The less abundant curves are for HDO (the three curves on the left). The curves of both isotopes show little variability, but there is an unexplained and consistent H_2O depletion around 85 km. **b**, The HDO/H_2O ratio vertical profiles are compared for the three orbits. While this ratio is 1/3,000 in Earth's sea water, Venus is known to be enriched in HDO by a factor of 150, as a result of preferential escape of the lighter isotope H[20,23–25]. The value of this ratio in the bulk lower atmosphere is measured to be 0.05, while SOIR finds a higher ratio in the whole range 70–95 km. The HDO/H_2O ratio increases with altitude, reaching about 0.12 (a factor of 2.5 above the bulk ratio). For a tentative explanation of this deuterium-super-enriched layer, see text.

©2007 Nature Publishing Group

equivalent liquid of water and D/H ratio ≈ 0.025 (enrichment 150) would imply a global quantity of water of only 4.5 m depth in the past compared to Earth's 2.8-km-deep ocean (if there were no D escape), a significant escape of D atoms could largely increase this lower limit. The observed presence of HDO in the photo-dissociation region indicates that D atoms must be present in the thermosphere, where they might suffer non-thermal escape mechanisms, as well as H atoms. It is therefore important to quantify the non-thermal escape mechanism of H and D atoms.

Received 28 February; accepted 22 May 2007.

1. Dickinson, R. E. & Ridley, E. C. Venus mesosphere and thermosphere temperature structure: II. Day-night variations. *Icarus* **30,** 163–178 (1977).
2. Bougher, S. W., Alexander, M. J. & Mayr, H. G. in *Venus II* (eds Bougher, S. W., Hunten, D. M. & Phillips, R. J.) 259–291 (Univ. Arizona Press, Tucson, 1997).
3. Connes, P., Connes, J., Benedict, W. S. & Kaplan, L. D. Traces of HCl and HF in the atmosphere of Venus. *Astrophys. J.* **147,** 1230–1237 (1967).
4. Bertaux, J.-L. *et al.* SPICAM on Mars Express: Observing modes and overview of UV spectrometer data and scientific results, *J. Geophys. Res.* **111,** E10S90, doi:10.1029/2006JE002690 (2006).
5. Quémerais, E. *et al.* Stellar occultations observed by SPICAM on Mars Express. *J. Geophys. Rev.* **111,** doi:10.1029/2005JE002604 (2006).
6. Clancy, R. T., Sandor, B. J. & Moriarty-Schieven, G. H. Observational definition of the Venus mesopause: vertical structure, diurnal variation, and temporal instability. *Icarus* **161,** 1–16 (2003).
7. Lellouch, E., Goldstein, J. J., Rosenqvist, J., Bougher, S. W. & Paubert, G. Global circulation, thermal structure, and carbon monoxide distribution in Venus's mesosphere in 1991. *Icarus* **110,** 315–339 (1994).
8. Zasova, L. V., Moroz, V. I., Linkin, V. M., Khatuntsev, I. V. & Maioriv, B. S. Structure of the Venusian atmosphere from surface up to 100 km. *Cosmic Res.* **44,** 364–383 (2006).
9. Jenkins, J. M., Steffes, P. G., Hinson, D. P., Twicken, J. D. & Tyler, G. L. Radio occultation studies of the Venus atmosphere with the Magellan spacecraft, 2. Results from the October 1991 experiments. *Icarus* **110,** 79–94 (1994).
10. Stewart, A. I. F., Gérard, J. C., Rusch, D. W. & Bougher, S. W. Morphology of the Venus ultraviolet night airglow. *J. Geophys. Res.* **85,** 7861–7870 (1980).
11. Gerard, J. C., Stewart, A. I. F. & Bougher, S. W. The altitude distribution of the Venus ultra-violet nightglow and implications on Vertical transport. *Geophys. Res. Lett.* **8,** 633–636 (1981).
12. Connes, P., Noxon, J. F., Traub, W. A. & Carleton, N. P. O_2 ($^1\Delta$) emission in the day and night airglow of Venus. *Astrophys. J.* **233,** L29–L32 (1979).
13. Drossart, P. *et al.* Venus upper atmospheric emissions from VIRTIS spectral imaging observations. *Nature* doi:10.1038/nature05xxx (this issue)..
14. Bougher, S. W., Gerard, J. C., Stewart, A. I. F. & Fessen, C. G. The Venus nitric oxide night airglow—Model calculations based on the Venus Thermospheric General Circulation Model. *J. Geophys. Res.* **95,** 6271–6284 (1990).
15. Keating, G. M. *et al.* VIRA (Venus International Reference Atmosphere) models of Venus neutral upper atmosphere: structure and composition. *Adv. Space Res.* **5** (11), 117–171 (1985).
16. Sandor, B. J. & Clancy, R. T. Water variations in the Venus mesosphere from microwave spectra. *Icarus* **177,** 129–143 (2005).
17. Gurwell, M. A. Melnick, G. J., Tolls, V., Bergin, E. A. & Patten, B. M. SWAS observations of water vapor in the Venus mesosphere. *Icarus* **188,** 288–304 (2007).
18. Nevejans, D. *et al.* Compact high-resolution spaceborne echelle grating spectrometer with acousto-optical tunable filter based order sorting for the infrared domain from 2.2 to 4.3 μm. *Appl. Opt.* **45,** 5191–5206 (2006).
19. Vandaele, A. C., Kruglanski, M. & De Mazière, M. Simulation and retrieval of atmospheric spectra using ASIMUT. In *ESA Atmospheric Sci. Conf. (Frascati, Italy, 2006)* [CD-ROM] 71.1 (ESA SP-628, Noordwijk, 2006).
20. Bézard, B., De Bergh, C., Crisp, D. & Maillard, J.-P. The deep atmosphere of Venus revealed by high-resolution nightside spectra. *Nature* **345,** 508–511 (1990).
21. Björaker, G. L. *et al.* Airborne observations of the gas composition of Venus above the cloud tops: measurements of H_2O, HDO, HF and the D/H and $^{18}O/^{16}O$ isotopic ratios. *Bull. Am. Astron. Soc.* **24,** 995 (1992).
22. Yung, Y. L. & DeMore, W. B. Photochemistry of the stratosphere of Venus: Implications for atmospheric evolution. *Icarus* **51,** 199–247 (1982).
23. Donahue, T. M., Hoffman, J. H., Hodges, R. R. & Watson, A. J. Venus was wet—a measurement of the ratio of deuterium to hydrogen. *Science* **216,** 630–633 (1982).
24. Donahue, T. M., Grinspoon, D. H., Hartle, R. E. & Hodges, R. R. in *Venus II* (eds Bougher, S. W., Hunten, D. M. & Phillips, R. J.) 385–414 (Univ. Arizona Press, Tucson, 1997).
25. De Bergh, C. *et al.* Deuterium on Venus: observations from Earth. *Science* **251,** 547–549 (1991).
26. Bertaux, J. L. & Montmessin, F. Isotopic fractionation through water vapor condensation: the deuteropause, a cold trap for deuterium in the atmosphere of Mars. *J. Geophys. Res. Planets* **106,** 32879–32884 (2001).
27. Cheng, B.-M. *et al.* Photo-induced fractionation of water isotopomers in the martian atmosphere. *Geophys. Res. Lett.* **26,** 3657–3660 (1999).
28. Young, L. D. G. High resolution spectra of Venus-A review. *Icarus* **17,** 632–658 (1972).

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

Acknowledgements Venus Express is a space mission from the European Space Agency (ESA). We wish to thank all ESA members who participated in this successful mission, and in particular H. Svedhem, D. McCoy, O. Witasse, A. Accomazzo and J. Louet. We also thank Astrium for the design and construction of the spacecraft, and in particular A. Clochet, responsible for the payload. We thank D. Hinson for communication of the Magellan radio-occultation atmospheric profile, and Y. Yung and T. Clancy for discussions. We thank our collaborators at Service d'Aéronomie/France, BIRA/Belgium and IKI/Moscow for the design and fabrication of the instrument. We thank CNRS and CNES for financing SPICAV/SOIR in France, the Belgian government, Roskosmos and the Russian Academy of Sciences. The Russian team acknowledges support from the Russian Foundation for Basic Research, and from the Russian Science Support Foundation.

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.-L.B. (bertaux@aerov.jussieu.fr).

***The SPICAV/SOIR team** Jean-Loup Bertaux[1], D. Nevejans[2], Oleg Korablev[3], F. Montmessin[1], Ann-Carine Vandaele[2], A. Fedorova[3], M. Cabane[1], E. Chassefière[1], J. Y. Chaufray[1], E. Dimarellis[1], J. P. Dubois[1], A. Hauchecorne[1], F. Leblanc[1], F. Lefèvre[1], P. Rannou[1], E. Quémerais[1], E. Villard[1], D. Fussen[2], C. Muller[2], E. Neefs[2], E. Van Ransbeeck[2], V. Wilquet[2], A. Rodin[3], A. Stepanov[3], I. Vinogradov[3], L. Zasova[3], F. Forget[4], S. Lebonnois[4], D. Titov[5], S. Rafkin[6], G. Durry[7], J. C. Gérard[8] & B. Sandel[9]

[1]Service d'Aéronomie du CNRS/IPSL, Verrières-le-Buisson 91371, France. [2]Belgian Institute for Space Aeronomy, 3 avenue Circulaire, B-1180 Brussels, Belgium. [3]Space Research Institute (IKI), 84/32 Profsoyuznaya, 117810 Moscow, Russia. [4]Laboratoire de Météorologie Dynamique, 4 place Jussieu, 75252 Paris cedex 05, France. [5]Max-Planck-Institut für Sonnensystemforschung Max-Planck-Strasse 2, D-37191 Katlenburg-Lindau, Germany. [6]Southwest Research Institute, Department of Geophysics, Astrophysics and Planetary Science, 1050 Walnut Avenue, Suite 400, Boulder, Colorado 80302-5143, USA. [7]Groupe de Spectrométrie Moléculaire et Atmosphérique, Université de Reims, Champagne-Ardennes B.P. 1039, 51687 Reims cedex, France. [8]Université de Liège, Institut d'Astrophysique et Geophysique—B5c, Allée du 6 Aout, 17 Sart Tilman, B-4000 Liege, Belgium. [9]Lunar and Planetary Laboratory, University of Arizona, 1541 E. University Boulevard, Tucson, Arizona 85721, USA.

©2007 Nature Publishing Group

LETTERS

The loss of ions from Venus through the plasma wake

S. Barabash[1], A. Fedorov[2], J. J. Sauvaud[2], R. Lundin[1], C. T. Russell[3], Y. Futaana[1], T. L. Zhang[4], H. Andersson[1], K. Brinkfeldt[1], A. Grigoriev[1], M. Holmström[1], M. Yamauchi[1], K. Asamura[5], W. Baumjohann[4], H. Lammer[4], A. J. Coates[6], D. O. Kataria[6], D. R. Linder[6], C. C. Curtis[7], K. C. Hsieh[7], B. R. Sandel[7], M. Grande[8], H. Gunell[9], H. E. J. Koskinen[10,11], E. Kallio[11], P. Riihelä[11], T. Säles[11], W. Schmidt[11], J. Kozyra[12], N. Krupp[13], M. Fränz[13], J. Woch[13], J. Luhmann[14], S. McKenna-Lawlor[15], C. Mazelle[2], J.-J. Thocaven[2], S. Orsini[16], R. Cerulli-Irelli[16], M. Mura[16], M. Milillo[16], M. Maggi[16], E. Roelof[17], P. Brandt[17], K. Szego[18], J. D. Winningham[19], R. A. Frahm[19], J. Scherrer[19], J. R. Sharber[19], P. Wurz[20] & P. Bochsler[20]

Venus, unlike Earth, is an extremely dry planet although both began with similar masses, distances from the Sun, and presumably water inventories. The high deuterium-to-hydrogen ratio in the venusian atmosphere relative to Earth's also indicates that the atmosphere has undergone significantly different evolution over the age of the Solar System[1]. Present-day thermal escape is low for all atmospheric species. However, hydrogen can escape by means of collisions with hot atoms from ionospheric photochemistry[2], and although the bulk of O and O_2 are gravitationally bound, heavy ions have been observed to escape[3] through interaction with the solar wind. Nevertheless, their relative rates of escape, spatial distribution, and composition could not be determined from these previous measurements. Here we report Venus Express measurements showing that the dominant escaping ions are O^+, He^+ and H^+. The escaping ions leave Venus through the plasma sheet (a central portion of the plasma wake) and in a boundary layer of the induced magnetosphere. The escape rate ratios are $Q(H^+)/Q(O^+) = 1.9$; $Q(He^+)/Q(O^+) = 0.07$. The first of these implies that the escape of H^+ and O^+, together with the estimated escape of neutral hydrogen and oxygen, currently takes place near the stoichometric ratio corresponding to water.

Our knowledge of the interaction of Venus with the solar wind and its role in the acceleration of ions is based mainly on the long-lived Pioneer Venus orbiter mission, which carried a limited plasma analyser with low time resolution[4] and an ion-neutral mass spectrometer with a restricted energy range of up to ~100 eV (ref. 5). These observations identified O^+ ions only indirectly, through their energy spectra. However, they showed that the distribution of escaping oxygen ions is controlled by the solar-wind convection electric field, a distinctive signature of the ion pick-up process[6]. The comprehensive plasma package on Venus Express comprises an ion mass analyser, an electron spectrometer, and energetic neutral atom imagers[7]. The energy range of the charged particle sensors, from 1 eV (measurements below 10 eV may be affected by the spacecraft potential) to 15 keV (electrons) and to 36 keV (ions), is well suited for measurements of plasma escape. The magnetic field data are provided by the Venus Express magnetometer[8]. The polar, highly elliptical orbit, with a pericentre of 250–350 km and an apocentre of 66,000 km, is optimal for studying the escaping planetary ions because it covers the region of the interaction of the solar wind near the terminator and the region directly downstream of the planet.

During pericentre flybys the spacecraft crosses the bow shock, where the supersonic solar-wind plasma is heated and deflected, and enters the magnetosheath, a region of heated subsonic solar-wind flow (Fig. 1a). The magnetic field properties of these regions are treated in another paper in this issue[9]. Then the spacecraft penetrates the induced magnetosphere boundary, where the solar-wind flux drops and the planetary ions O^+, He^+ and H^+ appear (Fig. 1e–g). These three ion species are reliably identified on all orbits, predominantly in the plasma sheet (Fig. 1b), a region separating the two lobes of the induced magnetotail possessing opposite magnetic field directions. In the magnetic field data the plasma sheet is readily identified by the change in the sign of the magnetic field component along the Venus–Sun line (B_x) and a decreased field magnitude (Fig. 1c). A plasma sheet should contain plasma whose thermal pressure balances the magnetic pressure from the surrounding lobes. In particle data it is most readily identified by the increase in the electron fluxes (Fig. 1d). The plane of the plasma sheet is aligned with the electric field formed from solar-wind convection around the planet (Fig. 1b). The O^+, He^+ and H^+ ions of planetary origin are also found close to the induced magnetosphere boundary but at a much lower energy (around 7:20 and 8:10 UT in Fig. 1).

Whereas the O^+ ions can only be of planetary origin, He^+ ions may originally have been solar-wind α-particles having undergone recombination, and the H^+ ions could also be solar-wind protons. However, the identical shapes of their energy spectra and similar spatial distributions confirm the planetary origin of these ion species. Inside the plasma sheet the ions have different energies, in accordance with their masses (Fig. 2). In the $+Z_{Vse}$ hemisphere (Vse being the Venus–solar–electrical coordinate system), ~0.5–1.0-keV O^+ ions are separated from the 100–200-eV He^+ ions, whereas the H^+ ions with energies below 100 eV fill both sides ($+Y_{Vse}$ and $-Y_{Vse}$) of the plasma sheet (Fig. 2b, c). At the induced magnetosphere boundary, all three components have an energy below 50 eV (Fig. 2d).

[1]Swedish Institute of Space Physics, S-98128 Kiruna, Sweden. [2]Centre d'Étude Spatiale des Rayonnements, BP-44346, F-31028 Tolouse, France. [3]IGPP, University of California, Los Angeles, California 90095, USA. [4]Space Research Institute, Austrian Academy of Science, A-8042 Graz, Austria. [5]Institute of Space and Astronautical Science, 3-1-1 Yoshinodai, Sagamihara 229-8510, Japan. [6]Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK. [7]University of Arizona, Tucson, Arizona 85721, USA. [8]University of Wales Aberystwyth, Penglais, Aberystwyth, Ceredigion SY23 3BZ, UK. [9]Department of Physics, West Virginia University, Morgantown, West Virginia 26506-6315, USA. [10]Department of Physical Sciences, University of Helsinki, Box 64, 00014 Helsinki, Finland. [11]Finnish Meteorological Institute, Box 503, FIN-00101 Helsinki, Finland. [12]Space Physics Research Laboratory, University of Michigan, Ann Arbor, Michigan 48109-2143, USA. [13]Max-Planck-Institut für Sonnensystemforschung, Max-Planck-Str. 2, D-37191 Katlenburg-Lindau, Germany. [14]Space Science Laboratory, University of California, Berkeley, California 94720-7450, USA. [15]Space Technology Ireland, National University of Ireland, Maynooth, Co. Kildare, Ireland. [16]Instituto di Fisica dello Spazio Interplanetari, I-00133 Rome, Italy. [17]Applied Physics Laboratory, Johns Hopkins University, Laurel, Maryland 20723-6099, USA. [18]KFKI Research Institute for Particle and Nuclear Physics, PO Box 49, H-1525 Budapest 114, Hungary. [19]Southwest Research Institute, San Antonio, Texas 78228-0510, USA. [20]University of Bern, Physikalisches Institut, CH-3012 Bern, Switzerland.

©2007 Nature Publishing Group

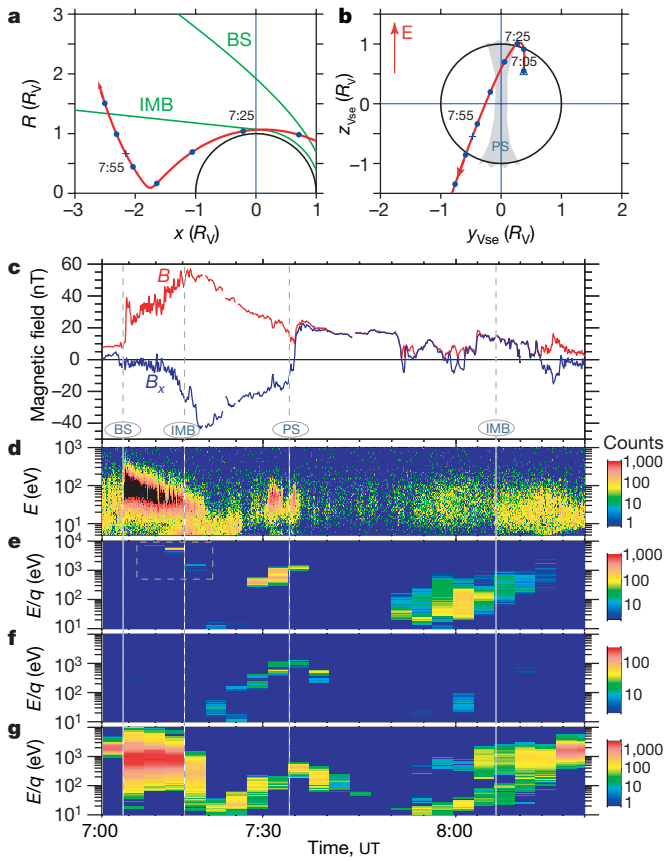


Figure 1 | Venus Express orbit, and plasma and magnetic field measurements on 2006 December 12. **a**, The Venus Express orbit in cylindrical coordinates, with x pointing towards the Sun, and R being the distance to the x axis (in Venus radii, $R_V = 6{,}051$ km). The green lines are the averaged[22] positions of the bow shock (BS) and induced magnetosphere boundary (IMB)[23], and the red line is the Venus Express orbit. **b**, The Venus Express orbit in the Venus–solar–electrical coordinate system (Vse). In this system x points to the Sun, and z is in the plane containing the convection electric field $E = -V_{sw} \times B$ (where V_{sw} is the velocity of the solar wind, B is the interplanetary magnetic field and $\times$ denotes the vector cross product); y completes the orthogonal coordinate system. The plasma sheet (PS) is identified by the grey area. The direction of the convection electric field is labelled E. **c**, The magnetic field magnitude (red line) and B_x component (blue line). **d–g**, Energy–time spectrograms for electrons (**d**), O^+ ions (**e**), He^+ ions (**f**) and protons (**g**) (see legend to Supplementary Fig. 1). Venus Express crosses the IMB at 7:02 UT, passes the PS at 7:34 UT and enters the magnetosheath again at 8:08 UT. The pick-up O^+ ions in the magnetosheath at about 7:13 UT are bounded by a dashed rectangle. E/q, energy/charge.

At Venus the ion acceleration occurs by three mechanisms: by ion pick-up (that is, by acceleration in the convection electric field), by instabilities at the induced magnetosphere boundary resulting in detached plasma clouds[10], and by polarization electric fields at low altitudes in the night-side ionosphere[11] where the field is nearly radial.

We observed two different populations of the accelerated planetary ions, namely ions escaping through the plasma sheet and ions escaping through the induced magnetosphere boundary layer. The energy distributions of these two populations are different. The ratio of the $O^+/He^+/H^+$ energies in the plasma sheet is about 4:2:1 (Fig. 2b, c). The ion energy is therefore a function of mass. That points to ion pick-up as an acceleration mechanism in the plasma sheet, in general agreement with the previous simulations for O^+ (ref. 12). However, ordinary pick-up would result in a 16:4:1 ratio of the $O^+/He^+/H^+$

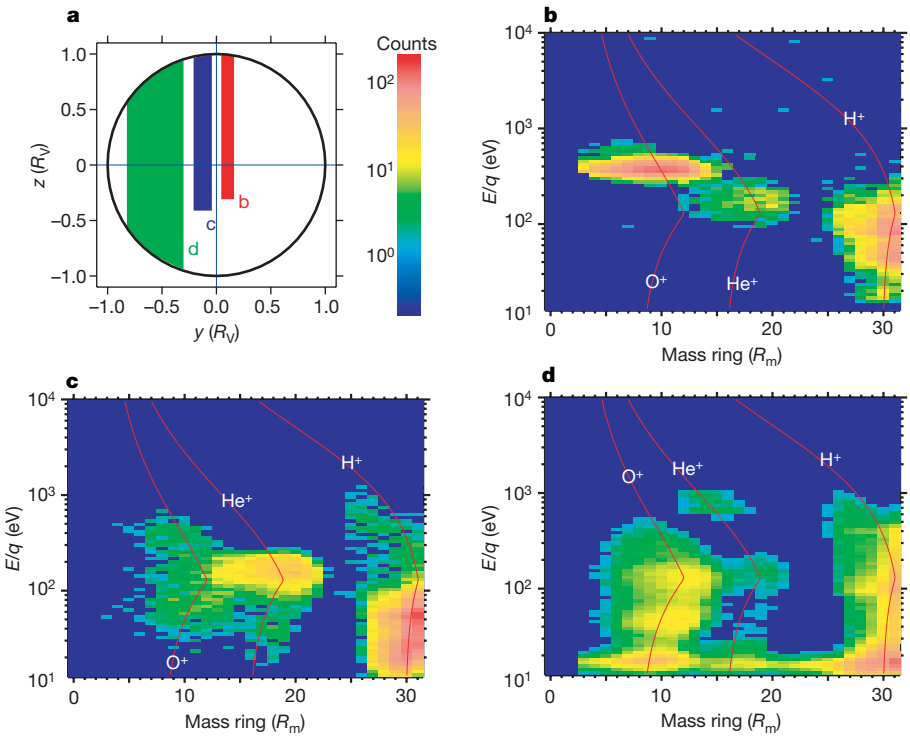


Figure 2 | Composition and energy distribution of the escaping plasma in three spatial regions. **a**, The regions labelled by c and d where energy-mass matrixes were collected in the Vse coordinate system (see legend to Supplementary Fig. 2). **b–d**, The energy–mass matrixes accumulated over 33 orbits during the period from 2006 May 18 to 2006 December 30 and collected in regions b-d in panel **a**, showing a plasma sheet matrix for the $+Y$ lobe (**b**), a plasma sheet matrix for the $-Y$ lobe (**c**), and a matrix for the IMB layer (**d**). Vertical axes are energy/charge (E/q), the horizontal axes are the position of the ion impact on the detector (a sensor mass ring, R_m), and the colour codes the counts that were accumulated over all directions and averaged over occurrences. The red lines represent constant mass.

©2007 Nature Publishing Group

energies. The difference may be due to species-dependent assimilation into the plasma flow. The ion pick-up process also operates, for example, in the magnetosheath region, in general agreement with the previous simulations for O^+ (ref. 12). See a clear example of 7–8-keV O^+ ions at 7:15 UT in Fig. 1e and 700–800-eV He^+ ions in Fig. 2d. The acceleration of the low-energy ions (<50 eV) detected at the induced magnetosphere boundary (Fig. 2d) is possibly due to a polarization electric field[11] resulting from the magnetic pressure gradient and the gravitational separation of ions and electrons along draped, stretched magnetic field lines. Because of the low plasma velocity in the boundary layer close to the induced magnetosphere boundary, the convection electric field is low and the accelerated ionospheric ions move away from the planet along magnetic field lines in a manner similar to that of the polar wind on Earth.

It is an open question whether the low-energy (<50-eV) ions detected at the induced magnetosphere boundary near the terminator might instead be related to the plasma clouds observed at the ionopause by the Pioneer Venus orbiter[10]. All of the atmospheric ions moving downstream are accelerated and their original morphological features disappear. However, for escape rate calculations their exact origins are not important as soon as the main escape region has been covered. The orbital scan through the induced magnetosphere does indeed show that it is inside the wake that we observe the highest fluxes of ions (Fig. 1e–g).

The ion escape considered here occurs through two main regions: the plasma sheet and a boundary layer at the induced magnetosphere boundary. These areas are labelled PS and BL in Fig. 3. As expected, the alignment of the plasma sheet plane and the ion energy are controlled by the electric field. Ions of higher energies are in the $+E$ hemisphere and ions of lower energies in the $-E$ hemisphere. All three species show similar spatial distributions, confirming that both H^+ and He^+ are indeed of planetary rather than solar-wind origin.

The escape rates for each species, namely $Q(O^+)$, $Q(He^+)$ and $Q(H^+)$, are obtained by integrating the flux over Y_{Vse} and Z_{Vse}. Although the spatial coverage is still not sufficient to obtain the total average escape rate, one can obtain reliable flux ratios. These are $Q(H^+)/Q(O^+) = 1.9$ and $Q(He^+)/Q(O^+) = 0.07$.

The escaping plasma consists primarily of H^+ and O^+ ions with some admixture of He^+. The measured composition differs significantly from the plasma composition in the ionosphere at an ionopause altitude of 300 km for solar minimum conditions[13], where the respective number-density ratios are $n(H^+)/n(O^+) = 0.1$ and $n(He^+)/n(O^+) = 4 \times 10^{-3}$. The enrichment of the escaping plasma in light species may result from two causes: first, an ion pick-up process operating primarily at higher altitudes, 600–700 km, in which the densities of light species are increased with respect to heavier ions[14]; and second, acceleration by the polarization electric field increasing the flux of lighter ions with respect to heavier because lighter ions gain higher velocities for the same energy.

The ratio of the escape rates of hydrogen and oxygen is a critical parameter in understanding both the dryness and the oxidation state of the venusian atmosphere[15]. Because the main source of hydrogen in the upper atmosphere of Venus is water vapour[16], two escaping hydrogen atoms would leave one oxygen atom in the atmosphere. Therefore, if the ratio of the escape rate of hydrogen to that of oxygen is greater than the stoichiometric ratio 2:1, there must be other sinks of oxygen, such as surface oxidation.

The measured ratios concern only ions and do not include the escape of the respective atoms. There are two processes that could remove neutral hydrogen and oxygen from the atmosphere to space: loss due to exogenic photochemical and charge-exchange reactions, and atmospheric sputtering. Photochemical and charge-exchange reactions result in a loss of hydrogen at a rate of 50% of the H^+ loss[14]. Atmospheric sputtering results in an oxygen loss at a rate of 25% of the O^+ loss[14]. Scaling the measured ratio $Q(H^+)/Q(O^+) = 1.9$ to account for the neutral losses, we arrive at a ratio for the total (neutral atoms plus ions) loss of 2.2:1, very close to the stoichiometric value 2:1. Some excess H^+ is due to solar-wind protons, which may penetrate through the induced magnetosphere boundary in small amounts and contribute to the measured H^+ flux in the boundary layer. That the present escape of hydrogen and oxygen through the wake takes place in the stoichiometric ratio of water implies that the atmosphere of Venus did not change its oxidation state after steady-state conditions had been reached, in contrast with that of Mars[17]. This result is thus consistent with the suggestion from atmospheric and ionospheric observations on the Pioneer Venus orbiter[15] that Venus loses the constituents of water in a ratio that maintains the system's oxidation state.

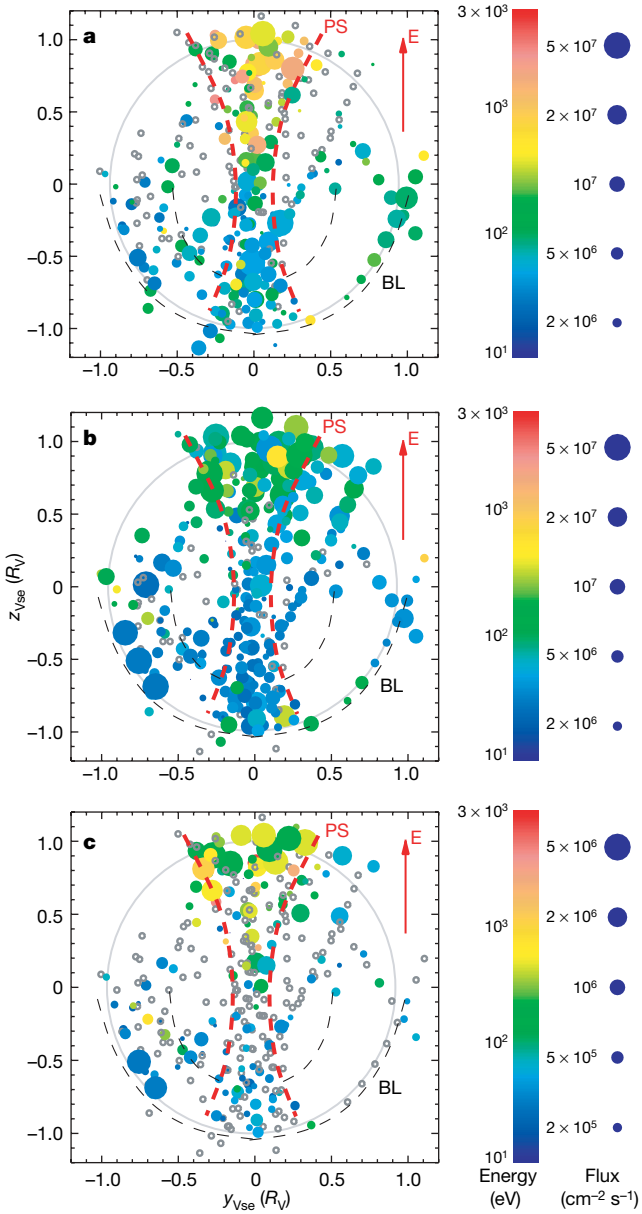


Figure 3 | Spatial distribution of the escaping plasma. The measured O^+ (**a**), H^+ (**b**) and He^+ (**c**) flux distributions in the tail region from 33 orbits were integrated over X_{Vse} [−0.5, −3.0] and are shown in a Y_{Vse}–Z_{Vse} plane across the tail. The geometrical eclipse of Venus is shown by the thin grey circle. To ensure that no solar-wind protons affect the mass composition measurements inside the IMB, we restrict the area of the analysis to $R < 1.2R_V$. Blank circles show measurements with zero flux. The flux scale for He^+ is tenfold lower than those for O^+ and H^+. The plasma sheet region is identified by red dashed lines and labelled PS, the boundary layer at the IMB is identified by black dashed lines and labelled BL, and the direction of the convection electric field is labelled E.

©2007 Nature Publishing Group

The absolute escape rates, including the implications for long-term water evolution, will be determined later. However, the initial analysis gives the reliable lower limit of $10^{25}\,s^{-1}$ for the escape of H^+ through the plasma wake. Measurements with the Pioneer Venus orbiter's ion and neutral mass spectrometer[18] and models of the observed tail magnetic and plasma structure[19] have inferred that there may be copious escape in the wake region. Evidence for hydrogen loss through the pick-up process outside the induced magnetosphere is also seen on Venus Express (M. Delva, personal communication) but the estimated rate of 2×10^{24} protons s^{-1} is much lower than that through the wake. Venus therefore loses its water through the plasma wake.

The observed high relative abundance of He^+ in the escaping plasma from Venus is surprising. At Mars He^+ is not detected in amounts comparable to the O^+ flux[20]. The calculated helium ionization rates at Mars[21] through two channels, photoionization and electron impact ionization, scaled to the total area of Venus, give only an increase in the escaping rate at Venus of a factor of 3 with respect to Mars. The higher abundance of He^+ may be due to the more effective acceleration caused by the polarization field[11]. The higher pressure gradients associated with the stronger gravitational field on Venus results in a higher polarization field that operates over larger distances at Venus than at Mars. Thus, whereas the stronger gravitational field of Venus reduces the Jeans escape of the heavier main constituents of the atmosphere, almost paradoxically it facilitates the loss of the lighter species.

The main result of these observations, however, is the establishment of potentially important ion escape from present-day Venus through the plasma wake region. Further observations in this region as the solar cycle progresses will establish the variability of these escape rates and will improve our ability to infer the history of Venus's water.

Received 26 June; accepted 29 October 2007.

1. Donahue, T. M., Grinspoon, D. H., Hartle, R. R. & Hodges, R. R. Jr. in *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment* (eds Bougher, W. W., Hunten, D. M. & Phillips, R. J.) 385–414 (Univ. of Arizona Press, Tucson, AZ, 1997).
2. Nagy, A. F., Cravens, T. E., Lee, J. H. & Stewart, A. I. P. Hot oxygen atoms in the upper atmosphere of Venus. *Geophys. Res. Lett.* **8,** 629–632 (1981).
3. Mihalov, J. D. & Brace, A. The distant interplanetary wake of Venus: Plasma observations from Pioneer Venus. *J. Geophys. Res.* **87,** 9045–9053 (1982).
4. Intrilligator, D. S., Wolf, J. H. & Michalov, J. D. The Pioneer Venus orbiter plasma analyzer experiment. *IEEE Trans. Geosci. Remote Sens.* **GE-18,** 39–43 (1980).
5. Kasprzak, W. T., Niemann, H. B., Hedin, A. E., Bougher, S. W. & Hunten, D. M. Neutral composition measurements by the Pioneer Venus neutral mass spectrometer during re-entry. *Geophys. Res. Lett.* **20,** 2747–2750 (1993).
6. Intriligator, D. S. Results of the first statistical studies of Pioneer Venus plasma observations in the distant Venus tail: Evidence for a hemispheric asymmetry in the pickup of ionospheric ions. *Geophys. Res. Lett.* **16,** 167–170 (1989).
7. Barabash, S. *et al.* The Analyser of Space Plasmas and Energetic Atoms (ASPERA-4) for the Venus Express mission. *Planet. Space Sci.* **55,** 1772–1792 (2007).
8. Zhang, T. L. *et al.* Magnetic field investigation of the Venus plasma environment: expected new results. *Planet. Space Sci.* **54,** 1336–1343 (2006).
9. Zhang, T. L. *et al.* Little or no solar wind enters Venus' atmosphere at solar minimum. *Nature* doi:10.1038/nature06026 (this issue).
10. Brace, L. H., Theis, R. F. & Hoegy, W. R. Plasma clouds above the ionopause of Venus and their implications. *Planet. Space Sci.* **30,** 29–37 (1982).
11. Hartle, R. E. & Grebowsky, J. M. Planetary loss from light ion escape on Venus. *Adv. Space Res.* **15,** 117–122 (1995).
12. Luhmann, J. G., Ledvina, S. A., Lyon, J. G. & Russell, C. T. Venus O^+ pickup ions: Collected PVO results and expectations for Venus Express. *Planet. Space Sci.* **54,** 1457–1471 (2006).
13. Fox, J. L. Advances in the aeronomy of Venus and Mars. *Adv. Space Res.* **33,** 132–139 (2004).
14. Lammer, H. *et al.* Loss of hydrogen and oxygen from the upper atmosphere of Venus. *Planet. Space Sci.* **54,** 1445–1456 (2006).
15. McElroy, M. B., Prather, M. J. & Rodriguez, J. M. Loss of oxygen from Venus. *Geophys. Res. Lett.* **9,** 649–651 (1982).
16. von Zahn, U., Kumar, S., Niemann, H. & Prinn, R. in *Venus* (eds Hunten, D. M., Colin, L., Donahue, T. M. & Moroz, V. I.) 299–430 (Univ. of Arizona Press, Tucson, AZ, 1983).
17. Lammer, H. *et al.* Loss of water from Mars: Implications for the oxidation of the soil. *Icarus* **165,** 9–25 (2003).
18. Luhmann, J. G., Kasprzak, W. T. & Russell, C. T. Space weather at Venus and its potential consequences for atmosphere evolution. *J. Geophys Res.* **112,** E04s10, doi:10.1029/2006JE002820 (2007).
19. McComas, D. J., Spence, H. E., Russell, C. T. & Saunders, M. A. The average magnetic field draping and consistent plasma properties of the Venus magnetotail. *J. Geophys. Res.* **91,** 7939–7953 (1986).
20. Barabash, S., Fedorov, A., Lundin, R. & Sauvaud, J.-A. Martian atmospheric erosion rates. *Science* **315,** 501–503 (2007).
21. Krasnopolsky, V. A. & Gladstone, G. R. Helium on Mars and Venus: EUVE observations and modeling. *Icarus* **176,** 395–407 (2005).
22. Zhang, T. L., Luhmann, J. G. & Russell, C. T. The solar cycle dependence of the location and shape of the Venus bow shock. *J. Geophys. Res.* **95,** 14961–14967 (1990).
23. Zhang, T. L., Luhmann, J. G. & Russell, C. T. The magnetic barrier at Venus. *J. Geophys. Res.* **96,** 11145–11153 (1991).

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

Acknowledgements We thank the European Space Agency for providing the Venus Express opportunity, and national space agencies and organizations for supporting the investigators who contributed to the success of the Venus Express plasma package.

Author Contributions S.B. is the principal investigator of the Venus Express plasma package, Analyser of Space Plasmas and Energetic Atoms (ASPERA)-4. J.A.S. is the co-principal investigator. A.F. is the leading co-investigator of the ion mass analyser of ASPERA-4. A.C. is the leading co-investigator of the electron spectrometer of ASPERA-4. T.L.Z. is the principal investigator of the magnetometer. The remaining authors are co-investigators on either of the plasma analyser or the magnetometer investigations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.B. (stas@irf.se).

©2007 Nature Publishing Group

LETTERS

Little or no solar wind enters Venus' atmosphere at solar minimum

T. L. Zhang[1,9], M. Delva[1], W. Baumjohann[1], H.-U. Auster[2], C. Carr[4], C. T. Russell[5], S. Barabash[6], M. Balikhin[7], K. Kudela[8], G. Berghofer[1], H. K. Biernat[1], H. Lammer[1], H. Lichtenegger[1], W. Magnes[1], R. Nakamura[1], K. Schwingenschuh[1], M. Volwerk[1], Z. Vörös[1], W. Zambelli[1], K.-H. Fornacon[2], K.-H. Glassmeier[2], I. Richter[2], A. Balogh[4], H. Schwarzl[5], S. A. Pope[7], J. K. Shi[9], C. Wang[9], U. Motschmann[3] & J.-P. Lebreton[10]

Venus has no significant internal magnetic field[1], which allows the solar wind to interact directly with its atmosphere[2,3]. A field is induced in this interaction, which partially shields the atmosphere, but we have no knowledge of how effective that shield is at solar minimum. (Our current knowledge of the solar wind interaction with Venus is derived from measurements at solar maximum[4–7].) The bow shock is close to the planet, meaning that it is possible that some solar wind could be absorbed by the atmosphere and contribute to the evolution of the atmosphere[8,9]. Here we report magnetic field measurements from the Venus Express spacecraft[4] in the plasma environment surrounding Venus. The bow shock under low solar activity conditions seems to be in the position that would be expected from a complete deflection by a magnetized ionosphere[10]. Therefore little solar wind enters the Venus ionosphere even at solar minimum.

Solar activity controls almost every aspect of the Venus plasma environment. Venus has no intrinsic magnetic field, so the solar wind is expected to interact directly with the upper atmosphere, which is partially ionized by solar extreme ultraviolet radiation and energetic particles that enter from surrounding space. These change with the phase of the solar cycle, as does the solar wind. Although the earlier Pioneer Venus Orbiter mission operated over a whole solar cycle, the Pioneer Venus Orbiter periapsis was too high (more than 2,000 km at solar minimum) to sample the near-Venus plasma environment. Since the Venus Express insertion at solar minimum, its periapsis altitude has been maintained at 250–350 km, allowing critical *in situ* measurements of the solar wind interaction with the ionosphere. Further, the near-polar orbit of Venus Express, with its high latitude of periapsis, fills critical gaps left by the Pioneer Venus Orbiter orbital sampling bias: the low-altitude region near the terminator (separating day and night sides) and mid-magnetotail downstream by 2–4 R_V (1 R_V = 6,051 km, one Venus radius). Venus Express also allows better exploration of the near-subsolar bow shock than did the Pioneer Venus Orbiter, whose trajectory did not penetrate the subsolar shock region. The Venus Express also has improved instrument temporal resolution. This increase in resolution enables us to study the plasma processes using the dynamic power spectra derived from magnetic field data sampled at 32 Hz.

Several notable features of the interaction can clearly be seen in such spectra (Fig. 1). Initially the satellite crosses the bow shock on the day side where the magnetic field strength and wave power increase (point A). It then moves through the shocked plasma of the magnetosheath to the night side where the spacecraft passes through closest approach (point C). Upstream waves are obvious at frequencies from 0.1 to 2 Hz in the solar wind in front of the quasi-parallel shock (Fig. 1a, far left) but are absent in front of the quasi-perpendicular shock (Fig. 1b) where the magnetic field lies nearly in the plane of the shock. Such waves can be either locally generated by backstreaming particles or generated at the bow shock and move upstream. The upstream wave and particle phenomena at Venus are generally similar to those at Earth. An example of the shock-generated waves seen at both planets are the upstream whistlers near 2 Hz, upstream of the quasi-parallel shock.

The size of the bow shock is largely determined by how completely the planetary obstacle deflects the solar wind. During April to August 2006, we obtained 147 clear crossings (Fig. 2). The best fit to the bow-shock location from a solar-zenith angle of 20° to 120° gives a terminator bow-shock location of 2.14 R_V, which is 1,600 km closer to Venus than the 2.40 R_V at solar maximum[11,12] but somewhat further than reported by Venera 9 and 10 (ref. 13). The best fit to the subsolar bow shock is 1.32 R_V, only 1,900 km above the surface of the planet, so we need a quantitative calculation to determine whether the solar wind is being partially absorbed by (added to) the planetary atmosphere, or is being fully deflected. To do this, we compare the altitude of the induced magnetic barrier with the altitude that would produce the observed location of the shock and deflect all the solar wind.

The induced magnetosphere of Venus consists of regions near the planet and its wake in which magnetic pressure dominates the other pressure contributions[6]. On the day side, the magnetic field piles up to form a magnetic barrier in the inner magnetosheath[14]. This magnetic barrier acts as an obstacle to the solar wind, in analogy to the Earth's magnetosphere. The magnetic barrier is bounded by the ionopause at its lower boundary and a 'magnetopause' at its upper boundary. Both ionopause and magnetopause extend to the night side. The magnetopause on the night side separates the magnetosheath from the magnetotail, which is formed by the anchored, draped magnetic fields. The upper boundary of the magnetic barrier, the induced magnetopause, had not been previously well-determined because of the insufficient temporal resolution of the Pioneer Venus Orbiter plasma instrument[14]. Phobos and the Mars Global Survey found that this boundary at Mars is associated with a drop in magnetic wave activity[15,16]. The dynamic spectra of Fig. 1 show that the Venus magnetosheath behaves similarly. Downstream of both the quasi-parallel and quasi-perpendicular shock, these waves are found with greater bandwidth behind the quasi-parallel shock. The wave

[1]Space Research Institute, Austrian Academy of Sciences, A-8042 Graz, Austria. [2]Institut für Geophysik und Extraterrestrische Physik, [3]Institut für Theoretische Physik, TU Braunschweig, D-3300 Germany. [4]Imperial College, London SW7 2BZ, UK. [5]IGPP, University of California, Los Angeles, California 90095, USA. [6]Swedish Institute of Space Physics, Kiruna, S-98128 Sweden. [7]University of Sheffield, Sheffield S1 3JD, UK. [8]Institute of Experimental Physics, Slovakia Academy of Sciences, Kosice, 04353 Slovakia. [9]State Key Laboratory of Space Weather, Chinese Academy of Sciences, 100080 China. [10]RSSD-ESTEC, Noordwijk 2000 AG, The Netherlands.

©2007 Nature Publishing Group

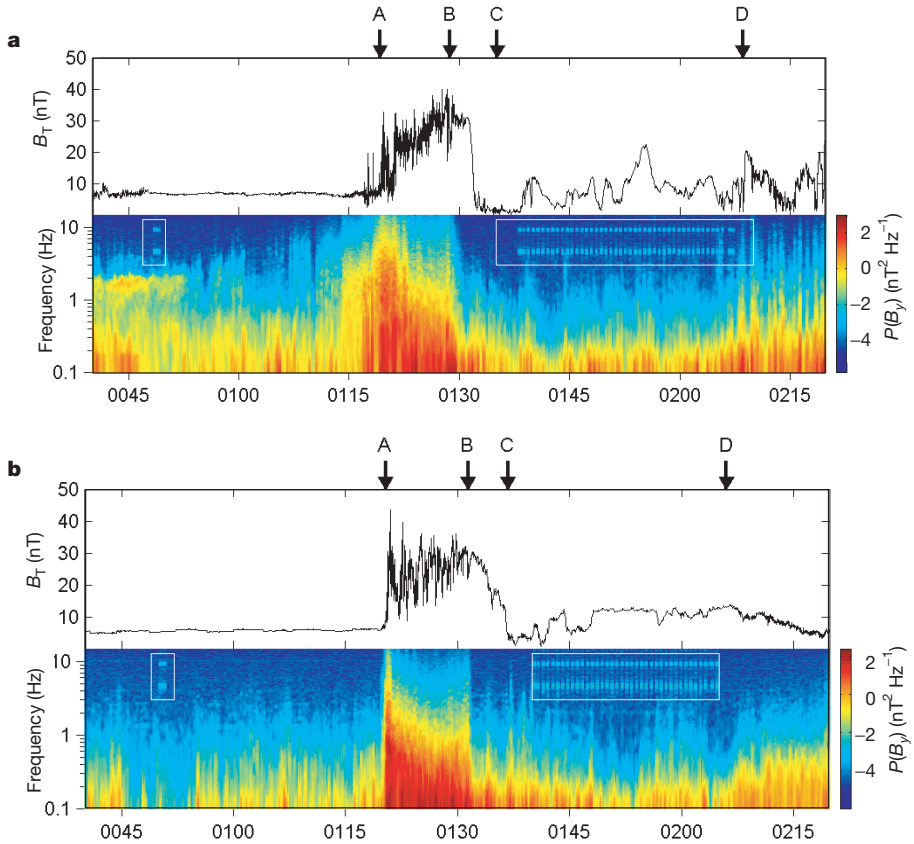


Figure 1 | Magnetic field measurements during pericentre fly-by. Data were obtained on 17 May 2006 (**a**) and 21 May 2006 (**b**). The top panels show the field strength B_T and the bottom panels show the dynamic power P spectrum. The x axis shows Universal Time in units of hours and minutes. The data used for the field time series are one-second-averaged data and the dynamic power spectra are calculated from the 32 Hz measurements. The 17 May shock is a quasi-parallel shock with a shock normal angle of 27°; the 21 May shock is quasi-perpendicular with a shock normal angle of 71°. The regular disturbances bounded by the white squares are the artificial effect of spacecraft reaction wheels. A, bow shock; B, magnetopause; C, closest approach; D, eclipse shadow boundary. B_T is the total magnetic field strength and B_y is the component of the magnetic field along the y-sensor, both in nanoTeslas.

activity drops sharply as the planet is approached, dividing the magnetosheath into two distinct regimes.

The observed magnetic field in the barrier wraps around the planet and applies a pressure gradient force back into the incoming solar wind, acting to deflect the solar wind. We have used the disappearance of waves from April to August 2006 to define the outer edge of the magnetic barrier, finding 137 clear crossings of this induced magnetopause. The orbital geometry of Venus Express means that most of the crossings are in the polar region with solar-zenith angle from 50° to 110°. The best-fit induced magnetopause (Fig. 2) gives an altitude of 1,013 km at the terminator and 300 km at the subsolar point. Thus the magnetic barrier is significantly lower than at solar maximum[14]. In fact, its upper boundary is the same altitude as the lower boundary of the magnetic barrier at solar maximum. We note that an earlier such study using a smaller number of Venera 9 and 10 shock locations at solar minimum concluded that there was a 10% absorption[17].

Near the planet the signature is quite different from that seen from Pioneer Venus[2,3]. Closer to periapsis, the field remains quiet, indicating that the spacecraft stays in the magnetic barrier. We do not see a drop in field strength indicative of a rise in plasma pressure located at the altitude at which the solar wind dynamic pressure is approximately balanced by the thermal pressure of the ionospheric plasma. If there is a well-defined ionopause in the subsolar region it must be below the current periapsis altitude. In fact, Pioneer Venus Orbiter radio occultations suggested that the altitude of the ionopause could be much depressed during the solar minimum, with an altitude of ~250 km on the day side[18]. We find that the ionosphere is completely magnetized and a well-defined magnetotail is formed in the Venus night side down to the lowest altitudes probed, 250 km (Fig. 2).

We can now evaluate the efficiency of the magnetized ionosphere in deflecting the solar wind with a quantitative comparison of the location of the observed magnetic barrier for models that fit the observed bow-shock locations. We use the gas dynamic model as described previously[10,19] with a ratio of specific heats of 5/3 and a

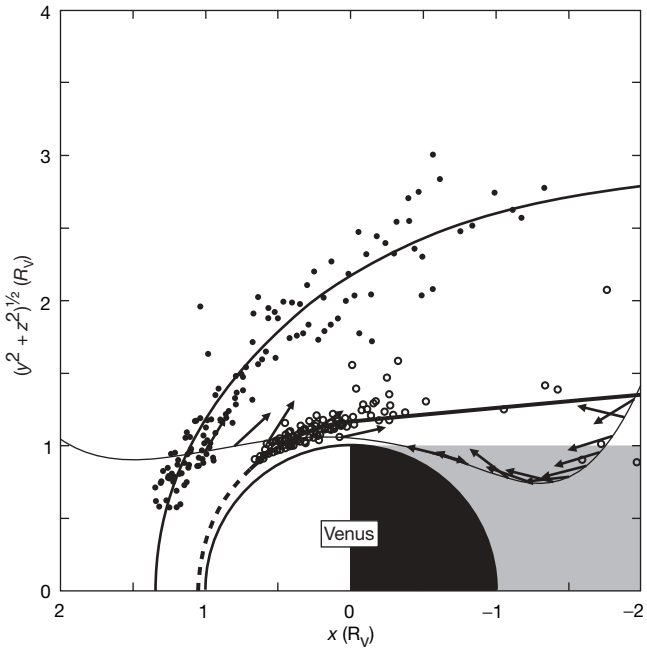


Figure 2 | Bow-shock location and induced magnetopause boundary at Venus. Solid circles are the bow-shock crossings and solid line is the best fit. The open circles are crossings of the magnetopause of the induced magnetosphere and the dashed-then-solid line is the best-fit model to these points. The shaded region is the optical shadow. The thin line is the trajectory along which arrows whose direction and length are proportional to the magnetic field projected in this plane. x is the direction to the Sun; y is the direction opposite to planetary motion; and z is northward, perpendicular to the orbit plane of Venus.

©2007 Nature Publishing Group

Mach number of 5.5 (the average value for solar minimum) and find that the solar wind is completely deflected around Venus: the obstacle in the simulation coincides with the top of the observed magnetic barrier (288 km), allowing little solar wind addition to the atmosphere even at solar minimum. The observed magnetic field is wrapped around the day side of the planet and it continues to hug the shape of the planet on the night side, reversing its direction to form a nearly complete torus. The anti-sunward force associated with curved magnetic fields in the night ionosphere and near-tail could lead to acceleration of plasma, resulting in further loss of the atmosphere. We await measurements of the plasma velocities in this region to confirm this inference and to determine whether these loss rates would have a major atmospheric impact when integrated over the age of the solar system.

METHODS

The Venus Express magnetometer measures the magnetic field vector with a cadence of 128 Hz and averages these measurements to lower rates to fit within its telemetry allocation. The magnetometer consists of two triaxial fluxgate sensors. Because Venus Express uses a design inherited from the magnetometer-less Mars Express mission, no substantive efforts were made to determine the magnetic cleanliness of the spacecraft and its payload. To obtain scientifically useful magnetic measurements in this unfavourable environment, the magnetometer has a dual triaxial sensor arranged in a gradiometer configuration. Both sensors take measurements simultaneously, to enable separation of spacecraft-generated stray field from the ambient field[20]. The outboard sensor is mounted at the tip of a one-metre deployable boom, while the inboard sensor is directly attached to the spacecraft with a separation of 10 cm from the top panel of the spacecraft. The magnetometer has a dynamic range varying between ± 32.8 nT and $\pm 8{,}388.6$ nT with a corresponding digital resolution from 1 to 128 pT. The default range for the outboard sensor is set to ± 262 nT with a resolution of 8 pT. The default range for the inboard sensor is ± 524 nT. A artificial magnetic field of $\pm 10{,}000$ nT can independently be applied to each sensor for compensation of any disturbing spacecraft stray field. The instrument operates continually at Venus. After switching on, the magnetometer automatically operates in a standard mode with both sensors at a 1 Hz data rate. During a typical science orbit, the magnetometer is switched to fast mode at 32 Hz one hour before periapsis and switched to standard mode one hour after periapsis. In addition, the instrument can operate in a high-resolution burst mode of 128 Hz to detect electromagnetic (whistler mode) waves associated with Venus lightning[21]. Initially, this operating mode was used for two minutes at periapsis.

Received 13 February; accepted 7 June 2007.

1. Phillips, J. L. & Russell, C. T. Upper limit on the intrinsic magnetic field of Venus. *J. Geophys. Res.* **92,** 2253–2263 (1987).
3. Luhmann, J. G. The solar wind interaction with Venus. *Space Sci. Rev.* **44,** 241–306 (1986).
4. Brace, L. H. & Kliore, A. J. The structure of the Venus ionosphere. *Space Sci. Rev.* **55,** 81–164 (1991).
5. Phillips, J. L. & McComas, D. L. The magnetosheath and magnetotail of Venus. *Space Sci. Rev.* **55,** 1–80 (1991).
6. Luhmann, J. G., Ledvina, S. A. & Russell, C. T. Induced magnetospheres. *Adv. Space Res.* **33,** 1905–1912 (2004).
7. Russell, C. T., Luhmann, J. G. & Strangeway, R. J. The solar wind interaction with Venus through the eyes of the Pioneer Venus Orbiter. *Planet. Space Sci.* **54,** 1482–1495 (2006).
8. Zhang, T. L. *et al.* Magnetic field investigation of the Venus plasma environment: expected new results. *Planet. Space Sci.* **54,** 1336–1343 (2006).
9. Titov, D. V. *et al.* Venus Express science planning. *Planet. Space Sci.* **54,** 1279–1297 (2006).
10. Spreiter, J. R., Summers, A. L. & Rizzi, A. W. Solar wind flow past nonmagnetic planets. *Venus Mars Planet. Space Sci.* **18,** 1281–1299 (1970).
11. Slavin, J. A. *et al.* The solar wind interaction with Venus: Pioneer Venus observations of bow shock location and structure. *J. Geophys. Res.* **85,** 7625–7641 (1980).
12. Russell, C. T. *et al.* Solar and interplanetary control of the location of the Venus bow shock. *J. Geophys. Res.* **93,** 5461–5469 (1988).
13. Verigin, M. I. *et al.* Plasma near Venus from the Venera 9 and 10 wide-angle analyzer data. *J. Geophys. Res.* **83,** 3721–3728 (1978).
14. Zhang, T. L., Luhmann, J. G. & Russell, C. T. The magnetic barrier at Venus. *J. Geophys. Res.* **96,** 11145–11153 (1991).
15. Riedler, W. *et al.* Magnetic fields near Mars: first results. *Nature* **341,** 604–607 (1989).
16. Bertucci, C. *et al.* Magnetic field draping enhancement at the Martian magnetic pileup boundary from Mars Global Surveyor observation. *Geophys. Res. Lett.* **30,** 1099, doi:10.1029/2002GL015713 (2003).
17. Slavin, J. A. *et al.* Solar wind flow about the terrestrial planets. 2. Comparison with gas dynamic theory and implications for solar-planetary interactions. *J. Geophys. Res.* **88,** 19–35 (1983).
18. Zhang, M. H. G., Luhmann, J. G., Kliore, A. J. & Kim, J. A post-Pioneer Venus reassessment of the Martian dayside ionosphere as observed by radio occultation methods. *J. Geophys. Res.* **95,** 14829–14839 (1990).
19. Zhang, T. L., Luhmann, J. G. & Russell, C. T. The solar cycle dependence of the location and shape of the Venus bow shock. *J. Geophys. Res.* **95,** 14961–14967 (1990).
20. Ness, N. F., Behannon, K. W., Lepping, R. P. & Schatten, K. H. Use of two magnetometers for magnetic field measurements on a spacecraft. *J. Geophys. Res.* **76,** 3565–3573 (1971).
21. Russell, C. T., Strangeway, R. J. & Zhang, T. L. Lightning detection on the Venus Express mission. *Planet. Space Sci.* **54,** 1344–1351 (2006).

Acknowledgements The work at Graz is partially supported by the ASAP. The work at TU-Braunschweig is supported by Astrium-EADS. The work in Imperial College is supported by the Particle Physics and Astronomy Research Council (PPARC). The work at UCLA was supported by the National Aeronautics and Space Administration. The work in Slovakia is supported by the Slovak Research and Development Agency. The work in China is supported by the CAS International Partnership Program for Creative Research Teams.

Author Contributions T.L.Z. is the principal investigator of the Venus Express magnetic field investigation.

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 T.L.Z. (tielong.zhang@oeaw.ac.at).

©2007 Nature Publishing Group

LETTERS

The structure of Venus' middle atmosphere and ionosphere

M. Pätzold[1], B. Häusler[2], M. K. Bird[3], S. Tellmann[1], R. Mattei[2], S. W. Asmar[4], V. Dehant[5], W. Eidel[2], T. Imamura[6], R. A. Simpson[7] & G. L. Tyler[7]

The atmosphere and ionosphere of Venus have been studied in the past by spacecraft with remote sensing[1–4] or *in situ* techniques[3,4]. These early missions, however, have left us with questions about, for example, the atmospheric structure in the transition region from the upper troposphere to the lower mesosphere (50–90 km) and the remarkably variable structure of the ionosphere. Observations become increasingly difficult within and below the global cloud deck (<50 km altitude), where strong absorption greatly limits the available investigative spectrum to a few infrared windows and the radio range. Here we report radio-sounding results from the first Venus Express Radio Science[5] (VeRa) occultation season. We determine the fine structure in temperatures at upper cloud-deck altitudes, detect a distinct day–night temperature difference in the southern middle atmosphere, and track day-to-day changes in Venus' ionosphere.

The atmosphere and ionosphere of Venus were sounded by VeRa at 2.3 and 8.4 GHz (the S and X bands, respectively) during the first occultation season from mid-July to the end of August 2006. As seen from the Earth, the spacecraft disappeared behind the planetary disk on every 24-hour orbit during this period[5]. The radio ray path is refracted according to the local state of the atmosphere, the degree of refraction being proportional to the electron density (ionosphere) or the neutral gas density (atmosphere). The VeRa observations are the first to use two simultaneous one-way downlink frequencies driven by an Ultrastable Oscillator for radio sounding of Venus' atmosphere. This enables observations during both ingress and egress, rather than being restricted to ingress (as with the Pioneer Venus Orbiter in the majority of cases). Forty-two profiles were obtained from 21 occultation passes during the first occultation season (one ingress and one egress per pass). These profiles reveal the ionospheric structure at altitudes of 100–500 km and the neutral atmosphere in the altitude range 50–90 km. Longitudes span the range from 296° E to 352° E for the daytime and from 189° E to 251° E for the night-time profiles. Additional information on the geometry and coverage during the first Venus Express occultation season, which covers the Venus Express orbits numbered 81–131 (2006 day of year: DOY 192–242), is provided in the Supplementary Information. Details on the experiment performance[5], the extraction of atmospheric profiles[6] and results from additional components of the VeRa investigation, including the first bistatic radar experiments, are published elsewhere.

To derive the temperature profile, however, we need to know the composition of the atmosphere (96.5% CO_2 and 3.5% N_2)[7] and the assumption of an upper boundary temperature. Figure 1 shows three VeRa temperature profiles derived from the same data set taken

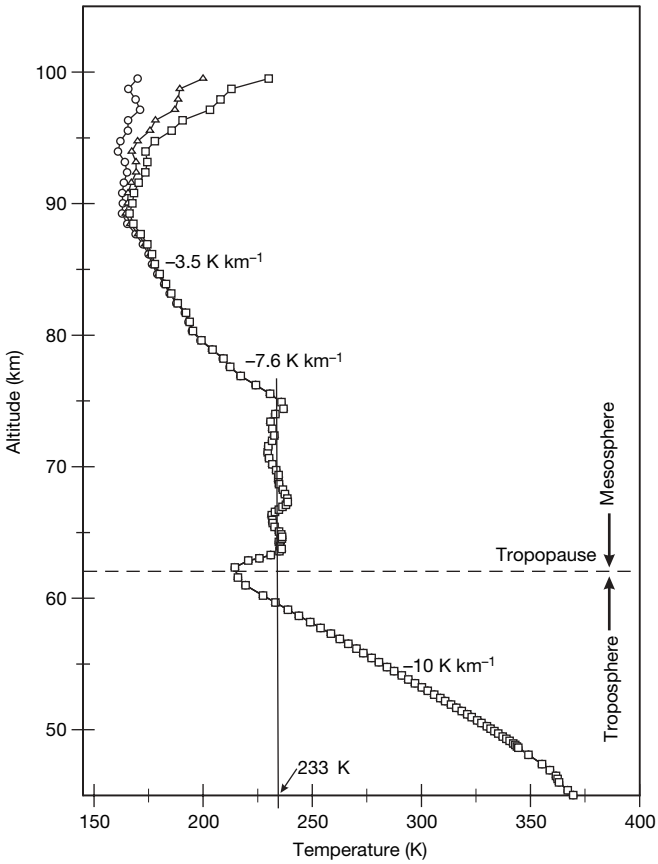


Figure 1 | Temperature profiles from DOY 234, 2006 at latitude 71° N, derived with three different upper boundary temperature conditions of 170, 200 and 230 K. Regardless of the upper boundary condition, all three profiles converge to the same temperature distribution below 90 km. The temperature shows a constant cooling (lapse rate) of $-10\ \mathrm{K\,km^{-1}}$ within the cloud deck below 60 km. The inversion at 62 km marks the tropopause, the transistion from the troposphere to the mesosphere. On average, the temperature is isothermal (233 K) within the upper cloud deck up to 75 km, but displays significant fine structure, a series of small inversions, with amplitudes larger than the measurement error. Above 75 km, the cooling rate observed above the transition is $-7.6\ \mathrm{K\,km^{-1}}$ up to 80 km altitudes and then drops to $-3.5\ \mathrm{K\,km^{-1}}$ above this height.

[1]Rheinisches Institut für Umweltforschung, Abt. Planetenforschung, Universität zu Köln, Aachener Strasse 201–209, 50931 Köln, Germany. [2]Institut für Raumfahrttechnik, Universität der Bundeswehr München, 85577 Neubiberg, Germany. [3]Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany. [4]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109, USA. [5]Observatoire Royal de Belgique, 3 Avenue Circulaire, 1180 Brussels, Belgium. [6]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, 3-1-1 Yoshinodai, Sagamihara, Japan. [7]Space, Telecommunications, and Radioscience Laboratory (MC 9515), Stanford University, Stanford, California 94305-4020, USA.

©2007 Nature Publishing Group

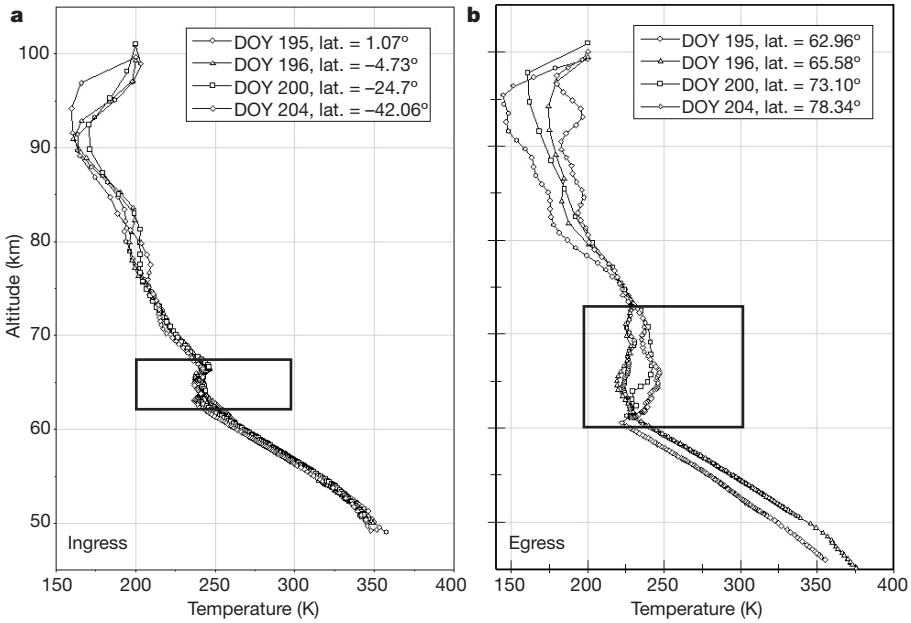


Figure 2 | Venus temperature height profiles from VeRa radio-sounding observations. Four temperature profiles are shown from occultation ingress at southern mid-latitudes (**a**) and egress at northern high polar latitudes (**b**). Reliable measurements of the received radio signal are not available at altitudes below about 50 km. The profiles are well determined below 90 km. Temperature inversions are detected at altitudes from 60 to 70 km. This inversion range (boxed area) occurs within the cloud deck and is more confined in the mid-latitudes (65–69 km) than at the poles (63–74 km).

during occultation egress on 2006 DOY 234 at northern polar latitudes with three different upper boundary temperatures. Regardless of the particular upper boundary temperature selected, all three profiles converge rapidly into the same temperature profile, providing confidence in the reliability of the temperature determination below 90 km. The sensitivity of VeRa extends the atmospheric profiles to higher altitudes, thus providing better continuity to the lower boundary of the Venus Express instrument SPICAV's atmospheric observations[8]. Figure 1 typically reveals the fine-scale thermal structure in the transition region from the upper troposphere to the lower mesosphere within the upper cloud deck. The Venus atmosphere reaches the one-bar pressure level at an altitude of about 50 km. Below this height, the radio signals also suffer significant absorption by the gas and cannot be used to sound the atmosphere at all below the super-refractive limiting height of about 35 km (ref. 9).

Figure 2 shows examples of temperature profiles derived from ingress (Fig. 2a) and egress (Fig. 2b) measurements. A decrease in temperature with increasing height (lapse rate, $-10\,\mathrm{K\,km^{-1}}$) is found below an altitude of 60−65 km. Above this level is a region of thermal inversions—a series of warmings and coolings within the range of altitudes attributed to the upper cloud deck. The series of inversions is constrained by two sharp lower and upper boundaries below 65 km and above 70 km, respectively, where again a clear cooling trend is observed. The distance between these boundaries is larger by a factor of three in the polar latitudes (Fig. 2b) than in the equatorial to mid-latitudes (Fig. 2a). The lower inversion is more pronounced in the polar latitudes and increases in altitude towards lower latitude, consistent with a convergence to the profiles seen in

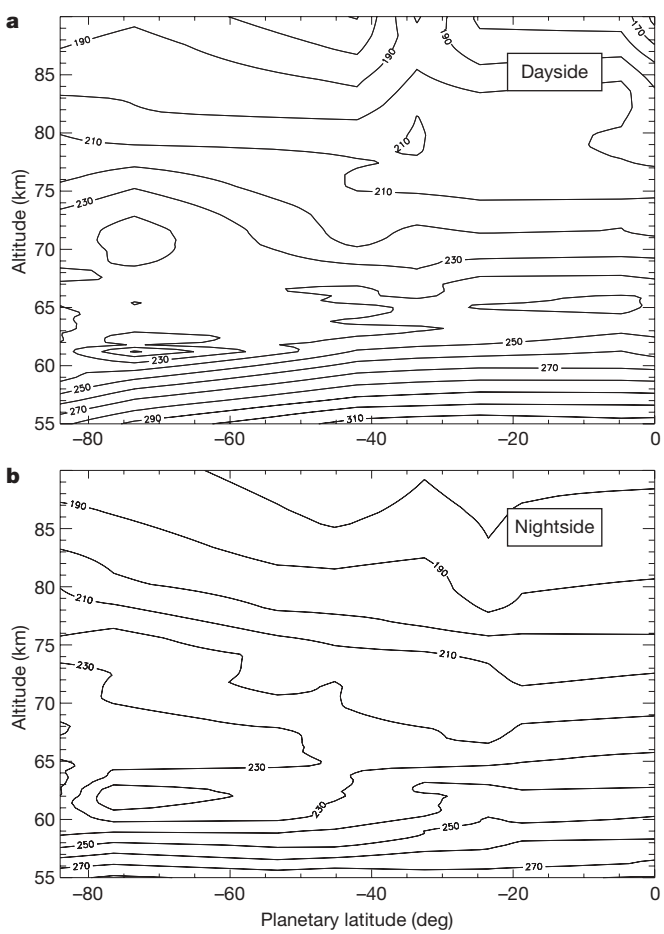


Figure 3 | Venus temperature maps derived from VeRa radio occultation data. Contours (in K) are generated using all available temperature profiles in the southern latitude range from 0° to 90° for the dayside (**a**, 50° < solar zenith angle < 90°) and nightside (**b**, 90° < solar zenith angle < 115°) of Venus. The range in altitude is from 55 to 90 km. The daytime temperature increases by about 30–40 K from the southern pole to the equator at altitudes below 60 km. This contrasts with the roughly constant night-time temperature below 60 km. The inferred result is a day–night equatorial temperature difference that reaches about 40 K at 55 km altitude. The "cold collar", a temperature depression encircling the pole in both hemispheres[15], is clearly observed on both the dayside and the nightside between 60° and 80° latitude at 63 km altitude. The collar extends to 40° latitude during the night in the southern hemisphere.

©2007 Nature Publishing Group

Fig. 2a. Although, on average, the temperature is roughly isothermal within these two boundaries at an average neutral scale height of 5.2 km, the series of inversions implies a thermal fine structure within the upper cloud deck never before seen with the clarity that is best exemplified in Fig. 1. Similar inversion behaviour was observed with the Pioneer Venus Orbiter[9], but was much more pronounced at polar latitudes than was observed with Venus Express.

All ingress profiles from the southern hemisphere are combined together in the contour plots of Fig. 3. Significant differences are found between the daytime and night-time temperature distributions (a colour plot of Fig. 3 is provided in the Supplementary Information).

As an example, Fig. 4a shows an electron density profile from the ionosphere from 2006 DOY 196, observed in the equatorial region at 15:20 local time and a solar zenith angle of 50°. The noise level of the retrieved ionospheric profiles is $\sim 2 \times 10^9\ m^{-3}$, which is 0.5% of the peak electron density. The VeRa ionospheric profiles are thus of the same high quality as those from the most recent Venus mission, Magellan[10]. We can readily identify the bottom and main layers, V1 and V2, respectively. Although the primary ionization product is CO_2^+ generated by photoionization of solar extreme-ultraviolet and soft X-rays, the main constituent of the V1 and V2 layers is O_2^+, formed by rapid molecular reactions of CO_2^+ with atomic oxygen[11]. The O^+ ion becomes important and controls the density above 180 km potentially in its own layer, V3 (ref. 11). The majority of electron density profiles display a bulge in the topside between 160 and 180 km that is not documented in theoretical models of ion and electron production[11].

Figure 4b shows a collection of daytime electron density profiles. Large variations can be seen in the topside density distribution at the same time that the lower V1 and V2 layers are relatively stable, with peak densities close to those expected for the solar minimum conditions of summer 2006 (ref. 12). A smaller peak density is observed for the profile on DOY 212 at high solar zenith angle. The base of the daytime ionosphere is stable at an altitude of 120 km. The peak density altitude of V1 at about 140 km marks the altitude where the soft X-ray intensity has decreased by e^{-1}; radiation from the broad X-ray spectrum penetrating deeper into the atmosphere does not produce detectable ionization below 120 km.

The ionopause, the distinct boundary between the solar wind flow and the planetary ionosphere, is present in the altitude range 225–375 km. These are quite typical for the applicable range of solar zenith angles near the solar activity minimum[12,13]. The location is clearly defined by the strong electron density drop into the noise level towards higher altitudes. The four profiles of Fig. 4b cover a time interval of two weeks. While the ionopause heights on DOY 196 and 200 are nearly the same, the height is almost 100 km higher on DOY 202. Ten days later the ionopause returned to its previously lower level. This may imply that the balance between the solar wind dynamic pressure and the ionospheric plasma pressure was stable over many days, but then changed significantly from one day to another. It is remarkable that all profiles display roughly the same electron density gradient near the ionopause—falling into the noise over an altitude range of about 30 km.

The ionosphere was also sounded on the Venus nightside (Fig. 5). We distinguish in this case between profiles from the 'deep' night with solar zenith angles greater than 98° and those close to the terminator (solar zenith angle ≈ 90–98°). Profiles from the deep night vary in character, sometimes showing significant unstructured ionization, while at other times no ionization can be detected. This is consistent with Pioneer Venus Orbiter results during solar

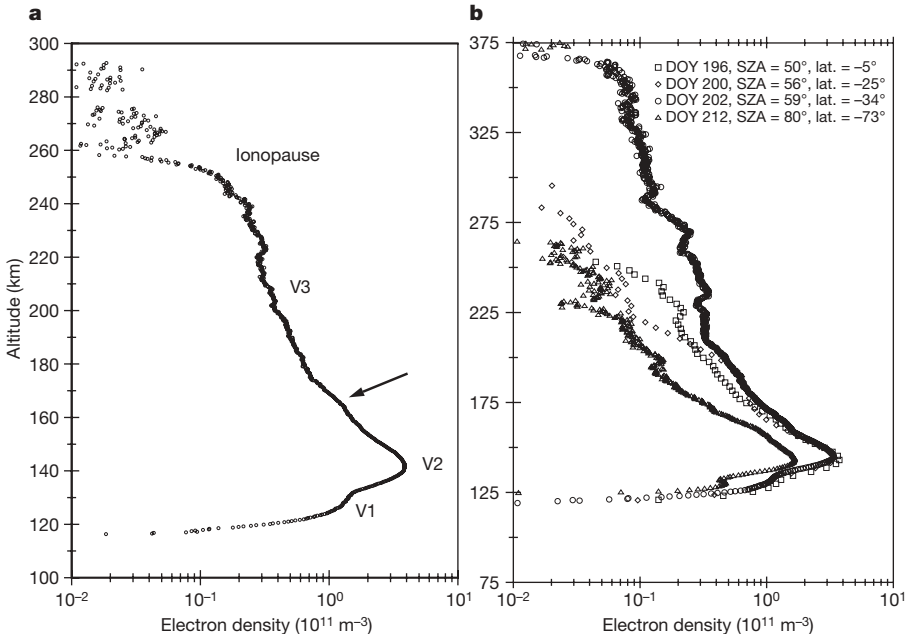


Figure 4 | Daytime electron density profiles in the ionosphere of Venus. **a**, The main features of a Venus ionospheric electron density height profile derived from the VeRa occultation data. This ingress example of 15 July 2006 (DOY 196), derived from simultaneous Doppler recordings of the occulted 2.3 and 8.4 GHz signals and from the associated differential Doppler data calculated from both sets of measurements, shows the ionosphere at solar zenith angle 50°, local time 15:20 and latitude −4.7°. The differential Doppler data are sensitive only to changes in the ionized medium along the radio propagation path—in this case, primarily in the Venus ionosphere. This profile is essentially identical with those derived from the X-band or S-band data sets alone, providing confidence in the quality of the derived electron density profiles even from single-frequency Doppler data. Electron density layers are identified as the secondary and main layers V1 and V2, respectively. A V3 region is formed above 180 km controlled by the O^+ ion[11]. A bulge in the topside between 160 and 180 km altitude (arrow) is apparent in the majority of the observed daytime profiles, and is not explained by models of the Venus ionosphere. The ionopause is located between 250 and 275 km in this example. **b**, VeRa daytime ionospheric electron density profiles on four different days. These ingress profiles were derived from the 8.4 GHz Doppler data. Clearly identified are the ionosopheric base at 120 km, the secondary and main layers V1 and V2, respectively, the V3 layer, a highly variable topside, and an ionopause at altitudes varying between 225 and 375 km. The latitude and solar zenith angle (SZA) associated with each profile is indicated.

©2007 Nature Publishing Group

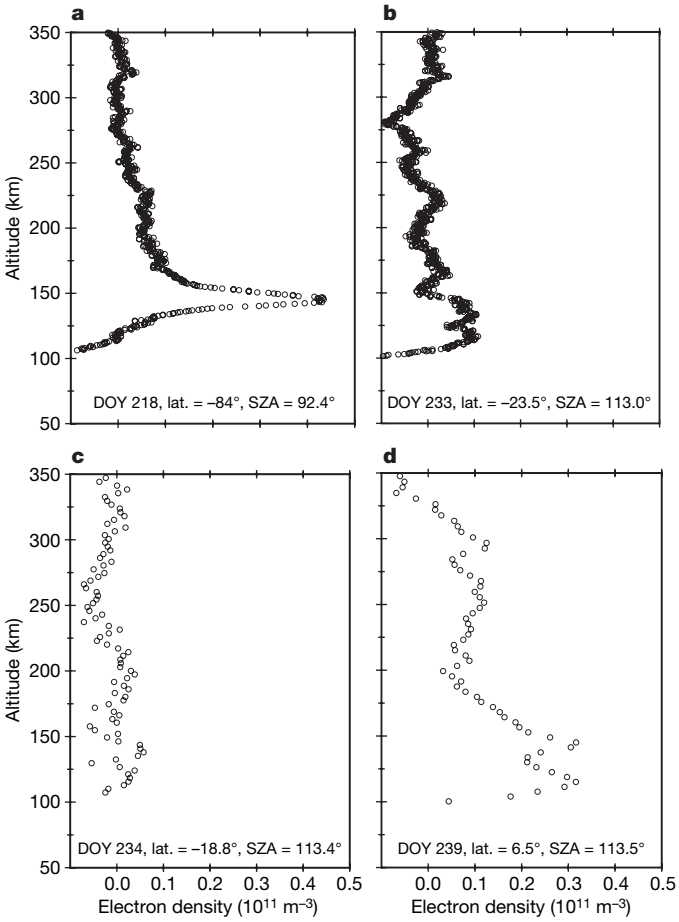


Figure 5 | Four electron density profiles from the nightside of Venus. a, An example from high polar latitudes at -84° and near the terminator at solar zenith angle solar zenith angle = 92.4°, when the upper atmosphere is still illuminated and a structured ionosphere has formed. The peak density at 145 km is about an order of magnitude less than that of the mid-latitude daytime profiles shown in Fig. 4. **b**–**d**, Profiles at various latitudes obtained for solar zenith angles near 113° and local time 05:00 (that is, unilluminated up to almost 500 km). Only weak, possibly sporadic, ionization was observed in some profiles. In contrast to the logarithmic representation of the daytime profiles in Fig. 4, these night-time profiles are plotted on a linear scale.

minimum, when ion convection from the dayside ionosphere is reduced[14]. Profiles from locations close to the terminator at polar latitudes show a layered structure similar to that observed during the day, albeit with much lower peak densities. The reduced densities result from the longer atmospheric absorption lengths and corresponding reduction in ionizing flux at these high solar zenith angles.

Received 23 May; accepted 5 September 2007.

1. Mariner Stanford Group. Venus: ionosphere and atmosphere as measured by dual-frequency radio occultation of Mariner V. *Science* **158,** 1678–1683 (1967).
2. Howard, H. T. *et al.* Venus: mass, gravity field, atmosphere, and ionosphere as measured by the Mariner 10 dual-frequency radio system. *Science* **183,** 1297–1301 (1974).
3. Hunten, D. M., Colin, L., Donahue, T. M. & Moroz, V. I. (eds) *Venus* (Univ. Arizona Press, Tucson, 1983).
4. Bougher, S. W., Hunten, D. M. & Phillips, R. J. (eds) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment* (Univ. Arizona Press, Tucson, 1997).
5. Häusler, B. *et al.* Radio science investigations by VeRa onboard the Venus Express spacecraft. *Planet. Space Sci.* **54,** 1315–1335 (2006).
6. Fjeldbo, G., Kliore, A. J. & Eshleman, V. R. The neutral atmosphere of Venus as studied with the Mariner V radio occultation experiments. *Astron. J.* **76,** 123–140 (1971).
7. von Zahn, U., Kumar, S., Niemann, H. & Prinn, R. in *Venus* (eds Hunten, D. M., Colin, L., Donahue, T. M. & Moroz, V. I.) 299–430 (Univ. Arizona Press, Tucson, 1983).
8. Bertaux, J. L. *et al.* A warm layer in Venus'cryosphere and high-altitude measurements of HF, HCl, H_2O and HDO. *Nature* doi:10.1038/nature05974 (this issue).
9. Kliore, A. J. & Patel, I. R. Vertical structure of the atmosphere of Venus from Pioneer Venus Orbiter radio occultations. *J. Geophys. Res.* **85,** 7957–7962 (1980).
10. Jenkins, J. M., Steffes, P. G., Hinson, D. P., Twicken, J. D. & Tyler, G. L. Radio occultation studies of the Venus atmosphere with the Magellan spacecraft. 2. Results from the October 1991 experiments. *Icarus* **110,** 79–93 (1994).
11. Fox, J. L. & Sung, K. Y. Solar activity variations of the Venus thermosphere/ionosphere. *J. Geophys. Res.* **106,** 21305–21336 (2001).
12. Fox, J. L. & Kliore, A. J. in *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment* (eds Bougher, S. W., Hunten, D. M. & Philips, R. J.) 161–188 (Univ. Arizona Press, Tucson, 1997).
13. Kliore, A. J. & Luhmann, J. G. Solar cycle effects on the structure of the electron density profiles in the dayside ionosphere of Venus. *J. Geophys. Res.* **96,** 21281–21289 (1991).
14. Kliore, A. J., Luhmann, J. G. & Zhang, M. H. G. The effect of the solar cycle on the maintenance of the nightside ionosphere of Venus. *J. Geophys. Res.* **96,** 11065–11071 (1991).
15. Taylor, F. W. *et al.* Structure and meteorology of the middle atmosphere of Venus: infrared remote sensing from the Pioneer orbiter. *J. Geophys. Res.* **85,** 7963–8006 (1980).

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

Acknowledgements We thank H. Svedhem, F. Jansen, the Project Science Team at ESTEC and the Flight Control Team at ESOC for continuous support. The German and the US part of VeRa are supported by DLR, Bonn-Oberkassel and by a contract with NASA, respectively.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.P. (paetzold@geo.uni-koeln.de).

©2007 Nature Publishing Group

LETTERS

Lightning on Venus inferred from whistler-mode waves in the ionosphere

C. T. Russell[1], T. L. Zhang[2], M. Delva[2], W. Magnes[2], R. J. Strangeway[1] & H. Y. Wei[1]

The occurrence of lightning in a planetary atmosphere enables chemical processes to take place that would not occur under standard temperatures and pressures[1–3]. Although much evidence has been reported for lightning on Venus[4–8], some searches have been negative[9–11] and the existence of lightning has remained controversial. A definitive detection would be the confirmation of electromagnetic, whistler-mode waves propagating from the atmosphere to the ionosphere. Here we report observations of Venus' ionosphere that reveal strong, circularly polarized, electromagnetic waves with frequencies near 100 Hz. The waves appear as bursts of radiation lasting 0.25 to 0.5 s, and have the expected properties of whistler-mode signals generated by lightning discharges in Venus' clouds.

We have begun again receiving scientific data from Venus with the arrival in April 2006 of the Venus Express spacecraft, inserted into a high inclination elliptical orbit with periapsis at 73° latitude and an altitude near 250 km. Among its instrumentation, Venus Express carries a magnetometer that can sample at rates as high as 128 Hz (ref. 12). When sampling at this rate, the magnetometer is weakly low-pass-filtered before digitization, permitting some 'aliasing', in which signals enter the telemetry above the highest frequency that can be reconstructed by the digital telemetry. This frequency is half the sampling frequency, or 64 Hz. This mode was specifically designed to capture lightning by taking advantage of the broad natural bandwidth of the magnetometer within the limitation of a fixed number of bits transmitted to ground from the spacecraft. Owing to the large data volume produced, this mode was initially limited to 2 min per orbit, centred on periapsis. Although the amplitudes and durations of the signals are preserved, the folding of the digitally sampled signals around half the sample period (aliasing) can affect the apparent dispersion of the signals and the handedness of the waves.

Here we discuss measurements from 37 initial orbits in May and June 2006, when the interference from the reaction wheels on the body of this spacecraft remained outside our analysis bandwidth of 42–60 Hz. We illustrate the properties of the wave events seen, with examples from day 160 (9 June 2006; Fig. 1). When the transmission began, signals appeared immediately, suggesting that the generation process was already active. The signals stopped after 15 s—probably when the spacecraft moved away from the source, or when the magnetic-field-aligned propagation path became disconnected from the source. The signals on other passes appeared to begin and end during the transmissions, and some continued to at least the end of the pass. All signals resemble those shown here. They are bursty, with rapidly varying amplitudes, and variable interburst spacings and durations. The lower wave amplitude in the x direction (see Fig. 1 legend) indicates that the direction of propagation of the wave is roughly parallel or antiparallel to the solar direction.

We can find the minimum variance direction (direction of phase propagation) of these signals quite precisely, and rotate the measurements into the principal axis coordinate system (Fig. 2a). The individual bursts each have slightly different directions. The direction of propagation in the example is aligned within 13° of the background magnetic field, as expected for a whistler-mode wave. This direction is very well determined, as the eigenvalues of the principal axis determination are in the ratio 13 to 10 to 1. The near equality of the two largest eigenvalues indicates that the wave is nearly circularly polarized, again as expected for a whistler-mode wave propagating close to parallel to the magnetic field. The wave components are very nearly 90° out of phase, as expected. A hodogram, in which the tip of the perturbation magnetic field is drawn for successive samples (Fig. 2b), confirms that the wave is nearly circularly polarized and its perturbation confined to a plane. Again, these are the characteristics expected for a whistler-mode wave, the only electromagnetic mode expected to propagate at these frequencies for the observed background magnetic field of 23 nT. Some of the events on the other three passes are more elliptically polarized, as would occur for propagation at an angle to the magnetic field.

These signals are not due to spacecraft interference. There are two magnetometers on board, one on the spacecraft deck and one on a one-metre boom. Spacecraft signals produce different amplitudes on the two sensors, whereas real signals produce the same amplitude.

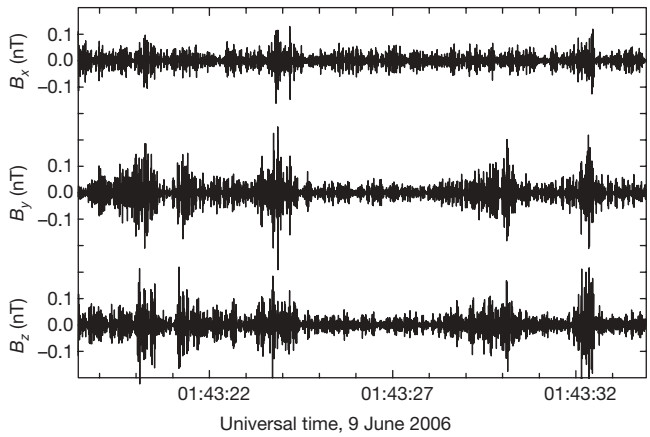


Figure 1 | Examples of the wave events. Signals recorded by the three sensors of the fluxgate magnetometer at an altitude of 305 km at 05:16 local time and a solar zenith angle of 91°. Data sampled at 128 Hz have been rotated into Venus solar orbital (VSO) coordinates (to give B_x, B_y, B_z) and bandpass filtered to display signals from 42 to 60 Hz to reduce spacecraft interference. The coordinate system has its x direction pointed towards the Sun, y is opposite planetary motion, and z is along the orbit pole. Signals appeared to be in progress at the start of the interval (shown here) but ceased after 01:43:32 UT (h:min:s). The spacecraft moves about 100 km in 15 s.

[1]Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90095-1567, USA. [2]Space Research Institute, Austrian Academy of Sciences, Graz, A-8042, Austria.

©2007 Nature Publishing Group

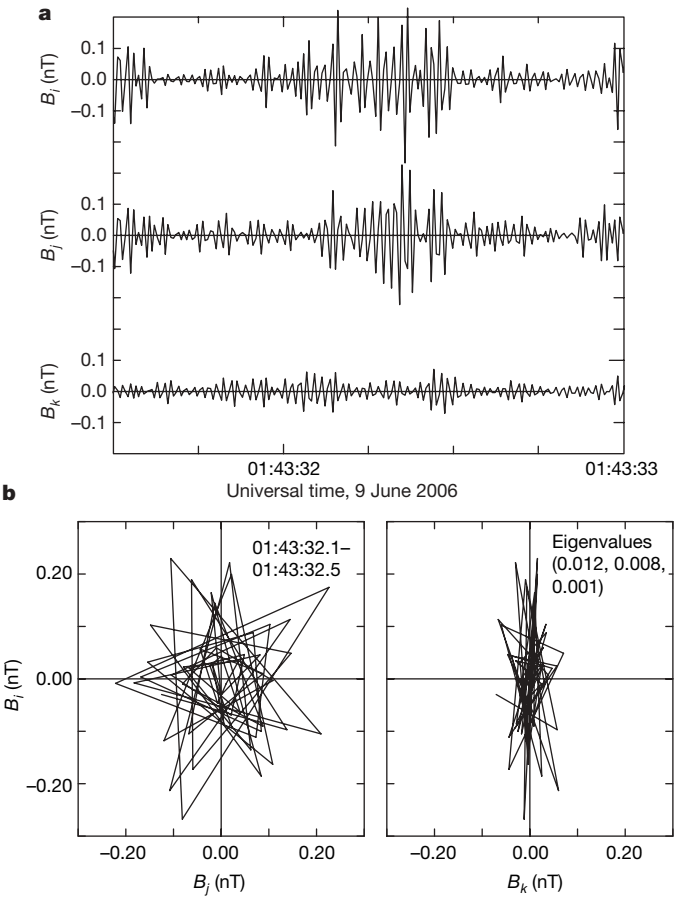


Figure 2 | Propagation of the signals. a, Burst of signal at 01:43:32 UT rotated into the principal axis coordinate system, and bandpass filtered as in Fig. 1; B_i, B_j and B_k are the field components in principal axis coordinates, with B_i along the direction of maximum variance and B_k along the direction of minimum variance. **b**, Hodogram of signal burst in the principal axis system. The wave is propagating along the direction (0.918, 0.340, 0.205) in VSO coordinates.

Subtracting the two sensor outputs then will exactly cancel the real signal, leaving noise. This test was performed, and showed that the bursts were registered identically by the two sensors over the entire 15-s period the waves were present.

As the electromagnetic energy flux of these waves diminished with altitude when surveyed by Pioneer Venus[13], these waves must have a source in Venus' atmosphere. We know of no other possible source of natural signals with these properties in Venus' ionosphere other than lightning in the atmosphere. The waveforms indicate an impulsive current source similar to terrestrial lightning. Also, the intermittent appearance of the bursts is like the occurrence pattern expected for a weather-associated phenomenon. As the mission proceeds, we will obtain further data and be able to gather occurrence statistics, including the occurrence rate across the dayside for which we have no information from previous missions. We further note that the night and morning local times at which these results were obtained coincide with the region of lowest expected occurrence rate from previous measurements[14].

On 50% of these passes, there are bursts of noise greater than 0.1 nT peak-to-peak, and on about 10% of the passes, there are bursts with amplitudes of greater than 0.2 nT peak-to-peak. The bursts have durations of about 0.2–0.5 s. If we assume that the magnetometer can detect signals over a footprint that has a radius equal to the spacecraft's altitude, it can see about 0.06% of the planet. If so, the burst rate of 0.03 s^{-1} observed on these orbits corresponds to a planet-wide rate of 50 s^{-1}, about half that at Earth[15]. However, the high-latitude region beneath the spacecraft at periapsis may not be representative of the entire planet.

In short, the initial data resolve the controversy concerning the presence of electromagnetic signals in Venus' ionosphere consistent with generation by atmospheric lightning. These signals are seen extensively, even well away from the regions of expected maximum occurrence based on the Pioneer Venus survey. They are bursty, and occur intermittently as on Earth, but occur at a lower frequency, in part because the ionosphere cannot transmit much higher frequencies. They are sufficiently intense to be well characterized by the fluxgate magnetometer at all local times examined, and the rate of occurrence may be similar to the terrestrial rate.

Received 23 January; accepted 1 May 2007.

1. Desch, S. J., Borucki, W. J., Russell, C. T. & Bar-Nun, A. Progress in planetary lightning. *Rep. Prog. Phys.* **65**, 955–997 (2002).
2. Krasnopolsky, V. A. A sensitive search for nitric oxide in the lower atmosphere of Venus and Mars: Detection on Venus and upper limit for Mars. *Icarus* **192**, 80–91 (2006).
3. Bar-Nun, A. Production of carbon and nitrogen species by thunderstorms on Venus. *Icarus* **42**, 338–342 (1990).
4. Ksanfomaliti, L. V. Electrical activity of the atmosphere of Venus, I. Measurements on descending probes. *Kosmich. Issled.* **21**, 279–296 (1983).
5. Krasnopol'sky, V. A. Lightning and nitric oxide on Venus. *Planet. Space Sci.* **31**, 1363–1369 (1983).
6. Russell, C. T. & Scarf, F. L. Evidence for lightning on Venus. *Adv. Space Res.* **10**, 125–136 (1990).
7. Gurnett, D. A. *et al.* Lightning and plasma wave observations from the Galileo flyby of Venus. *Science* **253**, 1522–1525 (1991).
8. Hansell, S. A., Wells, W. K. & Hunten, D. M. Optical detection of lightning on Venus. *Icarus* **117**, 345–351 (1995).
9. Sagdeev, R. Z. *et al.* Overview of VEGA Venus balloon in-situ meteorological measurements. *Science* **231**, 1411–1414 (1986).
10. Borucki, W. J., Dyer, J. W., Phillips, J. R. & Pham, P. Pioneer Venus Orbiter search for Venusian lightning. *J. Geophys. Res. A* **96**, 11033–11043 (1991).
11. Gurnett, D. A. *et al.* Non-detection at Venus of high-frequency radio signals characteristic of terrestrial lightning. *Nature* **409**, 313–315 (2001).
12. Zhang, T. L. *et al.* Magnetic field investigation of the Venus plasma environment: Expected new results from Venus Express. *Planet. Space Sci.* **54**, 1336–1343 (2006).
13. Russell, C. T., von Dornum, M. & Strangeway, R. J. VLF bursts in the night ionosphere of Venus: Estimates of the Poynting flux. *Geophys. Res. Lett.* **16**, 579–582 (1989).
14. Russell, C. T., Von Dornum, M. & Scarf, F. L. Impulsive signals in the night ionosphere of Venus: Comparison of results obtained below the local electron gyro frequency with those above. *Adv. Space Res.* **10**, 37–40 (1990).
15. Russell, C. T. Planetary lightning. *Annu. Rev. Earth Planet. Sci.* **21**, 43–87 (1993).

Acknowledgements This work was supported by NASA and the Austrian Academy of Science.

Author Contributions T.L.Z. is the Principal Investigator of the Venus Express magnetometer. M.D. led the development of the gradiometer technique used, and assisted in calibration and commissioning of the magnetometer. W.M. was the engineer responsible for the successful implementation of the investigation, and R.J.S. provided guidance in the design, based on Pioneer Venus observations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to C.T.R. (ctrussell@igpp.ucla.edu).

©2007 Nature Publishing Group

REVIEWS

Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes

Tom A. Rapoport[1]

A decisive step in the biosynthesis of many proteins is their partial or complete translocation across the eukaryotic endoplasmic reticulum membrane or the prokaryotic plasma membrane. Most of these proteins are translocated through a protein-conducting channel that is formed by a conserved, heterotrimeric membrane-protein complex, the Sec61 or SecY complex. Depending on channel binding partners, polypeptides are moved by different mechanisms: the polypeptide chain is transferred directly into the channel by the translating ribosome, a ratcheting mechanism is used by the endoplasmic reticulum chaperone BiP, and a pushing mechanism is used by the bacterial ATPase SecA. Structural, genetic and biochemical data show how the channel opens across the membrane, releases hydrophobic segments of membrane proteins laterally into lipid, and maintains the membrane barrier for small molecules.

For almost 40 years, researchers have been fascinated by the question of how proteins are transported across or are integrated into membranes. Pioneering work by G. Palade[1] demonstrated that in eukaryotic cells secretory proteins cross the endoplasmic reticulum membrane before being transported in vesicles to the plasma membrane. The laboratories of G. Blobel and C. Milstein then discovered that these proteins are directed to the endoplasmic reticulum membrane by signal sequences[2,3]. A little later, signal sequences were also found to direct the translocation of proteins across the bacterial plasma membrane[4,5]. Genetic experiments identified components required for translocation, initially in bacteria and later in yeast[6–8], and the establishment of an *in vitro* system initiated biochemical studies[9]. All of these achievements set the stage for investigations into the molecular mechanism of translocation, which will be the focus of this review.

Proteins transported across the eukaryotic endoplasmic reticulum membrane or the prokaryotic plasma membrane include soluble proteins, such as those ultimately secreted from the cell or localized to the endoplasmic reticulum lumen, and membrane proteins, such as those in the plasma membrane or in other organelles of the secretory pathway. Soluble proteins cross the membrane completely and usually have amino-terminal, cleavable signal sequences, the major feature of which is a segment of 7–12 hydrophobic amino acids. Membrane proteins have different topologies in the lipid bilayer, with one or more transmembrane segments composed of about 20 hydrophobic amino acids; the hydrophilic regions of these proteins either cross the membrane or remain in the cytosol. Both types of proteins are handled by the same machinery within the membrane: a protein-conducting channel. The channel allows soluble polypeptides to cross the membrane and hydrophobic transmembrane segments of membrane proteins to exit laterally into the lipid phase.

Structure of the translocation channel

The translocation channel is formed from a conserved heterotrimeric membrane protein complex, called the Sec61 complex in eukaryotes and the SecY complex in bacteria and archaea (for more details, see refs 10 and 11). The α- and γ-subunits show significant sequence conservation, and both subunits are essential for the function of the channel and for cell viability. The β-subunits are not essential; they are similar in eukaryotes and archaea, but show no obvious homology to the corresponding subunit in bacteria.

The α-subunit forms the pore of the channel, as initially shown by experiments in which photoreactive probes were systematically placed at different positions of a stalled translocating polypeptide[12]; all positions predicted to be within the membrane cross-linked only to the α-subunit of the Sec61 complex, indicating that this subunit surrounds the polypeptide chain during its passage across the membrane. In addition, experiments in which the purified Sec61/SecY complex was reconstituted into proteoliposomes showed that it is the essential membrane component for protein translocation[13–15]. The channel has an aqueous interior, as demonstrated by electrophysiology experiments[16] and by measurements of the fluorescence lifetime of probes incorporated into a translocating polypeptide chain[17,18].

The crystal structure of an archaeal SecY complex provided important insight into how the α-subunit forms the channel[10]. The structure is probably representative of complexes from all species, as indicated by sequence conservation and by the similarity to a lower-resolution structure of the *Escherichia coli* SecY complex, determined by electron microscopy from two-dimensional crystals[19,20]. Viewed from the cytosol, the channel has a square shape (Fig. 1a). The α-subunit is divided into two halves, transmembrane segments 1–5 and 6–10. The loop between transmembrane segments 5 and 6 at the back of the α-subunit serves as a hinge, allowing the α-subunit to open at the front—the 'lateral gate'. The γ-subunit links the two halves of the α-subunit at the back by extending one transmembrane segment diagonally across their interface. The β-subunit makes contact only with the periphery of the α-subunit, probably explaining why it is dispensable for the function of the complex.

The ten helices of the α-subunit form an hourglass-shaped pore that consists of cytoplasmic and external funnels, the tips of which meet about half way across the membrane (Fig. 1b). Whereas the

[1]Howard Hughes Medical Institute and Department of Cell Biology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, USA.

©2007 Nature Publishing Group

cytoplasmic funnel is empty, the external funnel is plugged by a short helix. The crystal structure therefore represents a closed channel but, as will be discussed later, biochemical data indicate how it can open and translocate proteins. The constriction of the hourglass-shaped channel is formed by a ring of six hydrophobic residues that project their side chains radially inward. The residues forming this 'pore ring' are amino acids with bulky, hydrophobic side chains.

Different modes of translocation

The channel alone is a passive pore; it must associate with partners that provide a driving force for translocation. Depending on the partner, there are three known ways in which the channel can function.

In co-translational translocation, the main partner is the ribosome. This mode of translocation is found in all cells and is used for the translocation of secretory proteins as well as for the integration of most membrane proteins. Co-translational translocation begins with a targeting phase. The signal or transmembrane sequence of a growing polypeptide chain is recognized by the signal-recognition particle (SRP); after this, the ribosome–nascent-chain–SRP complex binds to the membrane, first by an interaction between SRP and its membrane receptor, and then by an interaction between the ribosome and the translocation channel (Fig. 2; for review of the targeting phase, see refs 21 and 22). The elongating polypeptide chain subsequently moves directly from the tunnel inside the ribosome into the associated membrane channel. GTP hydrolysis is required for chain elongation by the ribosome, but polypeptide movement through the channel is independent of nucleotide hydrolysis[23]. In the case of membrane proteins, certain polypeptide segments do not enter the channel, but instead emerge from the ribosome–channel junction into the cytosol[24], generating a cytosolic domain.

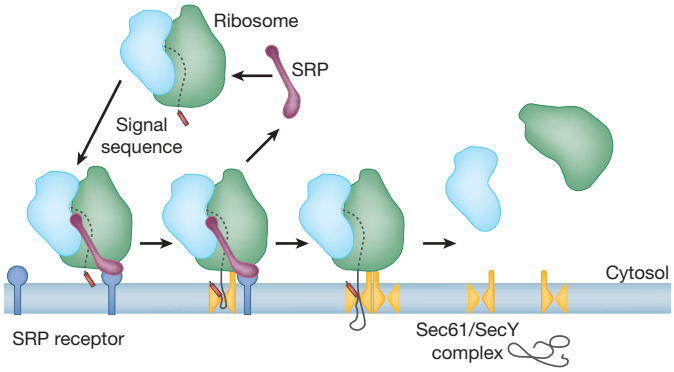


Figure 2 | Model of co-translational translocation. The scheme is mostly based on experiments with the eukaryotic system, but is probably similar for all organisms.

In most if not all cells, some proteins are transported after completion of their synthesis, that is, post-translationally. This pathway seems to be used by a larger fraction of proteins in simpler organisms, such as bacteria and yeast, perhaps because in these fast-growing cells translocation does not always keep pace with translation. This pathway is used mostly by soluble proteins, such as secretory proteins, which possess only moderately hydrophobic signal sequences that cause them to escape recognition by the SRP during their synthesis[25,26]. These proteins need to remain unfolded or loosely folded after their release from the ribosome[27]. Post-translational translocation occurs by different mechanisms in eukaryotes and bacteria.

During post-translational translocation in yeast, and probably in all eukaryotes, the channel partners with another membrane-protein complex, the tetrameric Sec62/Sec63 complex, and with the luminal chaperone BiP, a member of the Hsp70 family of ATPases[28,29]. In *Saccharomyces cerevisiae*, the Sec62/Sec63 complex consists of the essential Sec62 and Sec63 proteins as well as the dispensable Sec71 (also known as Sec66) and Sec72 proteins. Mammalian cells only have Sec62 and Sec63 (refs 30 and 31). Translocation begins with the binding of a translocation substrate to the channel (Fig. 3). During this step, all cytosolic chaperones are released from the substrate[32]. Once the polypeptide is inserted into the channel, its translocation occurs by a ratcheting mechanism[33]. The polypeptide chain in the channel can slide in either direction by brownian motion, but

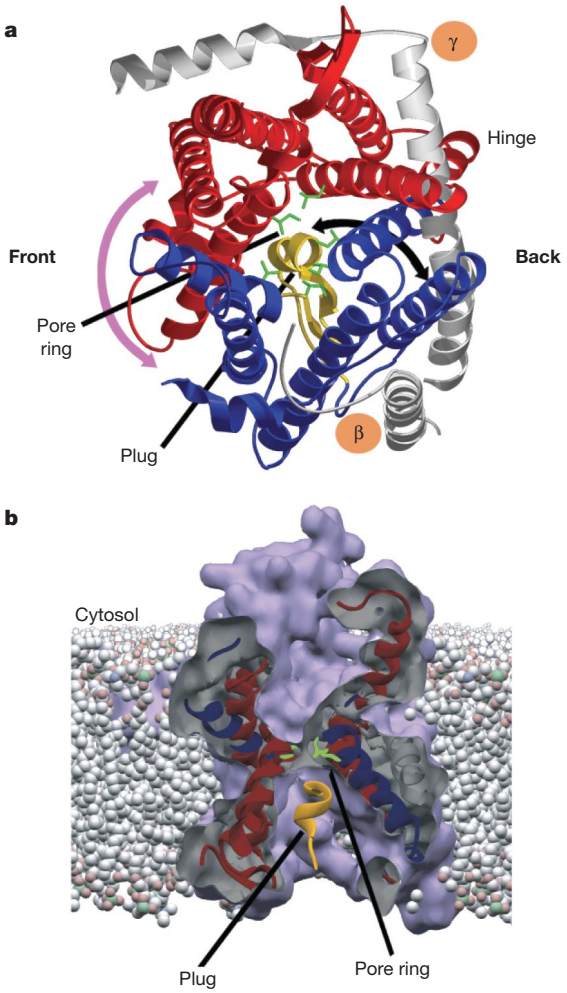


Figure 1 | The translocation channel. a, View from the cytosol of the crystal structure of the SecY complex from *Methanococcus jannaschii*. The α-subunit consists of two halves, transmembrane segments 1–5 and 6–10 (in blue and red, respectively), which can open the lateral gate at the front (purple double-headed arrow). The β- and γ-subunits are shown in grey. In the closed channel, the plug (in yellow) is in the centre of the α-subunit. Plug movement towards the back (black double-headed arrow) opens the channel across the membrane. The pore-ring residues are indicated in green. **b**, Cross-sectional view of the channel from the side.

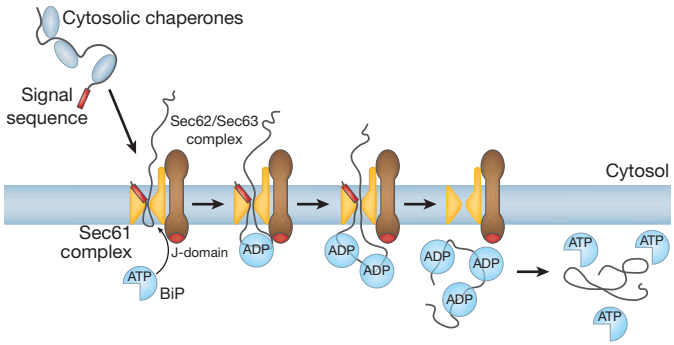


Figure 3 | Model of post-translational translocation in eukaryotes. It is possible that oligomers of the Sec61 complex mediate translocation, similar to the situation with the other modes of translocation (Figs 2 and 4).

©2007 Nature Publishing Group

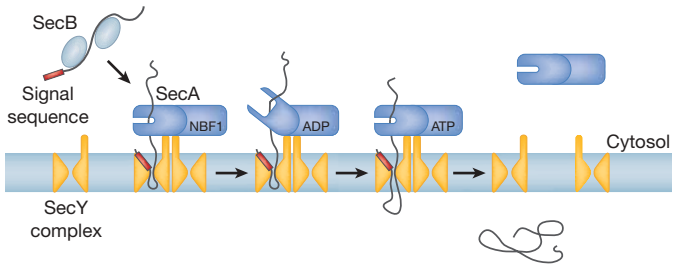


Figure 4 | Model of post-translational translocation in bacteria.

its binding to BiP inside the lumen of the endoplasmic reticulum prevents movement back into the cytosol, resulting in net forward translocation. ATP-bound BiP with an open peptide-binding pocket interacts with the J-domain of Sec63, which causes rapid ATP hydrolysis and closure of the peptide-binding pocket around the translocation substrate. J-domain-activated BiP has a low binding specificity[34], allowing it to interact with essentially any polypeptide segment that emerges from the channel into the lumen of the endoplasmic reticulum. When the polypeptide has moved sufficiently in the forward direction, the next BiP molecule can bind. This process is repeated until the polypeptide chain has completely traversed the channel. Finally, exchange of ADP for ATP opens the peptide-binding pocket and releases BiP.

In bacterial post-translational translocation, the channel partners with the cytosolic ATPase SecA. SecA has several domains, including two nucleotide-binding folds (NBF1 and NBF2) that bind the nucleotide between them and move relative to one another during the ATP hydrolysis cycle. The other domains also move, perhaps allowing SecA to alternate between the closed and open conformations that are observed in crystal structures[35,36]. A large groove in the open state might close around the polypeptide chain, because it is similar in dimensions to those seen in other proteins that interact with a wide range of substrates. Several experiments indicate that SecA functions as a monomer during translocation[37–40], but the issue is still controversial[41–43]. The translocation of many substrates begins with their binding to SecB, a cytosolic chaperone[44] (Fig. 4). Next, SecA interacts with SecB and accepts the polypeptide, probably binding both the signal sequence and the segment following it[45–47]. The subsequent transfer of the polypeptide into the channel requires a full cycle of ATP hydrolysis by SecA[48]. Once inserted into the channel, the substrate is translocated by a 'pushing' mechanism[49]. Although the details are not yet clear, a plausible mechanism assumes that the polypeptide-binding groove of SecA closes around the polypeptide chain and moves towards the channel, pushing the polypeptide into it (Fig. 4). The size of SecA means that it is unlikely that it inserts deeply into the SecY channel, as proposed earlier[49,50]. On nucleotide hydrolysis, the groove opens, releases the peptide, and moves away to 'grab' the next segment of the substrate. This cycle continues until the entire polypeptide is translocated. An electrochemical gradient across the membrane stimulates translocation *in vitro* and is required *in vivo*[51], but it is unclear how the gradient is used.

Archaea probably have both co- and post-translational translocation[52,53], but it is unknown how post-translational translocation occurs because these organisms lack SecA, the Sec62/Sec63 complex and BiP.

Opening the channel across the membrane

In all modes, the translocation of a secretory protein begins with its insertion into the channel. The polypeptide inserts as a loop (Fig. 5a), with the signal sequence intercalated into the walls of the channel and the segment distal to it located in the pore proper[54]. Opening of the channel for loop insertion probably occurs in two steps. The first is the binding of a channel partner—the ribosome, the Sec62/Sec63 complex or SecA. This event probably destabilizes interactions that keep the plug in the centre of the Sec61/SecY molecule. The ribosome and SecA interact with cytosolic loops in the carboxy-terminal half of Sec61/SecY[55], and these might transmit conformational changes to other parts of the molecule, resulting in transient displacement of the plug and continuous opening and closing of the lateral gate. This is supported by the observation of increased ion conductance when non-translating ribosomes are bound to the channel[16].

The second step is the intercalation of the hydrophobic segment of a signal sequence into the lateral gate. Photocrosslinking experiments with a yeast *in vitro* system show that the hydrophobic region of a bound signal sequence forms a helix of about two turns, which is intercalated between transmembrane segments 2b and 7 (ref. 56). The signal sequence can also be crosslinked to phospholipid molecules, indicating that it sits at the interface between channel and lipid. The binding of the signal sequence would separate transmembrane segments 2b and 7 and further destabilize plug interactions, causing the plug to move from the centre of Sec61/SecY into a cavity at the back of the molecule. Disulphide-bridge crosslinking shows that the plug indeed comes close to the transmembrane segment of the γ-subunit during translocation[57,58]. This model is also consistent with the observation that many mutations that allow the translocation of proteins with defective signal sequences (signal-suppressor mutations) would be expected to destabilize the closed channel[10,59].

Finally, the open state of the channel would be fixed by the insertion of the polypeptide segment distal to the signal sequence into the pore proper. During subsequent translocation, the signal sequence stays put, whereas the rest of the polypeptide moves through the pore. The plug could only return to the centre of Sec61/SecY when the polypeptide chain has left the pore. At some point, the signal sequence is cleaved by signal peptidase and is then further degraded by the signal peptide peptidase, a presenilin-like enzyme that cleaves the hydrophobic segment within the membrane[60].

The pore

The crystal structure indicates that a single copy of the Sec61/SecY complex forms the pore; according to this model, a polypeptide

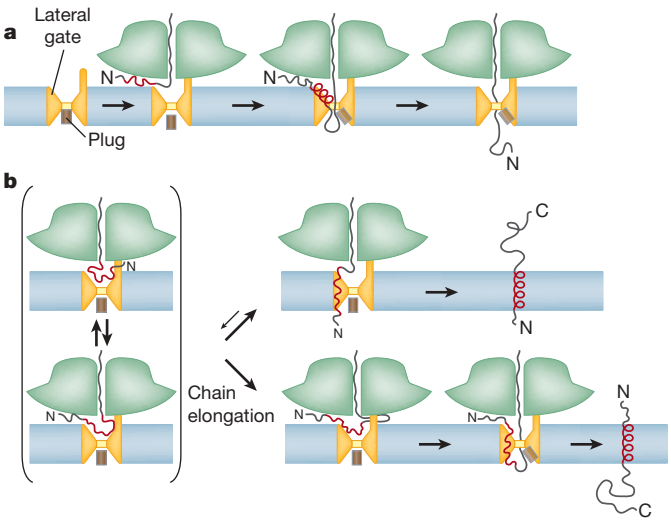


Figure 5 | Different stages of translocation. a, Translocation of a secretory protein. The red line indicates the hydrophobic region of a signal sequence. Depicted is the co-translational mode of translocation, but similar schemes can be envisioned for the other modes. For simplicity, only the translocating Sec61/SecY copy is shown. **b**, Translocation of membrane proteins. When a hydrophobic transmembrane sequence (in red) has emerged from the ribosome, it can bind reversibly in several conformations. If the hydrophobic sequence is long and the N terminus is not retained in the cytosol, it can flip across the membrane (upper panel). If the N terminus is retained in the cytosol and the polypeptide chain is further elongated by the translating ribosome (indicated by the loop between the ribosome and channel), the C terminus can translocate across the membrane (lower panel).

©2007 Nature Publishing Group

would move from the cytoplasmic funnel, through the pore ring, into the external funnel. Previously, it was thought that several copies of the Sec61/SecY complex would assemble to form a hydrophilic pore in the membrane, but the crystal structure shows that, similar to all other membrane proteins, a single SecY complex has an entirely hydrophobic belt around its exterior surface[10]. Systematic disulphide cross-linking experiments, in which one cysteine is placed in a translocation substrate and another at various positions of SecY, show that the polypeptide chain indeed passes through the centre of the SecY complex and makes contact with it only at the waist of the hourglass-shaped channel[61]. The aqueous interior of the channel, its shape and the lack of interactions of the hydrophobic pore residues with the hydrophilic polypeptide backbone all help to minimize the energy required to move a translocation substrate through the membrane.

The diameter of the pore ring, as observed in the crystal structure, is too small to allow the passage of most polypeptide chains. Because even peptide loops of 13 residues and bulky side chains can move through the channel[62,63], pore widening has to be postulated, which could be mediated by movements of the helices to which the pore-ring residues are attached. The flexibility of the pore ring is supported by molecular dynamics simulations as well as by electrophysiology experiments[64–67]. The intercalation of a signal sequence at the front of Sec61/SecY may cause additional widening of the pore, as is required for loop insertion of a polypeptide chain. The estimated maximum dimensions of the pore on the basis of the crystal structure are 15×20 ångström (Å) (ref. 10). A polypeptide in the pore could therefore form an α-helix, but no tertiary structure, in agreement with experimental data[68].

Results from fluorescence-quenching experiments have indicated that the pore is much larger (40–60 Å) than conceivable with a single Sec61/SecY molecule[69]. The fluorescent probes were incorporated into translocating nascent chains that are associated with stalled, membrane-bound ribosomes; quenching agents as large as NAD^+ were able to move through the channel to collide with the probes. These data could be reconciled with the crystal structure if two or more Sec61/SecY complexes associated at their front surfaces and opened their lateral gates to fuse their pores into a larger channel. Although a different arrangement—a back-to-back orientation—is suggested by the two-dimensional structure of the *E. coli* SecY complex and by cross-linking data[19,70], the model would be consistent with an electron microscopy structure in which a translating *E. coli* ribosome was proposed to be associated with two nearly front-to-front oriented SecY molecules[71]. However, this structure is based on a low-resolution electron-density map (~15 Å), and the docking of the SecY crystal structure required its drastic modification. The position and orientation of both SecY molecules are different from that of the single SecY molecule that is seen in more recent electron microscopy structures of non-translating ribosome–SecY complexes[72]. Disulphide-bridge cross-linking experiments argue against fusion of different pores because they show that, during SecA-mediated translocation, both the signal sequence and the mature region of a polypeptide chain are located in the same SecY molecule[48]. A detergent-solubilized translocation intermediate also contains just one copy of SecY associated with one SecA and one translocation substrate molecule[38].

If, then, the channel is formed from only one Sec61/SecY molecule, how can one rationalize the fluorescence-quenching results? Although there is currently no good explanation, it should be noted that the fluorescent probes were located deep inside the ribosome, and therefore the same large diameter (40–60 Å) must be assumed for the ribosome tunnel, a size that does not agree with that seen in ribosome structures (<20 Å) determined by crystallography or cryo-electron microscopy[73–75]. One might argue that the tunnel could widen under certain conditions, but this could not be caused simply by the arrival of a nascent polypeptide, because a chain is present in some of the structures. It is also difficult to see how the rigid ribosomal RNA that lines most of the tunnel could undergo such a dramatic change. It has been proposed that the structural methods give inaccurate answers because they are obtained in detergent[76], but the disulphide cross-linking experiments were performed with functional translocation intermediates in intact membranes[48].

Oligomeric translocation channels

Although the pore is formed by only one Sec61/SecY molecule, translocation of a polypeptide chain seems to be mediated by oligomers. This is best supported by the observation that a SecY molecule defective in SecA-mediated translocation can be rescued by linking it covalently with a wild-type SecY copy[48]. Disulphide-bridge cross-linking showed that SecA binds through its NBF1 domain to a non-translocating SecY copy, and moves the polypeptide chain through a neighbouring SecY copy[48] (Fig. 4). In this model, the static interaction with the non-translocating SecY molecule would prevent complete detachment of SecA when its peptide-binding domain moves away from the translocating SecY molecule to 'grab' the next polypeptide segment.

The Sec61/SecY complex probably forms oligomers during co-translational translocation as well. When a ribosome–nascent-chain–SRP complex binds to the SRP receptor, a domain of SRP undergoes a conformational change that exposes a site on the ribosome to which a single Sec61/SecY molecule could bind[77]—probably the one seen in recent electron microscopy structures of non-translating ribosome–SecY complexes[72]. The bound SecY molecule is close to the point where a polypeptide exits the ribosome and could thus become the translocating copy. At a later stage of translocation, SRP completely detaches from the ribosome, and one or more additional copies of the Sec61/SecY complex may associate, as suggested by cross-linking and freeze-fracture electron microscopy experiments[78,79]. These copies could stabilize the ribosome–channel junction and possibly recruit other components, such as signal peptidase and oligosaccharyl transferase, which are needed for polypeptide modification, or the translocon-associated protein complex (TRAP), the function of which is still unclear. On termination of translocation, dissociation of the Sec61/SecY oligomers could facilitate the release of the ribosome from the membrane[78]. Dissociable oligomers may also allow the Sec61/SecY complex to change channel partners and modes of translocation.

Electron microscopy structures of detergent-solubilized ribosome–channel complexes suggested the presence of three or four Sec61 molecules[80,81]. However, the low resolution of these structures makes it difficult to distinguish between protein and additional density contributed by lipid and detergent. It is therefore possible that only one Sec61 molecule is present, whereas the other Sec61 copies were lost during solubilization, similar to the dissociation seen with SecA-interacting SecY oligomers[38].

The emerging concept of homo-oligomeric channels, in which only one copy is active at any given time, may be a common theme in protein translocation. Such a situation may apply for PapC, which is involved in the secretion of pilus subunits across the outer membrane of *E. coli*[82], and for Tom40 and Tim22, which are involved in protein transport across the outer and inner mitochondrial membrane[83,84], respectively.

Membrane-protein integration

During the synthesis of a membrane protein, all transmembrane segments move from the aqueous interior of the channel, through the lateral gate, into the lipid phase. The gate is formed where short segments of four transmembrane segments at the front of Sec61/SecY link the two halves of the molecule[10]. The resulting seam in the wall of the channel is probably weak once these transmembrane segments are no longer contacted and stabilized by the plug. The lateral gate may therefore continuously open and close, exposing polypeptide segments located in the aqueous channel to the surrounding hydrophobic lipid phase. Segments that are sufficiently long and hydrophobic to span the entire membrane would exit the channel through

©2007 Nature Publishing Group

the lateral gate, simply by partitioning between aqueous and hydrophobic environments. This model is supported by photo-cross-linking experiments that examined the lateral exit of a transmembrane segment in different translocation intermediates[85], as well as by the agreement between a hydrophobicity scale derived from peptide partitioning into an organic solvent and the tendency of a peptide to span the membrane[86]. The model is also consistent with molecular dynamics simulations that indicate that lipid molecules do not pass the lateral gate rapidly[87]. The size of the channel indicates that transmembrane segments exit laterally one by one or in pairs. Hydrophilic segments between the transmembrane segments would move alternately from the ribosome, through the aqueous channel, to the external side of the membrane, or emerge between the ribosome and channel into the cytosol. Movement into the cytosol would utilize a gap between the ribosome and channel, which can be visualized in electron microscopy structures[81].

In contrast to a signal sequence, which always has its N terminus in the cytosol, the first transmembrane segment of a membrane protein can have its N terminus on either side of the membrane, depending on the amino-acid sequence of the protein. In a multi-spanning membrane protein, the first transmembrane segment often determines the orientation of the subsequent ones[88], although there are exceptions in which internal transmembrane segments have a preferred orientation regardless of the behaviour of preceding transmembrane segments (for a review, see ref. 89).

A model for how the orientation of the first transmembrane segment is determined is shown in Fig. 5. When a hydrophobic segment emerges from the ribosome, it can intercalate reversibly in two different orientations into the lateral gate (Fig. 5b). If the hydrophobic sequence is long and the N terminus is not retained in the cytosol by positive charges or by the folding of the preceding polypeptide segment (for a review, see ref. 89), it can flip across the channel and subsequently exit it laterally into the lipid phase. If the N terminus is retained in the cytosol and the polypeptide chain is further elongated, the C terminus can translocate across the channel, inserting the polypeptide as a loop, as in the case of a secretory protein.

Maintaining the permeability barrier

The channel must prevent the free movement of small molecules, such as ions or metabolites, both in its resting state and when translocating a polypeptide. Maintaining the membrane barrier is particularly important for prokaryotes, because the proton gradient across the membrane is the main source of their energy. The endoplasmic reticulum membrane may be somewhat leaky to small molecules[90], but it must also prevent the free flow of ions.

Results from fluorescence-quenching experiments with endoplasmic reticulum membranes suggest a complex molecular mechanism to maintain the membrane permeability barrier. In this model, the resting channel has a pore size of 9–15 Å, which is closed at the luminal end by BiP, either directly or indirectly[91]. During the translocation of a secretory protein, the channel widens to 40–60 Å, and the luminal seal is lost, replaced by a cytoplasmic seal from the translating ribosome[69,91]. When a multi-spanning membrane protein is synthesized, the seals provided by the ribosome and BiP alternate, depending on whether the nascent chain is directed to the endoplasmic reticulum lumen or the cytosol[92]. The pore is closed by BiP before the transmembrane segment in a nascent chain reaches the channel, implying that the ribosome recognizes the nascent chain as a membrane protein. Fluorescence resonance energy transfer (FRET) and chemical modification experiments support the idea that a transmembrane segment can form an α helix inside the ribosome[93,94], but it is difficult to see how the ribosome tunnel, with its mostly hydrophilic surface, could recognize a long hydrophobic sequence. In addition, consecutive transmembrane segments move in the same direction inside the ribosome, but would have to transmit opposing signals to the ribosome-associated channel. A tight seal between the ribosome and channel is also at odds with electron microscopy structures that reveal a gap of 12–15 Å between them[80,95]. Finally, this model does not explain how the membrane barrier is maintained in the absence of a ribosome (in post-translational translocation) or in the absence of BiP (in prokaryotes).

The crystal structure indicates a simpler model, in which the membrane barrier is formed by the channel itself, with both the plug and pore ring contributing to the seal[10]. Electrophysiology experiments show that the resting SecY channel, reconstituted in the absence of other components into a planar membrane, is indeed impermeable to ions and water, and opens on plug displacement[67]. In the active channel, the pore ring would fit like a gasket around the translocating polypeptide chain, thereby restricting the passage of small molecules during protein translocation. The seal would not be expected to be perfect; in fact, a partial loss of the electrochemical gradient is observed on accumulation of an arrested translocation intermediate in *E. coli* membranes[96]. Leakage is probably compensated for by powerful ion pumps. During the synthesis of a multi-spanning membrane protein, the seal would be provided in an alternating manner either by the nascent chain in the pore or—once the chain has left the pore—by the plug returning to the centre of Sec61/SecY. This model needs further experimental verification, but it would explain how the membrane barrier can be maintained in both co- and post-translational translocation, and why a gap between the ribosome and channel may not compromise the barrier.

Surprisingly, plug-deletion mutants are viable in *S. cerevisiae* and *E. coli*, and have only moderate translocation defects[97–99]. However, the crystal structure of these mutants shows that new plugs are formed from neighbouring polypeptide segments[99]. The new plugs still seal the closed channel, but they have lost many interactions that normally keep the plug in the centre of SecY. This results in continuous channel opening and closing, as observed in electrophysiology experiments[67]. In addition, facilitated channel opening in the plug-deletion mutants permits polypeptides with defective or even missing signal sequences to be translocated[98,99]. The plug sequences are only poorly conserved among Sec61/SecY channels, supporting the idea that promiscuous segments can seal the channel and lock it in its closed state.

Perspective

Progress during the past several years has led to a detailed understanding of protein translocation, in particular of the function of the Sec61/SecY channel. Nevertheless, major questions in the field are still controversial and unresolved. Further progress will require the combination of different approaches. To address how the channel maintains the permeability barrier, electrophysiology experiments are needed to complement the fluorescence-quenching method, particularly because results from the latter are difficult to reconcile with structural data. Important questions in co-translational translocation include how the SRP receptor and channel collaborate, how many Sec61/SecY complexes participate in translocation, and how the ribosome ultimately dissociates from the channel. The precise role of the Sec62/Sec63 components in post-translational translocation, and the mechanism by which SecA moves polypeptides, are also important issues for the future. Membrane-protein integration is still particularly poorly understood, and new methods are required to follow the integration of individual transmembrane segments during the synthesis of a multi-spanning membrane protein. Another unresolved issue concerns other translocation components, such as the translocon-associated membrane protein (TRAM) protein and the translocon-associated protein (TRAP) complex in mammalian cells, or the YidC, SecD and SecF proteins in prokaryotes. These components may be required as chaperones for the folding of transmembrane segments, or to increase the efficiency of translocation of some substrates, but their precise functions remain to be clarified. Much of the progress in the field will depend on the generation of high-resolution structures, with the 'holy grail' being a picture of the 'translocon in action', in which a channel associated with both a

©2007 Nature Publishing Group

partner and a translocating polypeptide chain is visualized at atomic detail. However, even lower resolution electron microscopy structures of active translocation complexes are eagerly awaited. The study of the Sec translocation system remains an exciting area of research and is likely to serve as a paradigm for other protein translocation systems, such as those in mitochondria, chloroplasts and peroxisomes.

1. Palade, G. Intracellular aspects of the process of protein synthesis. *Science* **189,** 347–358 (1975).
2. Blobel, G. & Sabatini, D. D. Ribosome-membrane interaction in eukaryotic cells. *Biomembranes* **2,** 193–195 (1971).
3. Milstein, C., Brownlee, G. G., Harrison, T. M. & Mathews, M. B. A possible precursor of immunoglobulin light chains. *Nature New Biol.* **239,** 117–120 (1972).
4. Inouye, H. & Beckwith, J. Synthesis and processing of an *Escherichia coli* alkaline phosphatase precursor *in vitro*. *Proc. Natl Acad. Sci. USA* **74,** 1440–1444 (1977).
5. Halegoua, S., Sekizawa, J. & Inouye, M. A new form of structural lipoprotein of outer membrane of *Escherichia coli*. *J. Biol. Chem.* **252,** 2324–2330 (1977).
6. Emr, S. D., Hanley-Way, S. & Silhavy, T. Suppressor mutations that restore export of a protein with a defective signal sequence. *Cell* **23,** 79–88 (1981).
7. Oliver, D. B. & Beckwith, J. E. *coli* mutant pleiotropically defective in the export of secreted proteins. *Cell* **25,** 765–772 (1981).
8. Deshaies, R. J. & Schekman, R. A yeast mutant defective at an early stage in import of secretory protein precursors into the endoplasmic reticulum. *J. Cell Biol.* **105,** 633–645 (1987).
9. Blobel, G. & Dobberstein, B. Transfer of proteins across membranes. II. Reconstitution of functional rough microsomes from heterologous components. *J. Cell Biol.* **67,** 852–862 (1975).
10. Van den Berg, B. *et al.* X-ray structure of a protein-conducting channel. *Nature* **427,** 36–44 (2004).
11. Osborne, A. R., Rapoport, T. A. & van den Berg, B. Protein translocation by the Sec61/SecY channel. *Annu. Rev. Cell Dev. Biol.* **21,** 529–550 (2005).
12. Mothes, W., Prehn, S. & Rapoport, T. A. Systematic probing of the environment of a translocating secretory protein during translocation through the ER membrane. *EMBO J.* **13,** 3937–3982 (1994).
13. Akimaru, J., Matsuyama, S., Tokuda, H. & Mizushima, S. Reconstitution of a protein translocation system containing purified SecY, SecE, and SecA from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **88,** 6545–6549 (1991).
14. Brundage, L., Hendrick, J. P., Schiebel, E., Driessen, A. J. & Wickner, W. The purified *E. coli* integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* **62,** 649–657 (1990).
15. Gorlich, D. & Rapoport, T. A. Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell* **75,** 615–630 (1993).
16. Simon, S. M. & Blobel, G. A protein-conducting channel in the endoplasmic reticulum. *Cell* **65,** 371–380 (1991).
17. Crowley, K. S., Reinhart, G. D. & Johnson, A. E. The signal sequence moves through a ribosomal tunnel into a noncytoplasmic aqueous environment at the ER membrane early in translocation. *Cell* **73,** 1101–1115 (1993).
18. Crowley, K. S., Liao, S. R., Worrell, V. E., Reinhart, G. D. & Johnson, A. E. Secretory proteins move through the endoplasmic reticulum membrane via an aqueous, gated pore. *Cell* **78,** 461–471 (1994).
19. Breyton, C., Haase, W., Rapoport, T. A., Kuhlbrandt, W. & Collinson, I. Three-dimensional structure of the bacterial protein–translocation complex SecYEG. *Nature* **418,** 662–665 (2002).
20. Bostina, M., Mohsin, B., Kuhlbrandt, W. & Collinson, I. Atomic model of the *E. coli* membrane-bound protein translocation complex SecYEG. *J. Mol. Biol.* **352,** 1035–1043 (2005).
21. Luirink, J. & Sinning, I. SRP-mediated protein targeting: structure and function revisited. *Biochim. Biophys. Acta* **1694,** 17–35 (2004).
22. Halic, M. & Beckmann, R. The signal recognition particle and its interactions during protein targeting. *Curr. Opin. Struct. Biol.* **15,** 116–125 (2005).
23. Connolly, T. & Gilmore, R. Formation of a functional ribosome–membrane junction during translocation requires the participation of a GTP-binding protein. *J. Cell Biol.* **103,** 2253–2261 (1986).
24. Mothes, W. *et al.* Molecular mechanism of membrane protein integration into the endoplasmic reticulum. *Cell* **89,** 523–533 (1997).
25. Ng, D. T., Brown, J. D. & Walter, P. Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J. Cell Biol.* **134,** 269–278 (1996).
26. Huber, D. *et al.* Use of thioredoxin as a reporter to identify a subset of *Escherichia coli* signal sequences that promote signal recognition particle-dependent translocation. *J. Bacteriol.* **187,** 2983–2991 (2005).
27. Huber, D. *et al.* A selection for mutants that interfere with folding of *Escherichia coli* thioredoxin-1 *in vivo*. *Proc. Natl Acad. Sci. USA* **102,** 18872–18877 (2005).
28. Deshaies, R. J., Sanders, S. L., Feldheim, D. A. & Schekman, R. Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. *Nature* **349,** 806–808 (1991).
29. Panzner, S., Dreier, L., Hartmann, E., Kostka, S. & Rapoport, T. A. Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. *Cell* **81,** 561–570 (1995).
30. Meyer, H. A. *et al.* Mammalian Sec61 is associated with Sec62 and Sec63. *J. Biol. Chem.* **275,** 14550–14557 (2000).
31. Tyedmers, J. *et al.* Homologs of the yeast Sec complex subunits Sec62p and Sec63p are abundant proteins in dog pancreas microsomes. *Proc. Natl Acad. Sci. USA* **97,** 7214–7219 (2000).
32. Plath, K. & Rapoport, T. A. Spontaneous release of cytosolic proteins from posttranslational substrates before their transport into the endoplasmic reticulum. *J. Cell Biol.* **151,** 167–178 (2000).
33. Matlack, K. E., Misselwitz, B., Plath, K. & Rapoport, T. A. BiP acts as a molecular ratchet during posttranslational transport of prepro-α factor across the ER membrane. *Cell* **97,** 553–564 (1999).
34. Misselwitz, B., Staeck, O. & Rapoport, T. A. J proteins catalytically activate Hsp70 molecules to trap a wide range of peptide sequences. *Mol. Cell* **2,** 593–603 (1998).
35. Hunt, J. F. *et al.* Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA. *Science* **297,** 2018–2026 (2002).
36. Osborne, A. R., Clemons, W. M. Jr & Rapoport, T. A. A large conformational change of the translocation ATPase SecA. *Proc. Natl Acad. Sci. USA* **101,** 10937–10942 (2004).
37. Or, E., Navon, A. & Rapoport, T. Dissociation of the dimeric SecA ATPase during protein translocation across the bacterial membrane. *EMBO J.* **21,** 4470–4479 (2002).
38. Duong, F. Binding, activation and dissociation of the dimeric SecA ATPase at the dimeric SecYEG translocase. *EMBO J.* **22,** 4375–4384 (2003).
39. Or, E., Boyd, D., Gon, S., Beckwith, J. & Rapoport, T. The bacterial ATPase SecA functions as a monomer in protein translocation. *J. Biol. Chem.* **280,** 9097–9105 (2004).
40. Alami, M., Dalal, K., Lelj-Garolla, B., Sligar, S. G. & Duong, F. Nanodiscs unravel the interaction between the SecYEG channel and its cytosolic partner SecA. *EMBO J.* **26,** 1995–2004 (2007).
41. Jilaveanu, L. B., Zito, C. R. & Oliver, D. Dimeric SecA is essential for protein translocation. *Proc. Natl Acad. Sci. USA* **102,** 7511–7516 (2005).
42. de Keyzer, J. *et al.* Covalently dimerized SecA is functional in protein translocation. *J. Biol. Chem.* **280,** 35255–35260 (2005).
43. Mitra, K., Frank, J. & Driessen, A. Co- and post-translational translocation through the protein-conducting channel: analogous mechanisms at work? *Nature Struct. Mol. Biol.* **13,** 957–964 (2006).
44. Randall, L. L. *et al.* Binding of SecB to ribosome-bound polypeptides has the same characteristics as binding to full-length, denatured proteins. *Proc. Natl Acad. Sci. USA* **94,** 802–807 (1997).
45. Lill, R., Dowhan, W. & Wickner, W. The ATPase activity of SecA is regulated by acidic phospholipids, SecY, and the leader and mature domains of precursor proteins. *Cell* **60,** 271–280 (1990).
46. Randall, L. L. & Hardy, S. J. SecB, one small chaperone in the complex milieu of the cell. *Cell. Mol. Life Sci.* **59,** 1617–1623 (2002).
47. Musial-Siwek, M., Rusch, S. L. & Kendall, D. A. Selective photoaffinity labeling identifies the signal peptide binding domain on SecA. *J. Mol. Biol.* **365,** 637–648 (2006).
48. Osborne, A. R. & Rapoport, T. A. Protein translocation is mediated by oligomers of the SecY complex with one SecY copy forming the channel. *Cell* **129,** 97–110 (2007).
49. Economou, A. & Wickner, W. SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. *Cell* **78,** 835–843 (1994).
50. Kim, Y. J, Rajapandi, T. & Oliver, D. SecA protein is exposed to the periplasmic surface of the *E. coli* inner membrane in its active state. *Cell* **78,** 845–853 (1994).
51. Schiebel, E., Driessen, A. J., Hartl, F. U. & Wickner, W. $\Delta\mu_H+$ and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell* **64,** 927–939 (1991).
52. Irihimovitch, V. & Eichler, J. Post-translational secretion of fusion proteins in the halophilic archaea *Haloferax volcanii*. *J. Biol. Chem.* **278,** 12881–12887 (2003).
53. Ortenberg, R. & Mevarech, M. Evidence for post-translational membrane insertion of the integral membrane protein bacterioopsin expressed in the heterologous halophilic archaeon *Haloferax volcanii*. *J. Biol. Chem.* **275,** 22839–22846 (2000).
54. Shaw, A. S., Rottier, P. J. & Rose, J. K. Evidence for the loop model of signal-sequence insertion into the endoplasmic reticulum. *Proc. Natl Acad. Sci. USA* **85,** 7592–7596 (1988).
55. Raden, D., Song, W. & Gilmore, R. Role of the cytoplasmic segments of Sec61α in the ribosome-binding and translocation-promoting activities of the Sec61 complex. *J. Cell Biol.* **150,** 53–64 (2000).
56. Plath, K., Mothes, W., Wilkinson, B. M., Stirling, C. J. & Rapoport, T. A. Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell* **94,** 795–807 (1998).
57. Harris, C. R. & Silhavy, T. J. Mapping an interface of SecY (PrlA) and SecE (PrlG) by using synthetic phenotypes and *in vivo* cross-linking. *J. Bacteriol.* **181,** 3438–3444 (1999).
58. Tam, P. C., Maillard, A. P., Chan, K. K. & Duong, F. Investigating the SecY plug movement at the SecYEG translocation channel. *EMBO J.* **24,** 3380–3388 (2005).
59. Smith, M. A., Clemons, W. M. Jr, DeMars, C. J. & Flower, A. M. Modeling the effects of prl mutations on the *Escherichia coli* SecY complex. *J. Bacteriol.* **187,** 6454–6465 (2005).

©2007 Nature Publishing Group

60. Weihofen, A., Binns, K., Lemberg, M. K., Ashman, K. & Martoglio, B. Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science* **296,** 2215–2218 (2002).
61. Cannon, K. S., Or, E., Clemons, W. M. Jr, Shibata, Y. & Rapoport, T. A. Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY. *J. Cell Biol.* **169,** 219–225 (2005).
62. Tani, K., Tokuda, H. & Mizushima, S. Translocation of proOmpA possessing an intramolecular disulfide bridge into membrane vesicles of *Escherichia coli*. Effect of membrane energization. *J. Biol. Chem.* **265,** 17341–17347 (1990).
63. Kurzchalia, T. V. *et al.* tRNA-mediated labelling of proteins with biotin. A nonradioactive method for the detection of cell-free translation products. *Eur. J. Biochem.* **172,** 663–668 (1988).
64. Gumbart, J. & Schulten, K. Molecular dynamics studies of the archaeal translocon. *Biophys. J.* **90,** 2356–2367 (2006).
65. Tian, P. & Andricioaei, I. Size, motion, and function of the SecY translocon revealed by molecular dynamics simulations with virtual probes. *Biophys. J.* **90,** 2718–2730 (2006).
66. Haider, S., Hall, B. A. & Sansom, M. S. Simulations of a protein translocation pore: SecY. *Biochemistry* **45,** 13018–13024 (2006).
67. Saparov, S. M. *et al.* Determining the conductance of the SecY protein translocation channel for small molecules. *Mol. Cell* **26,** 501–509 (2007).
68. Kowarik, M., Kung, S., Martoglio, B. & Helenius, A. Protein folding during cotranslational translocation in the endoplasmic reticulum. *Mol. Cell* **10,** 769–778 (2002).
69. Hamman, B. D., Chen, J. C., Johnson, E. E. & Johnson, A. E. The aqueous pore through the translocon has a diameter of 40–60 Å during cotranslational protein translocation at the ER membrane. *Cell* **89,** 535–544 (1997).
70. Kaufmann, A., Manting, E. H., Veenendaal, A. K., Driessen, A. J. & van der Does, C. Cysteine-directed cross-linking demonstrates that helix 3 of SecE is close to helix 2 of SecY and helix 3 of a neighboring SecE. *Biochemistry* **38,** 9115–9125 (1999).
71. Mitra, K. *et al.* Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature* **438,** 318–324 (2005).
72. Ménétret, J. F. *et al.* Ribosome binding of a single copy of the SecY complex: implications for protein translocation. *Mol. Cell* (in the press).
73. Voss, N. R., Gerstein, M., Steitz, T. A. & Moore, P. B. The geometry of the ribosomal polypeptide exit tunnel. *J. Mol. Biol.* **360,** 893–906 (2006).
74. Halic, M. *et al.* Following the signal sequence from ribosomal tunnel exit to signal recognition particle. *Nature* **444,** 507–511 (2006).
75. Mitra, K. *et al.* Elongation arrest by SecM via a cascade of ribosomal RNA rearrangements. *Mol. Cell* **22,** 533–543 (2006).
76. Johnson, A. E. Maintaining the permeability barrier during protein trafficking at the endoplasmic reticulum membrane. *Biochem. Soc. Trans.* **31,** 1227–1231 (2003).
77. Halic, M. *et al.* Signal recognition particle receptor exposes the ribosomal translocon binding site. *Science* **312,** 745–747 (2006).
78. Schaletzky, J. & Rapoport, T. A. Ribosome binding to and dissociation from translocation sites of the endoplasmic reticulum membrane. *Mol. Biol. Cell* **17,** 3860–3869 (2006).
79. Hanein, D. *et al.* Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. *Cell* **87,** 721–732 (1996).
80. Beckmann, R. *et al.* Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* **107,** 361–372 (2001).
81. Ménétret, J. F. *et al.* Architecture of the ribosome-channel complex derived from native membranes. *J. Mol. Biol.* **348,** 445–457 (2005).
82. Thanassi, D. G., Stathopoulos, C., Karkal, A. & Li, H. Protein secretion in the absence of ATP: the autotransporter, two-partner secretion and chaperone/usher pathways of Gram-negative bacteria. *Mol. Membr. Biol.* **22,** 63–72 (2005).
83. Ahting, U. *et al.* The TOM core complex: the general protein import pore of the outer membrane of mitochondria. *J. Cell Biol.* **147,** 959–968 (1999).
84. Rehling, P. *et al.* Protein insertion into the mitochondrial inner membrane by a twin-pore translocase. *Science* **299,** 1747–1751 (2003).
85. Heinrich, S. U., Mothes, W., Brunner, J. & Rapoport, T. A. The Sec61p complex mediates the integration of a membrane protein by allowing lipid partitioning of the transmembrane domain. *Cell* **102,** 233–244 (2000).
86. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433,** 377–381 (2005).
87. Gumbart, J. & Schulten, K. Structural determinants of lateral gate opening in the protein translocon. *Biochemistry* **46,** 11147–11157 (2007).
88. Wessels, H. P. & Spiess, M. Insertion of a multispanning membrane protein occurs sequentially and requires only one signal sequence. *Cell* **55,** 61–70 (1988).
89. Rapoport, T. A., Goder, V., Heinrich, S. U. & Matlack, K. E. Membrane-protein integration and the role of the translocation channel. *Trends Cell Biol.* **14,** 568–575 (2004).
90. Le Gall, S., Neuhof, A. & Rapoport, T. The endoplasmic reticulum membrane is permeable to small molecules. *Mol. Biol. Cell* **15,** 447–455 (2004).
91. Hamman, B. D., Hendershot, L. M. & Johnson, A. E. BiP maintains the permeability barrier of the ER membrane by sealing the lumenal end of the translocon pore before and early in translocation. *Cell* **92,** 747–758 (1998).
92. Liao, S., Lin, J., Do, H. & Johnson, A. E. Both lumenal and cytosolic gating of the aqueous ER translocon pore are regulated from inside the ribosome during membrane protein integration. *Cell* **90,** 31–41 (1997).
93. Woolhead, C. A., McCormick, P. J. & Johnson, A. E. Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins. *Cell* **116,** 725–736 (2004).
94. Lu, J. & Deutsch, C. Secondary structure formation of a transmembrane segment in Kv channels. *Biochemistry* **44,** 8230–8243 (2005).
95. Ménétret, J. F. *et al.* The structure of ribosome–channel complexes engaged in protein translocation. *Mol. Cell* **6,** 1219–1232 (2000).
96. Schiebel, E. & Wickner, W. Preprotein translocation creates a halide anion permeability in the *Escherichia coli* plasma membrane. *J. Biol. Chem.* **267,** 7505–7510 (1992).
97. Junne, T., Schwede, T., Goder, V. & Spiess, M. The plug domain of yeast Sec61p is important for efficient protein translocation, but is not essential for cell viability. *Mol. Biol. Cell* **17,** 4063–4068 (2006).
98. Maillard, A. P., Lalani, S., Silva, F., Belin, D. & Duong, F. Deregulation of the SecYEG translocation channel upon removal of the plug domain. *J. Biol. Chem.* **282,** 1281–1287 (2007).
99. Li, W. *et al.* The plug domain of the SecY protein stabilizes the closed state of the translocation channel and maintains a membrane seal. *Mol. Cell* **26,** 511–521 (2007).

Acknowledgements I thank A. Osborne, Y. Shibata, B. van den Berg and K. Matlack for critical reading of the manuscript, and W. Li and V. Goder for help with the figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T.A.R. (tom_rapoport@hms.harvard.edu).

©2007 Nature Publishing Group

Ediacaran oxidation and biotic evolution

Arising from: D. A. Fike *et al. Nature* **444,** 744–747 (2006).

The link between the radiation of various lineages of eukaryotes in the latest Proterozoic and massive environmental changes—oxygenation, global ice ages and bolide impact—is the focus of much research interest. Fike *et al.*[1] use carbon and sulphur isotope–chemostratigraphic data from Oman to propose three stages of oxidation in the Ediacaran oceans, and link the second and third stages to eukaryote diversification. The second stage, signalled by strongly ^{13}C-depleted sedimentary carbonates (the 'Shuram excursion'), is believed to result from oxidation of a large, deep-ocean reservoir of organic carbon[1]. Fike *et al.* use our data[2,3] to assert that a correlative carbon isotope excursion in Australia coincided with the initial diversification of acanthomorphic acritarchs. Peak diversity is claimed to have coincided with subsequent deposition of ^{13}C-enriched carbonate and the third oxidation stage. However, the authors seem to have misinterpreted our data, which instead indicate that diversification significantly preceded the Shuram excursion; this weakens their argument for a link between the inferred oxidation events and eukaryote evolution.

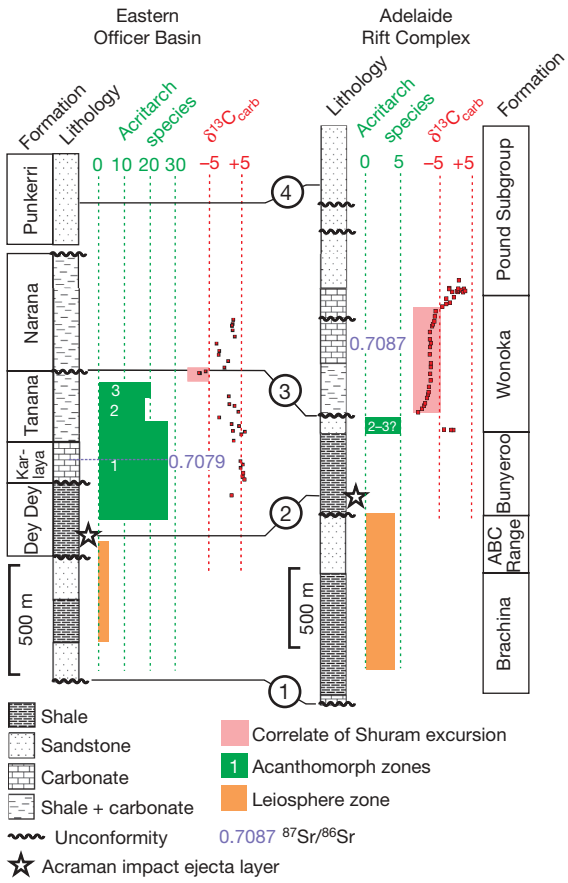


Figure 1 | Ediacaran successions in the eastern Officer Basin (data from Munta-1 drillhole[2,5]) and Adelaide Rift Complex (acritarch data from SCYW1A drillhole[2]; isotope data from Bunyeroo Gorge[3]). Selected correlation tie-lines (circled numbers) are (1) the basal Ediacaran sequence boundary; (2) the Acraman impact ejecta horizon[2,5]; (3) the 'canyon unconformity'[5]; and (4) occurrences of Ediacaran animal fossils. Marine ^{87}Sr/^{86}Sr increased monotonically through the Ediacaran; indicated least-altered values are thus consistent with the correlations shown[5].

In the type area of the Ediacaran system, the Adelaide Rift Complex, no acritarchs have been recovered from the strongly ^{13}C-depleted section (<−5‰, middle Wonoka Formation, Fig. 1) correlated with the Shuram excursion. This is probably due to high levels of thermal maturity and deep weathering. Nearby drillholes contain sparse acanthomorphs, indicating diversification in the underlying Bunyeroo and lower-most Wonoka Formations (Fig. 1), but the remainder of the Wonoka Formation is missing in these drillholes. The acritarch record is much better preserved and more complete in the Officer Basin. Here, acanthomorph diversification begins in the Dey Dey Mudstone (a correlative of the Bunyeroo Formation), within 50 m of the Acraman impact ejecta layer[4], and diverse assemblages continue through a thick overlying succession of mudstone and siltstone with interbedded ^{13}C-enriched carbonate (Karlaya Limestone and Tanana Formation) to reach peak diversity in the ^{13}C-enriched Wilari Dolomite Member[2,5]. The increase in acanthomorph species numbers does not coincide with a negative excursion in ^{13}C (Fig. 1). Not until some 700 m above the first acanthomorph record, and above at least three of our four acanthomorph zones[2], does strongly ^{13}C-depleted carbonate, presumably the onset of the Shuram excursion, appear in the Officer Basin[2,5].

Consistent with this, the Karlaya Limestone and Tanana Formation are older than the ^{13}C-depleted middle Wonoka Formation, from strontium (Sr) isotope stratigraphic and sequence stratigraphic evidence[5]. Acanthomorph abundances and species numbers apparently decline above this level, in contrast to the model of Fike *et al.*, although the fossil record is poor in this part of the Australian succession. Because of unconformities in the Officer Basin and non-preservation of acritarchs in the Adelaide Rift Complex, we are uncertain of the relationship of the peak diversity seen in the Wilari Dolomite Member with the Shuram excursion.

In contrast to the claim by Fike *et al.*, our data do not show peak acanthomorph diversity in the Julie Formation (Amadeus Basin). Not a single acritarch was recovered from the Julie Formation. In the Amadeus Basin, acanthomorph diversification begins in the underlying Pertatataka Formation (a correlative of the Bunyeroo Formation), with maximum recorded diversity about 50 m below the base of the Julie Formation. The Australian evidence for peak acanthomorph diversity coinciding with the third oxidation stage of Fike *et al.* is therefore unconvincing.

Kathleen Grey[1] **& Clive R. Calver**[2]
[1]Geological Survey of Western Australia, Department of Industry and Resources, East Perth, Western Australia 6004, Australia.
[2]Mineral Resources Tasmania, PO Box 56, Rosny Park, Tasmania 7018, Australia.
e-mail: ccalver@mrt.tas.gov.au

Received 5 April; accepted 29 August 2007.

1. Fike, D. A., Grotzinger, J. P., Pratt, L. M. & Summons, R. E. Oxidation of the Ediacaran Ocean. *Nature* **444,** 744–747 (2006).
2. Grey, K. in *Ediacaran Palynology of Australia.* Mem. 31 (ed. Laurie, J. R.) 1–439 (Association of Australasian Palaeontologists, Canberra, 2005).
3. Calver, C. R. Isotope stratigraphy of the Ediacarian (Neoproterozoic III) of the Adelaide Rift Complex, South Australia, and the overprint of water column stratification. *Precambr. Res.* **100,** 121–150 (2000).
4. Willman, S., Moczydłowska, M. & Grey, K. Neoproterozoic (Ediacaran) diversification of acritarchs—a new record from the Murnaroo 1 drillcore, eastern Officer Basin, Australia. *Rev. Palaeobot. Palynol.* **139,** 17–39 (2006).
5. Calver, C. R. & Lindsay, J. F. Ediacarian sequence and isotope stratigraphy of the Officer Basin, South Australia. *Aust. J. Earth Sci.* **45,** 513–532 (1998).

doi:10.1038/nature06360

©2007 Nature Publishing Group

Fike et al. *reply*

Replying to: K. Grey & C. R. Calver *Nature* **450,** doi:10.1038/nature06360 (2007).

Calver and Grey[1] point out the difficulties of relating the Australian acritarch record to the global record of environmental change. This results from the lack of documented stratigraphic sections that have overlapping chronologies of both acritarch evolution and global chemostratigraphic proxies, such as $\delta^{13}C_{carb}$, and we agree that coupled records are required to confirm our hypothesis[2]. Until such records exist, any correlations between Australian acritarch-bearing strata and strata containing records of global secular variations in $\delta^{13}C_{carb}$ are necessarily tenuous. This uncertainty, however, does not affect the principal conclusion of our paper, which emphasizes the evidence for sequential chemical oxidation of the Ediacaran ocean. Our inferred correlation between chemical oxidation and biological evolution is supported by subsequent results on strata containing strong geochronologic tie-points and a robust biostratigraphic and chemostratigraphic record[3,4].

More specifically, Calver and Grey contest our interpretation of an increase in acritarch diversity associated with the onset of the negative excursion in $\delta^{13}C_{carb}$ in the Wonoka Formation of the Adelaide Rift Complex, equivalent to our stage II oxidation. Despite this claim, existing data show that the onset of the Wonoka $\delta^{13}C_{carb}$ excursion is found in the most basal Wonoka, just above the Bunyeroo/Wonoka contact[5]. An examination of the Adelaide Rift Complex (figure 18 of ref. 6) acritarch record reveals an increase in acritarch diversity (appearance of the *Appendisphaera barbata* (*Ab*)/*Alicesphaeridium medusoidum* (*Am*)/*Gyalosphaeridium pulchrum* (Gp) assemblage) in the basal Wonoka. We therefore stand by our correlation between acritarch diversification in the basal Wonoka and the onset of the stage II oxidation.

Calver and Grey correctly point out our error in associating the Julie Formation with the peak in acritarch diversity in the Amadeus basin. The peak diversity is in the uppermost Pertatataka Formation, immediately underlying the Julie Formation. However, in the absence of geochronological constraints, we accept the interpretation that these formations are not broadly separated in time[7]. Therefore, this does not discount the central hypothesis of our paper, which argues for increasing biological diversity associated with progressive oxidation throughout Ediacaran time.

D. A. Fike[1], J. P. Grotzinger[2], L. L. Pratt[3] & R. R. Summons[1]
[1]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mail: dfike@mit.edu
[2]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.
[3]Department of Geological Sciences, Indiana University, Bloomington, Indiana 47405, USA.

1. Grey, K. & Calver, C. R. Ediacaran oxidation and biotic evolution. *Nature* doi:10.1038/nature06360 (2007).
2. Fike, D. A., Grotzinger, J. P., Pratt, L. M. & Summons, R. E. Oxidation of the Ediacaran Ocean. *Nature* **444,** 744–747 (2006).
3. Canfield, D. E., Poulton, S. W. & Narbonne, G. M. Late neoproterozoic deep ocean oxygenation and the rise of animal life. *Science* **315,** 92–95 (2007).
4. Yin, L. *et al.* Doushantuo embryos preserved inside diapause egg cysts. *Nature* **446,** 661–663 (2007).
5. Calver, C. R. Isotope stratigraphy of the Ediacarian (Neoproterozoic III) of the Adelaide Rift Complex, Australia, and the overprint of water column stratification. *Precambr. Res.* **100,** 121–150 (2000).
6. Grey, K. *Ediacaran palynology of Australia* Mem. 31 (ed. Laurie, J. R.) 52 (Association of Australasian Paleontologists, Canberra, 2005).
7. Walter, M. R., Veevers, J. J., Calver, C. R., Gorjan, P. & Hill, A. C. Dating the 840–544 Ma Neoproterozoic interval by isotopes of strontium, carbon, and sulfur in seawater, and some interpretative models. *Precambr. Res.* **100,** 371–433 (2000).

doi:10.1038/nature06361

©2007 Nature Publishing Group

Ediacaran oxidation and biotic evolution

Arising from: D. A. Fike *et al. Nature* **444,** 744–747 (2006).

The link between the radiation of various lineages of eukaryotes in the latest Proterozoic and massive environmental changes—oxygenation, global ice ages and bolide impact—is the focus of much research interest. Fike *et al.*[1] use carbon and sulphur isotope–chemostratigraphic data from Oman to propose three stages of oxidation in the Ediacaran oceans, and link the second and third stages to eukaryote diversification. The second stage, signalled by strongly ^{13}C-depleted sedimentary carbonates (the 'Shuram excursion'), is believed to result from oxidation of a large, deep-ocean reservoir of organic carbon[1]. Fike *et al.* use our data[2,3] to assert that a correlative carbon isotope excursion in Australia coincided with the initial diversification of acanthomorphic acritarchs. Peak diversity is claimed to have coincided with subsequent deposition of ^{13}C-enriched carbonate and the third oxidation stage. However, the authors seem to have misinterpreted our data, which instead indicate that diversification significantly preceded the Shuram excursion; this weakens their argument for a link between the inferred oxidation events and eukaryote evolution.

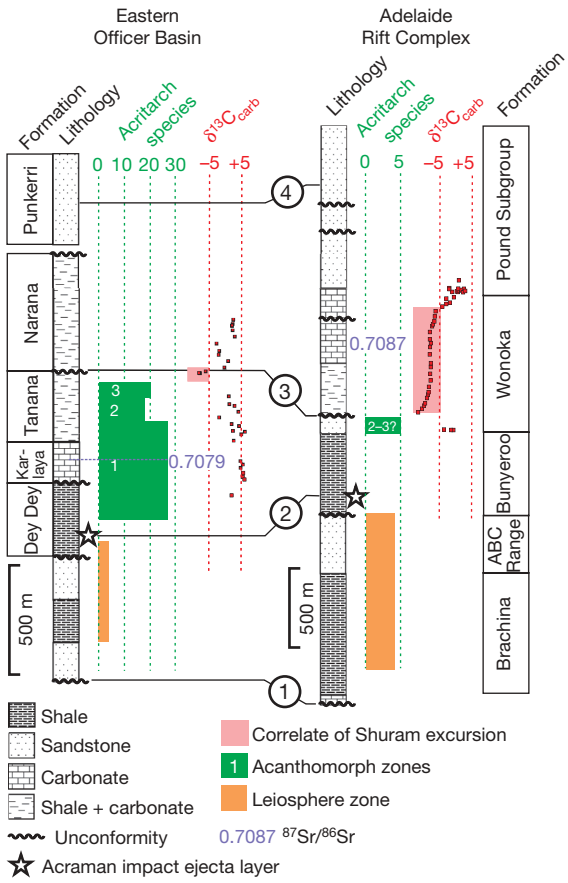


Figure 1 | Ediacaran successions in the eastern Officer Basin (data from Munta-1 drillhole[2,5]) and Adelaide Rift Complex (acritarch data from SCYW1A drillhole[2]; isotope data from Bunyeroo Gorge[3]). Selected correlation tie-lines (circled numbers) are (1) the basal Ediacaran sequence boundary; (2) the Acraman impact ejecta horizon[2,5]; (3) the 'canyon unconformity'[5]; and (4) occurrences of Ediacaran animal fossils. Marine ^{87}Sr/^{86}Sr increased monotonically through the Ediacaran; indicated least-altered values are thus consistent with the correlations shown[5].

In the type area of the Ediacaran system, the Adelaide Rift Complex, no acritarchs have been recovered from the strongly ^{13}C-depleted section (<−5‰, middle Wonoka Formation, Fig. 1) correlated with the Shuram excursion. This is probably due to high levels of thermal maturity and deep weathering. Nearby drillholes contain sparse acanthomorphs, indicating diversification in the underlying Bunyeroo and lower-most Wonoka Formations (Fig. 1), but the remainder of the Wonoka Formation is missing in these drillholes. The acritarch record is much better preserved and more complete in the Officer Basin. Here, acanthomorph diversification begins in the Dey Dey Mudstone (a correlative of the Bunyeroo Formation), within 50 m of the Acraman impact ejecta layer[4], and diverse assemblages continue through a thick overlying succession of mudstone and siltstone with interbedded ^{13}C-enriched carbonate (Karlaya Limestone and Tanana Formation) to reach peak diversity in the ^{13}C-enriched Wilari Dolomite Member[2,5]. The increase in acanthomorph species numbers does not coincide with a negative excursion in ^{13}C (Fig. 1). Not until some 700 m above the first acanthomorph record, and above at least three of our four acanthomorph zones[2], does strongly ^{13}C-depleted carbonate, presumably the onset of the Shuram excursion, appear in the Officer Basin[2,5].

Consistent with this, the Karlaya Limestone and Tanana Formation are older than the ^{13}C-depleted middle Wonoka Formation, from strontium (Sr) isotope stratigraphic and sequence stratigraphic evidence[5]. Acanthomorph abundances and species numbers apparently decline above this level, in contrast to the model of Fike *et al.*, although the fossil record is poor in this part of the Australian succession. Because of unconformities in the Officer Basin and non-preservation of acritarchs in the Adelaide Rift Complex, we are uncertain of the relationship of the peak diversity seen in the Wilari Dolomite Member with the Shuram excursion.

In contrast to the claim by Fike *et al.*, our data do not show peak acanthomorph diversity in the Julie Formation (Amadeus Basin). Not a single acritarch was recovered from the Julie Formation. In the Amadeus Basin, acanthomorph diversification begins in the underlying Pertatataka Formation (a correlative of the Bunyeroo Formation), with maximum recorded diversity about 50 m below the base of the Julie Formation. The Australian evidence for peak acanthomorph diversity coinciding with the third oxidation stage of Fike *et al.* is therefore unconvincing.

Kathleen Grey[1] **& Clive R. Calver**[2]
[1]Geological Survey of Western Australia, Department of Industry and Resources, East Perth, Western Australia 6004, Australia.
[2]Mineral Resources Tasmania, PO Box 56, Rosny Park, Tasmania 7018, Australia.
e-mail: ccalver@mrt.tas.gov.au

Received 5 April; accepted 29 August 2007.

1. Fike, D. A., Grotzinger, J. P., Pratt, L. M. & Summons, R. E. Oxidation of the Ediacaran Ocean. *Nature* **444,** 744–747 (2006).
2. Grey, K. in *Ediacaran Palynology of Australia.* Mem. 31 (ed. Laurie, J. R.) 1–439 (Association of Australasian Palaeontologists, Canberra, 2005).
3. Calver, C. R. Isotope stratigraphy of the Ediacarian (Neoproterozoic III) of the Adelaide Rift Complex, South Australia, and the overprint of water column stratification. *Precambr. Res.* **100,** 121–150 (2000).
4. Willman, S., Moczydłowska, M. & Grey, K. Neoproterozoic (Ediacaran) diversification of acritarchs—a new record from the Murnaroo 1 drillcore, eastern Officer Basin, Australia. *Rev. Palaeobot. Palynol.* **139,** 17–39 (2006).
5. Calver, C. R. & Lindsay, J. F. Ediacarian sequence and isotope stratigraphy of the Officer Basin, South Australia. *Aust. J. Earth Sci.* **45,** 513–532 (1998).

doi:10.1038/nature06360

©2007 Nature Publishing Group

Fike et al. *reply*

Replying to: K. Grey & C. R. Calver *Nature* **450,** doi:10.1038/nature06360 (2007).

Calver and Grey[1] point out the difficulties of relating the Australian acritarch record to the global record of environmental change. This results from the lack of documented stratigraphic sections that have overlapping chronologies of both acritarch evolution and global chemostratigraphic proxies, such as $\delta^{13}C_{carb}$, and we agree that coupled records are required to confirm our hypothesis[2]. Until such records exist, any correlations between Australian acritarch-bearing strata and strata containing records of global secular variations in $\delta^{13}C_{carb}$ are necessarily tenuous. This uncertainty, however, does not affect the principal conclusion of our paper, which emphasizes the evidence for sequential chemical oxidation of the Ediacaran ocean. Our inferred correlation between chemical oxidation and biological evolution is supported by subsequent results on strata containing strong geochronologic tie-points and a robust biostratigraphic and chemostratigraphic record[3,4].

More specifically, Calver and Grey contest our interpretation of an increase in acritarch diversity associated with the onset of the negative excursion in $\delta^{13}C_{carb}$ in the Wonoka Formation of the Adelaide Rift Complex, equivalent to our stage II oxidation. Despite this claim, existing data show that the onset of the Wonoka $\delta^{13}C_{carb}$ excursion is found in the most basal Wonoka, just above the Bunyeroo/Wonoka contact[5]. An examination of the Adelaide Rift Complex (figure 18 of ref. 6) acritarch record reveals an increase in acritarch diversity (appearance of the *Appendisphaera barbata* (*Ab*)/*Alicesphaeridium medusoidum* (*Am*)/*Gyalosphaeridium pulchrum* (Gp) assemblage) in the basal Wonoka. We therefore stand by our correlation between acritarch diversification in the basal Wonoka and the onset of the stage II oxidation.

Calver and Grey correctly point out our error in associating the Julie Formation with the peak in acritarch diversity in the Amadeus basin. The peak diversity is in the uppermost Pertatataka Formation, immediately underlying the Julie Formation. However, in the absence of geochronological constraints, we accept the interpretation that these formations are not broadly separated in time[7]. Therefore, this does not discount the central hypothesis of our paper, which argues for increasing biological diversity associated with progressive oxidation throughout Ediacaran time.

D. A. Fike[1], J. P. Grotzinger[2], L. L. Pratt[3] & R. R. Summons[1]
[1]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.
e-mail: dfike@mit.edu
[2]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.
[3]Department of Geological Sciences, Indiana University, Bloomington, Indiana 47405, USA.

1. Grey, K. & Calver, C. R. Ediacaran oxidation and biotic evolution. *Nature* doi:10.1038/nature06360 (2007).
2. Fike, D. A., Grotzinger, J. P., Pratt, L. M. & Summons, R. E. Oxidation of the Ediacaran Ocean. *Nature* **444,** 744–747 (2006).
3. Canfield, D. E., Poulton, S. W. & Narbonne, G. M. Late neoproterozoic deep ocean oxygenation and the rise of animal life. *Science* **315,** 92–95 (2007).
4. Yin, L. *et al.* Doushantuo embryos preserved inside diapause egg cysts. *Nature* **446,** 661–663 (2007).
5. Calver, C. R. Isotope stratigraphy of the Ediacarian (Neoproterozoic III) of the Adelaide Rift Complex, Australia, and the overprint of water column stratification. *Precambr. Res.* **100,** 121–150 (2000).
6. Grey, K. *Ediacaran palynology of Australia* Mem. 31 (ed. Laurie, J. R.) 52 (Association of Australasian Paleontologists, Canberra, 2005).
7. Walter, M. R., Veevers, J. J., Calver, C. R., Gorjan, P. & Hill, A. C. Dating the 840–544 Ma Neoproterozoic interval by isotopes of strontium, carbon, and sulfur in seawater, and some interpretative models. *Precambr. Res.* **100,** 371–433 (2000).

doi:10.1038/nature06361

©2007 Nature Publishing Group

ARTICLES

Shiga toxin induces tubular membrane invaginations for its uptake into cells

Winfried Römer[1,5], Ludwig Berland[2,6], Valérie Chambon[1,5], Katharina Gaus[8,9], Barbara Windschiegl[10], Danièle Tenza[3,5], Mohamed R. E. Aly[4,7], Vincent Fraisier[5], Jean-Claude Florent[4,7], David Perrais[11], Christophe Lamaze[1,5], Graça Raposo[3,5], Claudia Steinem[10], Pierre Sens[12], Patricia Bassereau[2,6] & Ludger Johannes[1,5]

Clathrin seems to be dispensable for some endocytic processes and, in several instances, no cytosolic coat protein complexes could be detected at sites of membrane invagination. Hence, new principles must in these cases be invoked to account for the mechanical force driving membrane shape changes. Here we show that the Gb3 (glycolipid)-binding B-subunit of bacterial Shiga toxin induces narrow tubular membrane invaginations in human and mouse cells and model membranes. In cells, tubule occurrence increases on energy depletion and inhibition of dynamin or actin functions. Our data thus demonstrate that active cellular processes are needed for tubule scission rather than tubule formation. We conclude that the B-subunit induces lipid reorganization that favours negative membrane curvature, which drives the formation of inward membrane tubules. Our findings support a model in which the lateral growth of B-subunit–Gb3 microdomains is limited by the invagination process, which itself is regulated by membrane tension. The physical principles underlying this basic cargo-induced membrane uptake may also be relevant to other internalization processes, creating a rationale for conceptualizing the perplexing diversity of endocytic routes.

Several endocytic routes do not involve clathrin-coated pits[1–6], and the mechanisms that initiate membrane uptake in these cases are still largely unknown. The bacterial Shiga toxin is one of the cargoes that are endocytosed independently of clathrin[7–9]. Intoxication of cells by this pathogenic product requires its binding to the glycosphingolipid receptor globotriaosyl ceramide (Gb3) in the exoplasmic membrane leaflet. To study one of the earliest steps of toxin uptake into cells, that is, the formation of membrane invaginations, we have applied various microscopy approaches to monitor dynamic membrane shape changes.

STxB uptake via tubular membrane invaginations

First, we analysed the localization in human HeLa cells of the non-toxic homopentameric Gb3-binding B-subunit of Shiga toxin (STxB). At early times of STxB internalization (5 min), tubular STxB-positive structures of variable length could be detected (Fig. 1a), of which some were clearly connected to the surface in the focal plane of observation (arrow). These tubules did not contain the transferrin receptor (TfR), which is internalized via clathrin-coated pits.

Dynamin has been ascribed a critical role in membrane scission in many, but not all, endocytic pathways[2,10]. Inhibiting dynamin function by overexpression of the dominant-negative dynamin protein Dyn2baK44A (Fig. 1b), or with dynasore—a small-molecule dynamin inhibitor[11], or by short interfering (si)RNA-mediated depletion of the protein (Supplementary Fig. 1) resulted in all cases in the increased occurrence of long surface-connected STxB-containing tubules, suggesting that their scission was inhibited under these conditions. Dyn2baK44A (Fig. 1b), but not TfR (Supplementary Fig. 2), localized to STxB-containing tubules. To bypass fixation or shifts in temperature—conditions that may affect tubule integrity—images were acquired on live cells at 37 °C. Again, STxB appeared in surface-connected tubules when dynamin function was inhibited (Fig. 1c), even at time points when STxB was almost exclusively in perinuclear Golgi membranes of control cells (Supplementary Fig. 3). The finding that dynasore partially protected cells from Shiga-toxin-induced protein biosynthesis inhibition (Supplementary Fig. 4) confirmed that a sizable fraction of toxin traffics through dynamin-dependent pathways, differentiating this uptake from the CLIC/GEEC[12,13] and flotillin[14] routes. Recently, strong evidence for a functional link between dynamin and actin in membrane scission has been presented[15]. Incubation of cells with latrunculin A, a small-molecule compound that inhibits actin polymerization, increased the occurrence of STxB-induced invaginations (Supplementary Fig. 5), suggesting that actin is required for their efficient scission.

Incubation of energy-depleted cells, in which molecular motors and other mechanoenzymes, including dynamin, are inactive, with STxB concentrations as low as 1 nM also led to the formation of tubules that were negative for TfR and clathrin (Supplementary Fig. 6a–c). Tubules were more frequently observed in energy-depleted cells than in control cells, probably because dynamin was unable to process them, despite being localized on them (Fig. 1d). Immunogold labelling on cryosections of energy-depleted cells clearly confirmed that invaginations were surface-connected (Fig. 1e, and Supplementary Fig. 7). Furthermore, the dynamic growth of surface-connected tubules could be imaged on energy-depleted live cells (Fig. 1f, Supplementary Fig. 6d, e; and Supplementary Movies 1–3). Cholesterol extraction, which has been described to affect cell entry of STxB[16], also resulted in the accumulation of surface-connected tubules (Fig. 1g), the length of which was

[1]Institut Curie, Centre de Recherche, Laboratoire Trafic, Signalisation et Ciblage Intracellulaires, [2]Laboratoire Physico-Chimie, [3]Laboratoire Structure et Compartiments Membranaires, [4]Laboratoire Chimie Organique (Vectorisation), 26 rue d'Ulm, 75248 Paris Cedex 05, France. [5]CNRS UMR144, [6]Université P. et M. Curie/CNRS UMR168, [7]CNRS UMR176, [8]Centre for Vascular Research, University of New South Wales, 2052 Sydney, Australia. [9]Department of Haematology, Prince of Wales Hospital, 2031 Sydney, Australia. [10]Institut für Organische und Biomolekulare Chemie, Georg-August Universität, Tammannstr. 2, 37077 Göttingen, Germany. [11]Laboratoire de Physiologie Cellulaire de la Synapse, CNRS UMR 5091 et Université Bordeaux 2, Institut François Magendie, 33077 Bordeaux, France. [12]UMR Gulliver CNRS-ESPCI 7083, 10 rue Vauquelin, 75231 Paris Cedex 05, France.

©2007 Nature Publishing Group

increased when cells were additionally energy-depleted (Fig. 1h). How cholesterol functions in membrane scission remains to be determined, but resistance to detergent extraction indicated that tubular STxB was in a raft-like lipid environment (Supplementary Fig. 8).

In summary, in different experimental conditions that all affect STxB entry into cells, STxB was localized in surface-connected tubules, strongly suggesting that these are intermediates of STxB uptake. This was directly demonstrated with evanescent field fluorescence microscopy: the reversion of energy depletion by washout led to the scission of STxB-containing tubules (Fig. 1i, and Supplementary Movie 4).

STxB induces tubular membrane invaginations

The presence of tubules in energy-depleted cells points to the possibility that STxB itself induces these structures. The following experiment was performed to test this hypothesis. Incubation of energy-depleted cells with the live-cell membrane dye FM4-64 at 37 °C in the absence of STxB (Supplementary Fig. 9) or at 4 °C in

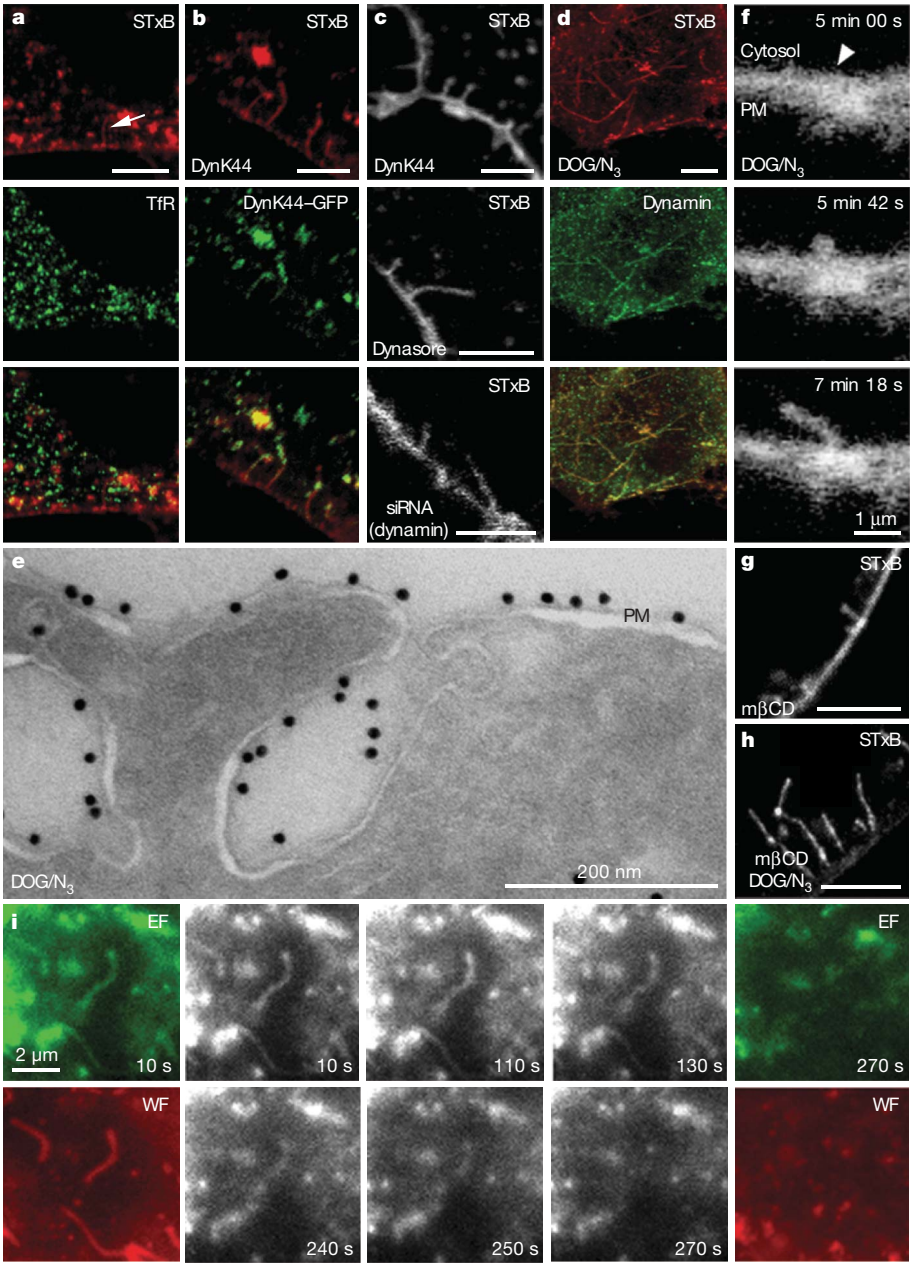


Figure 1 | STxB is found on endocytic membrane invaginations. **a**, Cy3-coupled STxB (20 nM, red) was bound to HeLa cells at 4 °C, a temperature at which no endocytosis occurs. The cells were then shifted to 37 °C for 5 min, fixed and stained for TfR (green). **b**, Incubation as in **a**, with HeLa cells that were transfected for 16 h with Dyn2baK44A. GFP-tagged mutant dynamin protein (green) is localized on STxB-containing tubules (red). **c**, Cy3-STxB was imaged after incubation on live HeLa cells in the indicated conditions (16 h Dyn2baK44A expression, 80 μM dynasore, or 3 days dynamin siRNA transfection). **d**, Endogenous dynamin can be detected on STxB-containing tubules in energy-depleted HeLa cells treated with deoxyglucose (DOG) and azide (N_3). **e**, STxB incubation (10 min) with energy-depleted HeLa cells, followed by immunogold labelling for STxB (15 nm gold particles) on cryosections. **f**, Live-cell analysis of tube formation after STxB injection at 37 °C onto energy-depleted cells. After STxB binding to the plasma membrane (PM), tube growth is followed over time. **g**, Cells were cholesterol extracted with methyl-β-cyclodextrin (mβCD) and incubated with Cy3-STxB. **h**, Incubation as in **g**, with cells that were also energy-depleted. **i**, Tubule fate was recorded by evanescent field (EF) microscopy during reversion of energy depletion by washout. In the left and right columns, evanescent-field and wide-field (WF) images are shown for initial (10 s) and late (270 s) time points of observation. Black and white images show evanescent field acquisitions between these time points. **a**–**i**, Scale bars, 5 μm, except if indicated otherwise.

©2007 Nature Publishing Group

its presence (STxB binding, Fig. 2a) revealed that no membrane invaginations could be detected under these conditions. Only when STxB was present during incubations at 37 °C could FM4-64 be found in invaginations (Fig. 2b). We conclude that STxB did not diffuse into preexisting invaginations, but induced them.

The specificity of the tubulation process was tested in two ways. First, we generated an STxB variant in which the Gb3 binding site III was inactivated by mutagenesis. The binding capacity of the W34A mutant is only reduced fourfold with an unchanged K_d, whereas a two-log reduction in cytotoxicity has been described in Vero cells[17]. Even when used at high concentrations that ensured efficient binding to cells, the W34A mutant was only rarely found in tubules (Fig. 2c). Second, energy-depleted cells were incubated with a monoclonal IgM antibody against the cellular Shiga toxin receptor, Gb3. The antibody bound to cells, but even after cross-linking with secondary antibodies, no membrane invaginations could be detected (Fig. 2d). These two observations demonstrated that binding to cells alone is not sufficient to induce membrane deformation.

Caveolins are peripheral membrane proteins that have been implicated in clathrin-independent endocytosis, including that of cholera toxin[18]. We found here that caveolin 1 was absent from STxB-induced invaginations (Fig. 2e), and that these could form on two caveolin-free cell systems: energy-depleted human huh7 cells (Fig. 2f) and fibroblasts from caveolin 1 knockout mice (Fig. 2g). These results established that caveolin 1 was not required for the formation of toxin-induced invaginations.

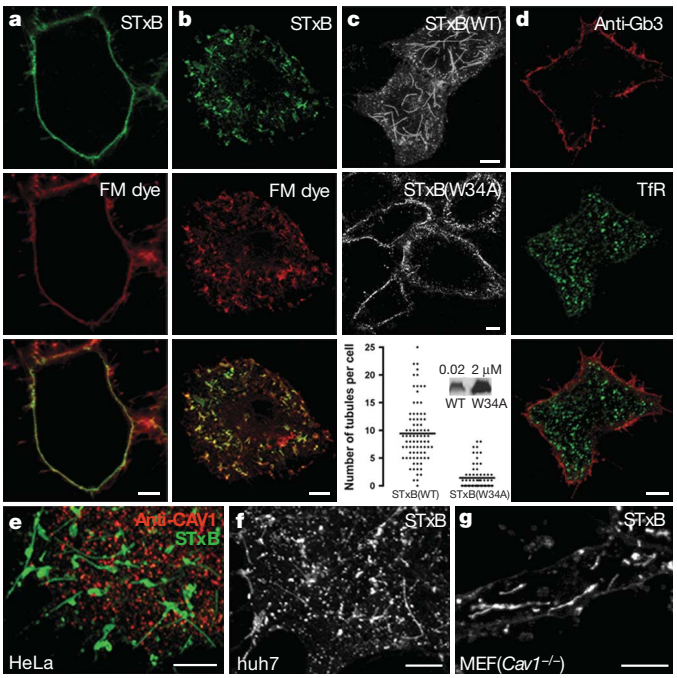


Figure 2 | STxB induces tubular invaginations on cells. **a**, No FM4-64 (red)-positive invaginations can be detected in live energy-depleted cells at 4 °C, even in the presence of STxB (green). **b**, Same as **a**, but after a shift to 37 °C. FM4-64 (red) was now found in STxB (green)-containing invaginations. **c**, On incubation with energy-depleted cells, the W34A mutant (2 μM) induces many fewer tubular invaginations than wild-type (WT) STxB (20 nM), despite the fact that more W34A was bound to cells under these conditions, as revealed by western analysis. In the histogram, each symbol represents one cell; horizontal bars indicate means. Note that 55% of data points are on the base line (0 tubules) in the W34A condition. **d**, An anti-Gb3 IgM antibody (15 μg ml^{-1}), cross-linked with secondary antibody (red), does not induce tubular invaginations on energy-depleted HeLa cells labelled for TfR (green). **e**, STxB-induced tubular invaginations on energy-depleted cells do not co-localize with immuno-stained caveolin 1. **f**, **g**, Tubular STxB-induced invaginations in caveolin-1-negative energy-depleted huh7 cells (**f**) or in Gb3-synthase-transfected mouse embryonic fibroblasts (MEF) from caveolin 1 knockout mice (**g**). Scale bars, 5 μm.

Membrane invaginations on model membranes

The cell experiments strongly indicated that STxB could induce membrane invaginations even when the cellular membrane deformation machinery was inhibited. To study whether STxB induces invaginations in the total absence of cellular proteins and to analyse the physical and biochemical conditions of tubule formation in a controlled manner, the process was reconstituted in a minimal system. Cytosol-free giant unilamellar vesicles (GUVs; typical diameters of 10–30 μm) composed of 1,2-dioleoylphosphatidyl choline (DOPC; 64 mol%; spiked with 1% Bodipy$_{FL}$-C$_5$-HPC, green), cholesterol (30 mol%), and purified porcine Gb3 (5 mol%) were observed by fluorescence microscopy on incubation with fluorophore-labelled STxB. After addition, STxB (red) bound within seconds to GUVs and was initially uniformly distributed (Fig. 3a, top panel, and Supplementary Movie 5). Spots of intense STxB labelling subsequently became apparent on the same GUV (Fig. 3a, middle panel, arrows). A decrease in membrane tension owing to GUV surface increase on Bodipy$_{FL}$-C$_5$-HPC photoactivation (J. Solon, personal communication) resulted within 1 min in the formation of tubular membrane invaginations (Fig. 3a, bottom panel). Similar observations were made when membrane tension was decreased by increasing the osmotic pressure of the bath. Narrow tubules up to several micrometres in length formed in the presence or absence of cholesterol,

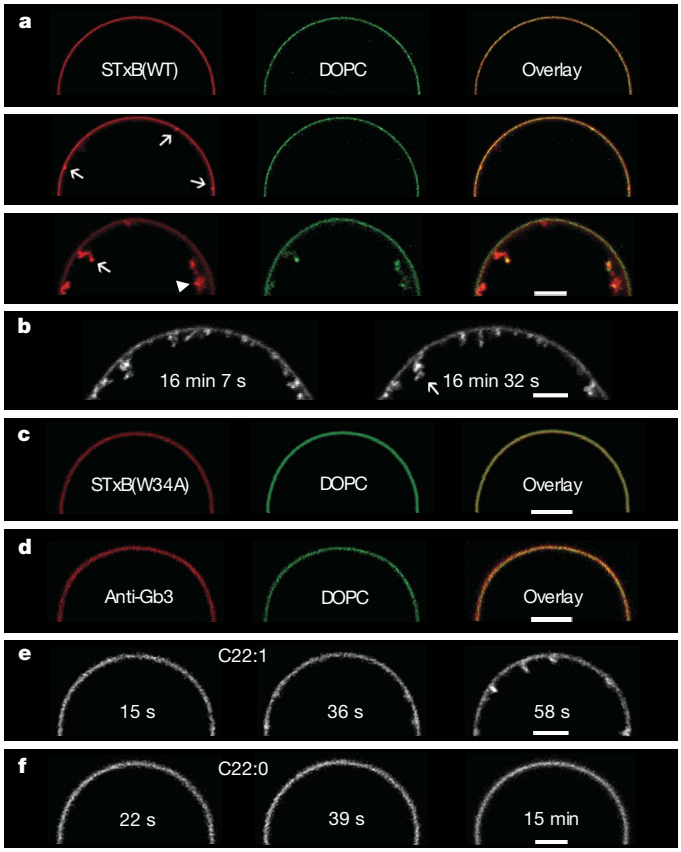


Figure 3 | Reconstitution of tubule formation on GUVs. **a**, STxB (red, 200 nM) binding at 37 °C to GUVs (upper panel) composed of DOPC (green), cholesterol and Gb3 (5 mol%); STxB spontaneous clustering (middle panel, arrows); after a decrease in membrane tension, induction of membrane invaginations (lower panel), which appear as coils (arrowhead) or spiral structures (arrow). **b**, Real time imaging as in **a**, at steady-state once invaginations have formed. **c**, The W34A mutant (200 nM) fails to induce tubular invaginations, despite its efficient binding to GUVs that were prepared as in **a**. **d**, Anti-Gb3 antibodies (15 μg ml^{-1}) do not induce invaginations, even when cross-linked with secondary antibodies (red). **e**, STxB induces tubules on GUVs that are made with Gb3 carrying a C22:1 acyl chain. **f**, STxB fails to induce tubules on GUVs with C22:0 Gb3. Scale bars, 5 μm.

©2007 Nature Publishing Group

with as little as 20 nM STxB and 0.1 mol% Gb3 (Supplementary Fig. 10). Tubules often had spiral morphology (Fig. 3a, bottom panel; arrow), and longer ones formed supercoiled structures (arrowhead). These different shapes were easily identified by time-resolved imaging of confocal sections at the medial plane of the GUVs (Fig. 3b, and Supplementary Movie 6), or by three-dimensional reconstruction of spinning disk confocal microscopy images (Supplementary Movie 7). In the absence of STxB (Supplementary Movie 8) or Gb3 (Supplementary Fig. 10), no invaginations could be detected on flaccid GUVs.

The W34A mutant (Fig. 3c and Supplementary Movie 9) and cross-linked antibodies against Gb3 (Fig. 3d) failed to induce tubular invaginations on flaccid GUVs, despite efficient binding to GUV membranes, mirroring the findings in energy-depleted cells (Fig. 2c, d), and suggesting that the invagination process requires specific structural conditions that are not met in the Gb3–antibody complex.

To study the relationship between the structure of the STxB–Gb3 complex and its function in membrane invagination, we obtained defined molecular Gb3 species by chemical synthesis (see Methods). On GUVs containing Gb3 with a C22:1 (Δ^{13}) single unsaturated acyl chain in *cis*-geometry, tubule formation could readily be observed (Fig. 3e, and Supplementary Movie 10). In contrast, no tubules were formed when Gb3 was made with a saturated C22:0 acyl chain (Fig. 3f, and Supplementary Movie 11) or lyso Gb3 (lacking one acyl chain; not shown), despite efficient STxB binding to these GUVs.

Our observations so far, and further evidence presented below, suggest that two concomitant, and probably related processes contribute to the formation of STxB-enriched membrane tubules, namely the coupling of STxB to a change of local membrane curvature (Supplementary Fig. 11) and the formation of high-density clusters of STxB–Gb3 complexes (Supplementary Fig. 12). The curvature question will be considered first.

Several mechanisms—protein scaffolding, helix insertion, lipid composition and membrane protein shape—have been described to account for membrane curvature in biological systems (summarized in refs 19–21). Considering that STxB lacks intrinsic curvature[22,23], and that glycosphingolipids have large polar head groups that are not expected to adopt negative membrane curvature[24], it seems unlikely that any of these mechanisms apply to the STxB–Gb3 system. In contrast, we propose that it is the specific organization of STxB–Gb3 complexes that imposes a preferred negative curvature to the membrane. STxB—which can bind up to 15 Gb3 molecules per pentamer[25]—may locally create an asymmetric compressive stress in the external leaflet of the lipid bilayer, which leads to membrane bending towards the protein (see schematic representation in Supplementary Fig. 11). The findings that unsaturated Gb3 species with voluminous lipid tails induce membrane invaginations (Fig. 3e), whereas saturated Gb3 species (Fig. 3f) or the binding-site mutant W34A (Fig. 2c and 3c) have a strongly reduced propensity to do so, are strong arguments in favour of this hypothesis.

STxB clustering on cell and model membranes

The above-mentioned curvature effect may result in large-scale membrane invaginations only if STxB forms high-density clusters and if the membrane tension is sufficiently low[26] (see Supplementary Fig. 12). In agreement with this prediction, spots of intense STxB labelling were frequently observed at invagination sites before tubule growth (Fig. 3a, and Supplementary Movies 5 and 10). When membrane invagination was prevented by keeping GUVs under high tension, STxB clusters reached micrometre scales (Fig. 4a, and Supplementary Fig. 13), suggesting that: (1) STxB clustering is independent of macroscopic membrane deformation, and (2) the lateral extension of STxB domains is limited by the invagination process. No STxB clusters formed on incubation of the W34A mutant with GUVs made with standard porcine Gb3, or on incubation of wild-type STxB with GUVs made with C22:0 Gb3 species (Supplementary Fig. 13). As described above (Figs 2c and 3c, f), the formation of membrane invaginations was also inhibited under these conditions, confirming that the invagination and clustering processes are related.

Atomic force microscopy on solid supported lipid bilayers was used to obtain high-resolution information on cluster structure. Protein clusters formed on the addition of STxB to membranes composed of DOPC, cholesterol and Gb3 (Fig. 4b). These structures were not detected in the absence of STxB (Fig. 4c) or Gb3 (not shown). The difference in height between protein-free and protein-covered membrane, as determined from contact mode images, was 2.3 nm ($\pm$0.4 nm, $n = 27$), which represents a 15% increase over the 2 nm height that was reported from the crystal structure of STxB[22,27]. This fact could be attributed to STxB-induced membrane thickening

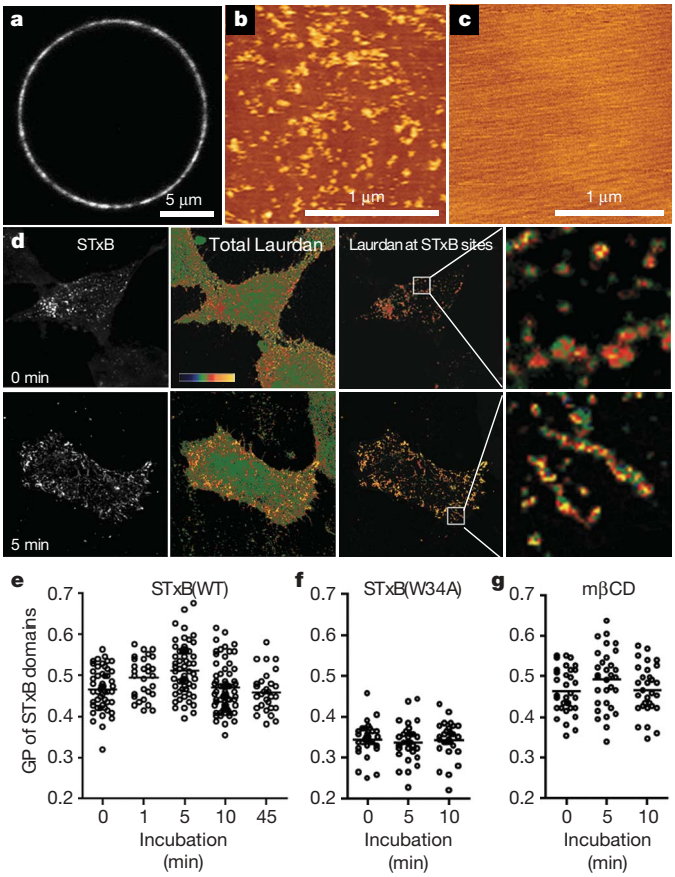


Figure 4 | STxB-induced domain formation. **a**, Experiment as in Fig. 3a, in which GUVs were kept at high membrane tension (osmotic pressure of the bath at 300 mOsm, no photoactivation), leading to the formation of extended STxB clusters. **b**, Atomic force microscopy image of STxB binding (100 nM) to supported bilayers composed of DOPC:cholesterol:Gb3 (65:30:5 mol%, respectively). STxB forms homogeneously distributed protein clusters. **c**, Experiment as in **b**, but in the absence of STxB. **d**, Laurdan generalized polarization images of fixed, energy-depleted cells after STxB binding (0 min) or after a 5 min shift to 37 °C. Red and yellow colours indicate high membrane order. **e**, Quantitative analysis of generalized polarization (GP) values at membrane domains enriched in STxB. Each symbol represents one cell; horizontal bars indicate means. STxB domains initially condense further when incubated at 37 °C, reaching peak generalized polarization values after 5 min ($P < 0.01$ for 5 min versus 0 min; Tukey's multiple comparison test), before becoming more fluid again at later time points ($P < 0.001$ for 45 min versus 5 min). **f**, Generalized polarization values of W34A STxB-enriched domains. Mean generalized polarization values of W34A domains are significantly lower ($P < 0.001$) compared to wild-type STxB domains at all time points and do not increase when incubated at 37 °C. The mean generalized polarization value of cell membranes without STxB is ~0.30, suggesting that W34A does not induce ordered domains. **g**, Cholesterol depletion with mβCD has no effect on STxB-induced membrane condensation, with no significant differences between **e** and **g** at any time point.

©2007 Nature Publishing Group

owing to higher lipid density, and is consistent with the view that STxB exerts a compressive stress on the outer-membrane leaflet, ultimately leading to the invagination of free membranes. Furthermore, thickness mismatch provides a driving force for the formation of protein-rich lipid domains[28] (see below). STxB clusters seem to be smaller on supported bilayers than on GUVs, which is in agreement with a recent report on the influence of solid substrate interactions on domain size[29].

The effect of STxB clustering on cell membrane order was analysed by two-photon laser-scanning microscopy in conjunction with the polarization-sensitive membrane dye Laurdan[30]. Membrane order is characterized by generalized polarization values ranging from -1 to $+1$ with condensed domains typically between 0.3–0.55 (ref. 31). Membrane reorganization was observed in energy-depleted cells. After binding to cells on ice and during tube formation upon a shift to 37 °C, STxB co-localized predominantly with ordered domains of high generalized polarization values (Fig. 4d). Membrane ordering was significantly more pronounced at early time points, concomitant with tubule formation (Fig. 4e, 0 versus 5 min). Analysis of global generalized polarization distribution revealed that the abundance of ordered domains was increased by STxB in all conditions, and that this effect was again strongest at early time points after a temperature shift to 37 °C (Supplementary Fig. 14). Taken together, these data suggest that membrane ordering takes place on STxB binding, and peaks at the time of formation of invaginations. The W34A mutant—which has a strongly reduced propensity to induce invaginations in cells (Fig. 2c) and model membranes (Fig. 3c)—did not increase membrane order (Fig. 4f), consistent with its reduced clustering capacity (Supplementary Fig. 13). In contrast, even after cholesterol extraction, generalized polarization values at wild-type STxB domains were high and increased during a temperature shift to 37 °C (Fig. 4g), in line with the formation of invaginations under these conditions (Fig. 1g, h).

The formation of STxB clusters is associated with the existence of a line energy between the STxB–Gb3 complexes and the surrounding membrane (see schematic representation in Supplementary Fig. 12). Line energy occurs at interfaces between membrane regions with different physical properties and may stem from higher lipid order, as suggested by our Laurdan studies (Fig. 4d, e), and increased membrane thickness in the vicinity of STxB, as deduced by atomic force microscopy (Fig. 4b). Line energy alone can only explain membrane deformations in the micrometre size range[32–34], and must be coupled to additional factors, such as the bending effects described above, to account for the high membrane curvatures that are typical for cellular membrane invaginations. Our finding that STxB domain formation is sensitive to steric constraints within the lipid tails under STxB pentamers, owing to variation of acyl chain number (W34A mutant) or saturation (C22:0 Gb3 species) (Supplementary Fig. 13), provides strong experimental evidence for a model according to which curvature-mediated interactions between STxB–Gb3 complexes could favour clustering[35,36] (see schematic representation in Supplementary Fig. 12). The observation that STxB-induced tubules have spiral morphologies (Fig. 3a, b and Supplementary Movies 6, 7) further suggests that STxB–Gb3 complexes also impose chiral properties on the membrane, which constitutes another driving force for submicrometre membrane invagination[37–39]. Both mechanisms may function in concert in the STxB-induced invagination process.

Collectively, the findings reported here demonstrate the capacity of STxB to induce membrane invaginations on cell and model membranes. When STxB binds with high affinity to its glycolipid receptor Gb3, its membrane-ordering capacity leads to the dynamic formation of STxB–Gb3 cluster domains with an intrinsic propensity to impose negative curvature on the membrane. The lateral growth of these STxB–Gb3 cluster domains seems to be regulated by the invagination process, as recently proposed for membrane rafts[40]. Narrow invaginations can form from these domains without the help of the cellular protein machinery. This cargo-induced mechanism may also apply to other extracellular pathogens, such as viruses, and more generally to other types of endocytosis events, although the factors implicated in the formation of two-dimensional membrane microdomains may differ in different contexts. Processing machineries would be recruited onto the invaginations to influence the dynamics and shapes of the forming transport intermediates.

METHODS SUMMARY

Immunofluorescence studies[41] and evanescent-field microscopy[42] were carried out as described. Cells were fixed at 37 °C to maintain tubule integrity. Cellular energy was depleted by incubating HeLa cells in PBS^{++} supplemented with 10 mM 2-deoxy-D-glucose and 10 mM NaN_3 for 30 min at 37 °C. Residual ATP levels were 2.1% ($\pm$0.46%, $n = 3$) under energy block conditions, similar to a previous report[43]. Dynasore was used at 20-80 μM [11]. Dyn2baK44A plasmid was transfected into HeLa cells for 16 h. Cholesterol was extracted using 10 mM methyl-β-cyclodextrin for 30 min[16]. GUVs were grown using the electroformation technique[44]. Supported lipid bilayers were prepared as described[45], but at a temperature of 50 °C. Membrane order of STxB-enriched membranes was analysed with the membrane dye Laurdan[30].

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

Received 15 August; accepted 5 October 2007.

1. Mayor, S. & Pagano, R. E. Pathways of clathrin-independent endocytosis. *Nature Rev. Mol. Cell Biol.* **8**, 603–612 (2007).
2. Kirkham, M. & Parton, R. G. Clathrin-independent endocytosis: new insights into caveolae and non-caveolar lipid raft carriers. *Biochim. Biophys. Acta* **1745**, 273–286 (2005).
3. Pelkmans, L. & Helenius, A. Endocytosis via caveolae. *Traffic* **3**, 311–320 (2002).
4. Conner, S. D. & Schmid, S. L. Regulated portals of entry into the cell. *Nature* **422**, 37–44 (2003).
5. Johannes, L. & Lamaze, C. Clathrin-dependent or not: Is it still the question? *Traffic* **3**, 443–451 (2002).
6. Nichols, B. J. & Lippincott-Schwartz, J. Endocytosis without clathrin coats. *Trends Cell Biol.* **11**, 406–412 (2001).
7. Nichols, B. J. *et al.* Rapid cycling of lipid raft markers between the cell surface and Golgi complex. *J. Cell Biol.* **153**, 529–541 (2001).
8. Lauvrak, S. U., Torgersen, M. L. & Sandvig, K. Efficient endosome-to-Golgi transport of Shiga toxin is dependent on dynamin and clathrin. *J. Cell Sci.* **117**, 2321–2331 (2004).
9. Saint-Pol, A. *et al.* Clathrin adaptor epsinR is required for retrograde sorting on early endosomal membranes. *Dev. Cell* **6**, 525–538 (2004).
10. Praefcke, G. J. & McMahon, H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nature Rev. Mol. Cell Biol.* **5**, 133–147 (2004).
11. Macia, E. *et al.* Dynasore, a cell-permeable inhibitor of dynamin. *Dev. Cell* **10**, 839–850 (2006).
12. Sabharanjak, S., Sharma, P., Parton, R. G. & Mayor, S. GPI-anchored proteins are delivered to recycling endosomes via a distinct cdc42-regulated, clathrin-independent pinocytic pathway. *Dev. Cell* **2**, 411–423 (2002).
13. Kirkham, M. *et al.* Ultrastructural identification of uncoated caveolin-independent early endocytic vehicles. *J. Cell Biol.* **168**, 465–476 (2005).
14. Glebov, O. O., Bright, N. A. & Nichols, B. J. Flotillin-1 defines a clathrin-independent endocytic pathway in mammalian cells. *Nature Cell Biol.* **8**, 46–54 (2006).
15. Roux, A., Uyhazi, K., Frost, A. & De Camilli, P. GTP-dependent twisting of dynamin implicates constriction and tension in membrane fission. *Nature* **441**, 528–531 (2006).
16. Falguières, T. *et al.* Targeting of Shiga toxin B-subunit to retrograde transport route in association with detergent resistant membranes. *Mol. Biol. Cell* **12**, 2453–2468 (2001).
17. Bast, D. J., Banerjee, L., Clark, C., Read, R. J. & Brunton, J. L. The identification of three biologically relevant globotriaosyl ceramide receptor binding sites on the Verotoxin 1 B subunit. *Mol. Microbiol.* **32**, 953–960 (1999).
18. Henley, J. R., Krueger, E. W., Oswald, B. J. & McNiven, M. A. Dynamin-mediated internalization of caveolae. *J. Cell Biol.* **141**, 85–99 (1998).
19. Farsad, K. & De Camilli, P. Mechanisms of membrane deformation. *Curr. Opin. Cell Biol.* **15**, 372–381 (2003).
20. McMahon, H. T. & Gallop, J. L. Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature* **438**, 590–596 (2005).
21. Zimmerberg, J. & Kozlov, M. M. How proteins produce cellular membrane curvature. *Nature Rev. Mol. Cell Biol.* **7**, 9–19 (2006).
22. Stein, P. E., Boodhoo, A., Tyrrell, G. J., Brunton, J. L. & Read, R. J. Crystal structure of the cell-binding B oligomer of verotoxin-1 from *E. coli*. *Nature* **355**, 748–750 (1992).
23. Hegnerelle, X. *et al.* Two-dimensional structures of the Shiga toxin B-subunit and of a chimera bound to the glycolipid receptor Gb3. *J. Struct. Biol.* **139**, 113–121 (2002).

©2007 Nature Publishing Group

24. Israelachvili, J. *Intramolecular and Surface Forces*. 2nd edn, pt 3 (Academic Press, 1991).
25. Ling, H. *et al.* Structure of Shiga-like toxin I B-pentamer complexed with an analogue of its receptor Gb3. *Biochemistry* **37,** 1777–1788 (1998).
26. Sens, P. & Turner, M. S. Theoretical model for the formation of caveolae and similar membrane invaginations. *Biophys. J.* **86,** 2049–2057 (2004).
27. Fraser, M. E., Chernaia, M. M., Kozlov, Y. V. & James, M. N. Crystal structure of the holotoxin from *Shigella dysenteriae* at 2.5 Å resolution. *Nature Struct. Biol.* **1,** 59–64 (1994).
28. Sens, P. & Safran, S. A. Inclusions induced phase separation in mixed lipid film. *Eur. Phys. J. E* **1,** 237–248 (2000).
29. Jensen, M. H., Morris, E. J. & Simonsen, A. C. Domain shapes, coarsening, and random patterns in ternary membranes. *Langmuir* **23,** 8135–8141 (2007).
30. Gaus, K. *et al.* Visualizing lipid structure and raft domains in living cells with two-photon microscopy. *Proc. Natl Acad. Sci. USA* **100,** 15554–15559 (2003).
31. Gaus, K., Zech, T. & Harder, T. Visualizing membrane microdomains by Laurdan 2-photon microscopy. *Mol. Membr. Biol.* **23,** 41–48 (2006).
32. Julicher, F. Domain induced budding of vesicles. *Phys. Rev. Lett.* **70,** 2964–2967 (1993).
33. Baumgart, T., Hess, S. T. & Webb, W. W. Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension. *Nature* **425,** 821–824 (2003).
34. Bacia, K., Schwille, P. & Kurzchalia, T. Sterol structure determines the separation of phases and the curvature of the liquid-ordered phase in model membranes. *Proc. Natl Acad. Sci. USA* **102,** 3272–3277 (2005).
35. Antonny, B. Membrane deformation by protein coats. *Curr. Opin. Cell Biol.* **18,** 386–394 (2006).
36. Reynwar, B. J. *et al.* Aggregation and vesiculation of membrane proteins by curvature-mediated interactions. *Nature* **447,** 461–464 (2007).
37. Oda, R., Huc, I., Schmutz, M., Candau, S. J. & MacKintosh, F. C. Tuning bilayer twist using chiral counterions. *Nature* **399,** 566–569 (1999).
38. Sarasij, R. C. & Rao, M. Tilt texture domains on a membrane and chirality induced budding. *Phys. Rev. Lett.* **88,** 088101 (2002).
39. Sarasij, R. C., Mayor, S. & Rao, M. Chirality induced budding: a raft-mediated mechanism for endocytosis and morphology of caveolae? *Biophys. J.* **92,** 3140–3158 (2007).
40. Hancock, J. F. Lipid rafts: contentious only from simplistic standpoints. *Nature Rev. Mol. Cell Biol.* **7,** 456–462 (2006).
41. Johannes, L., Tenza, D., Antony, C. & Goud, B. Retrograde transport of KDEL-bearing B-fragment of Shiga toxin. *J. Biol. Chem.* **272,** 19554–19561 (1997).
42. Merrifield, C. J., Perrais, D. & Zenisek, D. Coupling between clathrin-coated-pit invagination, cortactin recruitment, and membrane scission observed in live cells. *Cell* **121,** 593–606 (2005).
43. Zha, X. *et al.* Sphingomyelinase treatment induces ATP-independent endocytosis. *J. Cell Biol.* **140,** 39–47 (1998).
44. Mathivet, L., Cribier, S. & Devaux, P. F. Shape change and physical properties of giant phospholipid vesicles prepared in the presence of an AC electric field. *Biophys. J.* **70,** 1112–1121 (1996).
45. Menke, M., Gerke, V. & Steinem, C. Phosphatidylserine membrane domain clustering induced by annexin A2/S100A10 heterotetramer. *Biochemistry* **44,** 15296–15303 (2005).

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

Acknowledgements We thank S. Mayor and M. Rao for helpful discussions and sharing unpublished data, M. McNiven for advice, and J.-B. Sibarita and B. Stechmann for assistance with experiments. The following colleagues are acknowledged for providing materials: T. Kirchhausen, M. Bornens, M. McNiven and A. Smith. Our laboratories were supported by: Ligue Nationale contre le Cancer, Association de Recherche Contre le Cancer, Curie Institute (PIC Vectorisation), European Commission (SoftComp), CNRS (ACI Dynamique et réactivité des assemblages biologiques) and the Human Frontier Science Program Organization. W.R. holds a postdoctoral fellowship from the CNRS, and L.B. is supported by a grant from the Direction Générale pour l'Armement (DGA).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.J. (johannes@curie.fr).

©2007 Nature Publishing Group

METHODS

Materials. The following materials were purchased from the indicated suppliers: anti-dynamin antibody (BD Transduction Laboratories); anti-human-TfR antibody (Zymed); anti-CD77 (Gb3) antibody and all secondary antibodies conjugated with FITC, Cy3 and Cy5 (Beckman Coulter); 1,2-dioleoyl-*sn*-glycero-3-phosphocholine, purified porcine Gb3 (Matreya); cholesterol, 2-deoxy-D-glucose, and sodium azide (Sigma); FM4-64 (Molecular Probes). Anti-CTR433 antibody was provided by M. Bornens. STxB was purified as previously described[41]. The W34A mutant was constructed by site-directed mutagenesis, as described[41].

Cell culture and transfection. Dyn2baK44A plasmid was transfected into HeLa cells by electroporation. Cells were used 16 h after transfection. Dynasore was used at 20–80 μM, as described[11]. siRNA against dynamin 2 (GGA CAU GAU CCU GCA GUUdTdT) was from Proligo and was transfected according to the manufacturer's instructions with Oligofectamine (Invitrogen) for 72 h.

Depletion of cellular energy and cholesterol. Cellular energy was depleted by incubating HeLa cells in PBS^{++} supplemented with 10 mM 2-deoxy-D-glucose and 10 mM NaN_3 for 30 min at 37 °C. Residual ATP levels were 2.1% (±0.46%, $n = 3$) under energy-block conditions, similar to a previous report[43]. The treatment was reversible in that energy-depleted cells resumed growth after removal of the energy block. Control cells were incubated with PBS^{++}, supplemented with 5 mM glucose. Cholesterol was extracted using 10 mM methyl-β-cyclodextrin for 30 min, as described[16].

Wide-field, confocal and evanescent-field microscopy. Immunofluorescence studies were carried out, as previously described[41]. Cells were fixed at 37 °C to maintain tubule integrity. Fixed samples were examined under a confocal microscope (Leica Microsystems). For live cell imaging, HeLa cells were grown to subconfluence on FluoroDish chambers with integrated glass coverslips (World Precision Instruments). When indicated, cells were incubated in the presence of 5 μg ml^{-1} of FM4-64 dye. GUVs were transferred into a chamber containing various concentrations of labelled STxB or anti-Gb3 antibody in PBS buffer. Live-cell imaging of cells or GUVs was performed using a Zeiss inverted confocal microscope (LSM 510, Carl Zeiss) equipped with a Zeiss ×63 PL APO HCX, 1.4 numerical aperture oil immersion objective. Observations were made at room temperature or at 37 °C. Montages were prepared using Metamorph Software 6.2.6 and Photoshop 7.0 (Adobe). Only Movie 6 is based on images recorded with a Spinning Disk Confocal Microscope from Andor.

For evanescent-field microscopy, Alexa488-STxB (20 nM) was bound to energy-depleted cells, which were then shifted for 15 min to 37 °C in the continued presence of the energy-depletion compounds. The cells were placed on the stage of an IX71 microscope (Olympus) equipped with a Zeiss 1.45 numerical aperture objective. Evanescent-field illumination was obtained with a 488 nm laser (Sapphire, Coherent) adjusted for a penetration depth of 100 nm. Cells were visualized with both evanescent-field illumination and wide-field illumination (obtained with a xenon lamp and an excitation filter). During DOG/N_3 washout (at 37 °C), we performed time-lapse recordings with evanescent-field illumination every 5 s. After tubule disappearance, another image using wide-field illumination was taken.

Laurdan-2 photon microscopy and data analysis. Laurdan intensity images were recorded with a DM IRE2 Microscope (Leica) equipped with photon-multiplier tubes and Leica acquisition software. Laurdan fluorescence was excited at 800 nm with a Verdi/Mira 900 multi-photon laser system and intensity images were recorded simultaneously in the emission range of 400–460 nm and 470–530 nm[30]. Microscopy calibrations were performed as described previously[30]. For confocal STxB images of the same focal depth, a helium–neon laser was used to excite Cy3 (excitation, 543 nm; emmision, 550–620 nm) with appropriate cut-off filters and pinhole widths. For fixed cells, a 100× oil objective, 1.4 numerical aperture was used; for live cells, a 63× water objective, 1.3 numerical aperture was used.

To calculate generalized polarization (GP) images, defined as:

$$GP = \frac{I_{(400-460)} - I_{(470-530)}}{I_{(400-460)} + I_{(470-530)}}$$

intensity images were converted into floating point format, and the GP value for each pixel calculated and converted back into an 8-bit unsigned format. Final GP images were pseudo-coloured in Adobe Photoshop (see colour scale in Fig. 4d).

To determine GP values at sites enriched in STxB, confocal STxB images of the same focal depth were used to mask the GP images; the confocal images defined the regions of interest and the mean GP value of the regions of interest was determined for each image. GP values were corrected using the G-factor obtained for Laurdan in DMSO for each experiment[30].

To analyse changes in global membrane structure, GP distributions were obtained from the histograms of typically >30 GP images from a single experiment to minimize differences in Laurdan distribution between intracellular membranes, normalized (sum = 100) and fitted to two gaussian distributions using the nonlinear fitting algorithm (Microsoft Excel). The quality of the fit was tested as previously described[31].

GUV formation. Giant unilamellar vesicles (GUVs) were grown using the electroformation technique, essentially as previously described[44]. The lipid mixtures were prepared in chloroform at a total concentration of 0.5 mg ml^{-1} with the appropriate lipid ratio. Vesicles were grown in a solution of sucrose (typically at a concentration of 298 mM) adjusted to 300 mOsm, corresponding to the osmotic pressure of the protein solution. The size of observed vesicles ranged between 10 and 30 μm. The difference in density between the outside (PBS) and the inside of the vesicles led to their sedimentation.

Experiments on supported lipid bilayers. Mixed lipid films were prepared by drying lipids dissolved in chloroform under a stream of nitrogen, followed by 3 h under vacuum at 60 °C. Multilamellar vesicles were obtained by allowing the lipid films to swell in PBS at 50 °C and then vortexing several times. The resulting multilamellar vesicles were subsequently converted into large unilamellar vesicles by extrusion using a mini-extruder (LiposoFast, Avestin) supplied with a 50 nm polycarbonate membrane at 50 °C. The large unilamellar vesicles were deposited on a freshly cleaved mica plate and were incubated for 1 h at 50 °C. Surface images were obtained in an open Teflon fluid chamber using a JPK NanoWizard scanning force microscope (JPK Instruments). Measurements were performed in PBS buffer in contact/tapping mode using microfabricated silicon tips with a typical resonant frequency of 10/21 kHz and a nominal spring constant of 0.03/0.3 N m^{-1}. Usual scan rates were set between 0.6–1.1 Hz. The image resolution was 512 × 512 pixels. Height was analysed from a number of areas of several images.

Chemical methods. To synthesize the target Gb3 analogues **1–4**, 3-*O*-benzoyloxyazidosphingosine is a typical acceptor for coupling with globotriosyl donors. The acceptor is readily available from azidosphingosine by regioselective silylation of the primary OH-group followed by benzoylation then desilylation[46]. We attempted here, in Method A (below), to ensue direct regioselective acetylation of the secondary OH-group in azidosphingosine using the triethylorthoacetate method[47]. Thus azidosphingosine **5** (Supplementary Fig. 15) was treated with triethylorthoacetate in the presence of *p*-TsOH to afford an intermediate cyclic orthoester that was opened by treatment with AcOH. Unfortunately, a mixture of two compounds of the same mass corresponding to 1-acetoxy and 3-acetoxyazidosphingosines respectively was formed in the ratio of 2:1. These compounds were assigned easily from the COSY-spectrum in which the olefinic protons, centred at 5.8 and 5.5 p.p.m. showed a cross signal with H-3 at 4.13 p.p.m. in the case of **6** and at higher shift, 5.34 p.p.m. in the case of **7**. To overcome this non-regioselective acetylation and low yield of **7**, compound **5** was treated with thexyl chloride in pyridine at −40 °C to assist the selective silylation of the primary hydroxyl group followed by acetylation with Ac_2O in pyridine. Monitoring of these two steps by TLC showed only one spot, although ensuing desilylation with TBAF-AcOH in THF showed a major product of lower R_f value corresponding to **7,** and a minor product of higher R_f value corresponding to **6**. In this way (Method B; Supplementary Fig. 15), compound **5** was acetylated with high selectivity at the 3-OH group and compound **7** was obtained as an alternative to the benzoylated analogue in 68% yield over three steps along with the formation of **6** in 3% yield, which was separated easily by Flash chromatography. Compound **7** was then coupled with a known globotriaosyl trichloroacetamidate donor **8**, for which we introduced recently a facile approach for its synthesis[48], in the presence of TMSOTf (6 mol%) as promoter under the inverse procedure at temperature gradient from 0 °C to 18 °C to afford glycoside **9** in 48% yield. The relatively acidic conditions used to avoid formation of orthoester may be responsible for this moderate yield due to hydrolysis of the donor, which becomes a competitive reaction. The *β*-anomeric configuration of the newly formed stereogenic centre at the reducing terminal was easily deduced from the signal of H-2a (^{1}H NMR) δ 4.88 ($J_{1,2}$ 7.8 Hz). Deacetylation of **9** under Zemplén conditions afforded simply the acyl free Gb3, compound **1**.

To complete the synthesis of the desired analogues **2–4**, we chose the Staudinger reduction of the azido group of **9** in the presence of the corresponding acid either as anhydride, if commercially available like lauric anhydride, or as *p*-nitrophenyl ester in the case of docosanoic acid and erucic acid to quench the liberated amine as soon as it is formed, thus to avoid *trans*-acetylation of the free amine as a competitive reaction.

To prepare the novel *p*-nitrophenyl esters of **10** and **11**, the acids were condensed with *p*-nitrophenol in the presence of DIEA and DCC as dehydrating agent to produce esters **12** and **13** in very good yields (Supplementary Fig. 16). To finalize the synthesis of Gb3-analogues **2–4**, compound **9** was treated with PPh_3 in pyridine-H_2O and lauric anhydride to produce **14** (32%), or *p*-nitrophenyl ester **12** to yield **15** (17%), or *p*-nitrophenyl ester **13** to afford **16** (16%). Generally, acid anhydride afforded better yield than the *p*-nitrophenyl esters for which the yield in the case of lauric anhydride was in agreement with previous

©2007 Nature Publishing Group

coupling for other chains[49]. Assignment of intermediate acetylated Gb3-analogues **14–16** was proven easily from MS as well as ^{1}H NMR by virtue of the amide proton doublet centred at ≈5.65 p.p.m. in all analogues with nearby $J_{2,NH}$ values of ≈9.1 Hz that were in good agreement with related known analogues of Gb3[49]. Despite the low yield of amide bond formation, deacetylation of the available amounts of esters **14–16** under Zemplén conditions afforded the desired Gb3-analogues **2–4** smoothly in nearly quantitative yield to provide sufficient materials for studying their relation to the STx-endocytosis pathway. Attempts to modify the acylation step using DCC alternatives[50] were not satisfactory and purification was extremely difficult, data are not shown.

General chemical methods. Solvents were purified in the usual way. Thin layer chromatography was performed on plastic plates Silica Gel 60 F_{254} (E. Merck, layer thickness 0.2 mm). The detection was achieved by treatment with a solution of 20 g ammonium molybdate and 0.4 g cerium(IV) sulphate in 400 ml 10% H_2SO_4 or with 15% H_2SO_4, and heating at 150 °C. Flash chromatography was carried out on silica gel (Baker, 30–60 μm), type 1 silica, or Lichroprep Si 60 (Merck; diameter 15–25 μm), type-2 silica. Optical rotation was determined at 20 °C with a Perkin-Elmer 241/MC polarimeter (1 dm cell). Nuclear magnetic resonance (NMR) spectra were recorded with a Bruker AC 300 DRX instrument with Me_4Si as internal standard; *J* values are given in Hz. The assignment of ^{1}H NMR (300 MHz) spectra was based on chemical shift correlation (DQFCOSY), whereas the assignment of ^{13}C NMR (75 MHz) spectra were based on Carbon-Proton Shift-Correlation Heteronuclear Multiple Quantum Coherence (HMQC). Mass spectrometry spectra were recorded with MALDI-Kompakt (Kratos) using α-cyano-3,5-dihydroxycinamic acid (CHCA), whereas Fast Atom Bombardment (FAB) spectra were recorded with a Jeol MS700 apparatus from the Ecole Normale Supérieure (ENS, Paris).

***(2S,3R,4E)* 1-Acetyl-2-azido-4-octadecene diol and *(2S,3R,4E)* 3-Acetyl-2-azido-4-octadecene diol (6–7).** Method A: a mixture of **5** (0.48 g, 1.5 mmol), triethylorthoacetic acid (1.0 ml, 5.4 mmol) and *p*-TsOH (50 mg, 0.26 mmol) in dry CH_3CN was stirred at room temperature overnight then treated dropwise with AcOH (80%, 5 ml) in an ice-bath. Stirring was continued at room temperature overnight, then the mixture was co-evaporated with toluene *in vacuo*. The residue was purified by flash chromatography type 2 silica (5:1 petroleum ether: ethyl acetate) to yield **6** (0.24 g, 44%) as colourless oil. $[\alpha]_D$ −16.5 (*c* 0.29, $CHCl_3$); R_f 0.3 (5:1 petroleum ether:ethyl acetate); ^{1}H NMR ($CDCl_3$): δ 5.80 (m, 1-H, H-5), 5.50 (dd, 1-H, *J* 15.4, 7.4 Hz, H-4), 4.27 (dd, 1-H, J_{gem} 11.7, $J_{1,2}$ 3.6 Hz, H-1′), 4.18–4.09 (m, 2-H, H-1, H-3), 3.67 (m, 1-H, H-2), 2.13 (s, 3-H, CH_3CO), 2.11-2.03 (m, 2-H, CH_2-6), 1.94 (d, 1-H, *J* 4.3 Hz, OH), 1.25 (m, 22-H, 11 CH_2), 0.87 (t, 3-H, *J* 6.0, 7.0 Hz, CH_3); ^{13}C NMR ($CDCl_3$): δ 171.80 (C = O), 136.57 (C-5), 127.13 (C-4), 72.61 (C-3), 64.68 (C-1), 63.75 (C-2), 32.33–28.87 [11-(CH_2)-], 22.72 (CH_2-6), 20.81 (CH_3CO), 14.17 (CH_3); (high resolution mass spectrum (HRMS), FAB$^+$) *m/z* calculated for $C_{20}H_{37}N_3O_3$, [M + Na]$^+$ 390.2732. Found: 390.2733; and **7** (0.12 g, 23%) as colourless oil. $[\alpha]_D$ −100.0 (*c* 0.07, $CHCl_3$); R_f 0.19 (5:1 petroleum ether:ethyl acetate); ^{1}H NMR ($CDCl_3$): δ 5.85 (m, 1-H, H-5), 5.46 (m, 1 H, H-4), 5.35 (dd, 1-H, $J_{2,3}$ 4.8, $J_{3,4}$ 7.9 Hz, H-3), 3.73–3.62 (m, 2-H, H-1′, H-2), 3.55 (dd, 1-H, J_{gem} 10.6, $J_{1,2}$ 4.5 Hz, H-1), 2.19-2.01 (m, 5-H, CH_3CO, CH_2-6), 1.88 (m, 1-H, OH), 1.37–1.25 (m, 22-H, 11 CH_2), 0.88 (t, 3-H, *J* 6.2, 7.2 Hz, CH_3); ^{13}C NMR ($CDCl_3$): δ 169.97 (C = O), 138.61 (C-5), 123.30 (C-4), 74.10 (C-3), 65.80 (C-2), 61.93 (C-1), 32.33–28.68 [11-(CH_2)-], 22.69 (CH_2-6), 21.11 (CH_3CO), 14.13 (CH_3); (HRMS, FAB$^+$) *m/z* calculated for $C_{20}H_{37}N_3O_3$, [M + Na]$^+$ 390.2732. Found: 390.2736.

Method B: TDSCl (0.14 g, 0.71 mmol) was added dropwise with stirring under Ar to a solution of **5** (0.17 g, 0.52 mmol) in pyridine (1 ml) at −40°. The mixture was allowed to reach room temperature gradually and then co-evaporated with toluene *in vacuo* after being stirred for 48 h. The residue was purified by flash chromatography (20:1 petroleum ether:ethyl acetate) and the dry residue (0.2 g) was stirred with AcOH-pyridine (2:1, 9 ml) overnight then co-evaporated with toluene *in vacuo*. The residue was purified by flash chromatography (40:1 petroleum ether:ethyl acetate) to afford a residue (0.2 g), which was taken in dry THF (1 ml) to an ice bath and treated with AcOH (40 μl) and TBAF (1 M in THF, 0.47 ml). After reaching room temperature and stirring overnight, the mixture was evaporated *in vacuo* and the residue was taken in ethyl acetate (20 ml) and washed with brine, dried over $MgSO_4$ and evaporated *in vacuo*. The residue was purified easily by flash chromatography as described in Method A on type 1 silica to afford **6** in 3% yield and **7** in 68% yield over three steps.

***(2S,3R,4E)* 3-Acetoxy-2-azido-1-[(2,3,4,6-tetra-*O*-acetyl-*α*-D-galactopyranosyl)-(1→4)-(2,3,6-tri-*O*-acetyl-*β*-D-galactopyranosyl)-(1→4)-(2,3,6-tri-*O*-acetyl-*β*-D-glucopyranoside)]-4-octadecene (9).** A mixture of **7** (50 mg, 0.14 mmol), $Et_2O.BF_3$ (1 M in DCM, 0.09 ml) and 4 Å powdered MS (0.2 g) in dry DCM (1 ml) was stirred in an ice-bath while **8** (0.17 g, 0.16 mmol) in dry DCM (1.5 ml) was added dropwise. The ice-bath was removed and the mixture was stirred at room temperature overnight. The mixture was neutralized with Et_3N, filtered through celite and evaporated *in vacuo*. The residue was purified by flash chromatography (5:1 toluene:acetone) to yield **9** (83 mg, 48%, rise to 65%, based on recovery of the unreacted acceptor) as yellowish oil. $[\alpha]_D$ + 33.3 (*c* 0.15, $CHCl_3$); R_f 0.2 (4:1 toluene:acetone); ^{1}H NMR ($CDCl_3$): δ 5.80 (m, 1-H, H-5_{cer}), 5.56 (d, 1-H, $J_{3,4} = J_{4,5}$ 2.2 Hz, H-4_c), 5.43–5.34 (m, 2-H, H-4_{cer}, H-3_c), 5.28 (dd, 1-H, $J_{4,1}$, 8.0 Hz, H-3_{cer}), 5.20–5.14 (m, 2-H, H-3_a, H-2_c), 5.08 (dd, 1-H, $J_{1,2}$ 7.8, $J_{2,3}$ 10.7 Hz, H-2_b), 4.96 (d, 1-H, $J_{1,2}$ 3.4 Hz, H-1_c), 4.88 (dd, 1-H, $J_{1,2}$ 7.8, $J_{2,3}$ 8.9 Hz, H-2_a), 4.71 (dd, 1-H, $J_{2,3}$ 10.9, $J_{3,4}$ 2.3 Hz, H-3_b), 4.51–4.37 (m, 5-H, H-1_b, H-5_c, H-1_a, H-$\check{6}_b$, H-$\check{6}_a$), 4.14–4.04 (m, 4-H, H-$\check{6}_c$, H-6_c, H-6_a, H-6_b), 3.99 (d, 1-H, $J_{3,4}$, 2.3, $J_{4,5}$ = 1.4 Hz, H-4_b), 3.84–3.71 (m, 4-H, H-2_{cer}, H-$1'_{cer}$, H-4_a, H-5_b), 3.63 (m, 1-H, H-5_a), 3.50 (dd, 1-H, $J_{1,2}$ 4.9, J_{gem} 9.7 Hz, H-1_{cer}), 2.16–1.97 (m, 35-H, 11 CH_3CO, CH_2-6_{cer}), 1.40–1.23 [m, 22-H, 11-(CH_2)-], 0.86 (t, 3-H, *J* 6.1, 6.9 Hz, CH_3); ^{13}C NMR ($CDCl_3$): δ 170.64–168.88 (11 C = O), 138.71 (C-5_{cer}), 122.73 (C-4_{cer}), 101.11 (C-1_b), 100.09 (C-1_a), 99.58 (C-1_c), 76.24, 73.87, 73.07, 72.73, 72.52, 71.77, 71.46, 68.91, 68.77, 68.14, 67.84, 67.04, 63.01, 61.28. 60.22 (18-C), 32.28-28.67 [11-(CH_2)-], 21.03-20.46 (11 CH_3CO, CH_2-6_{cer}), 14.07 (CH_3); (HRMS, FAB$^+$) *m/z* calculated for $C_{58}H_{87}N_3O_{28}$, [M + Na]$^+$ 1,296.5274. Found: 1,296.5360.

***(2S,3R,4E)* 2-Azido-3-hydroxy-1-[(*α*-D-galactopyranosyl)-(1→4)-(*β*-D-galactopyranosyl)-(1→4)-(*β*-D-glucopyranoside)]-4-octadecene (1).** Compound **9** (0.103 mg, 81 μmol) was stirred with NaOMe (5.38 M, 50 μl) in MeOH (10 ml) for 24 h then neutralized with Amberlite IR resin 120 (H^+-form) and evaporated *in vacuo*. The residue was purified by flash chromatography (5:3:0.5 ethyl acetate:isopropanol-H_2O) to yield **1** (60 mg, 92%) as amorphous mass. R_f 0.29 (5:3:0.5 ethyl acetate:isopropanol-H_2O); (MS, MALDITOF) *m/z* calculated for $C_{36}H_{65}N_3O_{17}$, [M + Na]$^+$ 834.4211; [M + K]$^+$ 850.0322. Found: 834.4130; 850.3851.

General procedure for the synthesis of *p*-nitrophenyl esters of fatty acids (12 and 13) (Supplementary Fig. 16). A mixture of the acid **10** or **11** (0.6 mmol), *p*-nitrophenol (1.2 eq) and DIEA (1.2 eq) in dry THF (2 ml) was stirred at room temperatue while DCC (1 M in DCM, 1.2 eq) was added dropwise, and stirring was continued overnight until evaporation *in vacuo*. The residue was solubilized with ethyl acetate (20 ml), filtered at the pump and evaporated *in vacuo*.

***p*-Nitrophenyl docosanoate (12).** The residue was purified by flash chromatography (40:1 petroleum ether:ethyl acetate) to afford **12** (82%) as colourless crystals. R_f 0.18 (40:1 petroleum ether:ethyl acetate); m.p. 72–73 °C; ^{1}H NMR ($CDCl_3$): δ 8.25 (d, 2-H, *J* 9.1 Hz, H-3_{Ar}, H-5_{Ar}), 7.26 (d, 2-H, *J* 9.1 Hz, H-2_{Ar}, H-6_{Ar}), 2.59 (t, 2-H, *J* 7.5 Hz, CH_2CO), 1.73 (m, 2-H, CH_2-3), 1.40–1.25 (m, 36-H, 18 CH_2), 0.87 (t, 3-H, *J* 6.0, 6.9 Hz, CH_3)); ^{13}C NMR ($CDCl_3$): δ 171.30 (C = O), 155.52 (C-4_{Ar}), 145.21 (C-1_{Ar}), 125.16 (C-3_{Ar}, C-5_{Ar}), 122.42 (C-2_{Ar}, C-6_{Ar}), 34.32 (CH_2CO), 31.93–22.70 (19 CH_2), 14.12 (CH_3).

***p*-Nitrophenyl docosenoate (13).** The residue was purified by flash chromatography (40:1→20:1 petroleum ether:ethyl acetate) to afford **13** (82%) as semisolid residue. R_f 0.32 (20:1 petroleum ether:ethyl acetate); ^{1}H NMR ($CDCl_3$): δ 8.23 (d, 2-H, *J* 9.0 Hz, H-3_{Ar}, H-5_{Ar}), 7.24 (d, 2-H, *J* 9.0 Hz, H-2_{Ar}, H-6_{Ar}), 5.33 (m, 2-H, CH = CH), 2.58 (t, 2-H, *J* 7.5 Hz, CH_2CO), 2.00 (m, 4-H, $CH_2CH = CHCH_2$) 1.72 (m, 2-H, CH_2-3), 1.26 (m, 28-H, 14 CH_2), 0.86 (t, 3-H, 6.9 Hz, CH_3); ^{13}C NMR ($CDCl_3$): δ 171.15 (C = O), 155.42 (C-4_{Ar}), 145.07 (C-1_{Ar}), 129.79, 129.73 (CH = CH), 125.03 (C-3_{Ar}, C-5_{Ar}), 122.32 (C-2_{Ar}, C-6_{Ar}), 34.19 (CH_2CO), 31.83–22.61 (17 CH_2), 14.03 (CH_3).

***(2S,3R,4E)* 3-Acetoxy-2-(dodecanoylamino)-1-[(2,3,4,6-tetra-*O*-acetyl-*α*-*D*-galactopyranosyl)-(1→4)-(2,3,6-tri-*O*-acetyl-*β*-*D*-galactopyranosyl)-(1→4)-(2,3,6-tri-*O*-acetyl-*β*-*D*-glucopyranoside)]-4-octadecene (14).** A mixture of **9** (36 mg, 28 μmol), lauric anhydride (32 mg, 84 μmol), PPh_3 (30 mg, 0.11 mmol) in pyridine:H_2O (4.5:0.5, 2.5 ml) was stirred at room temperature for 24 h then co-evaporated with toluene *in vacuo*. The residue was purified by flash chromatography (1.5:1 ethyl acetate:petroleum ether) to yield **14** (12.7 mg, 32%) as colourless residue. $[\alpha]_D$ +26.9 (*c* 0.13, $CHCl_3$); R_f 0.2 (1.5:1 ethyl acetate:petroleum ether); ^{1}H NMR ($CDCl_3$): δ 5.76 (m, 1-H, H-5_{cer}), 5.64 (d, 1-H, *J* 9.0 Hz, NH), 5.57 (d, 1-H, $J_{3,4}$ 3.5, $J_{4,5}$ 2.3 Hz, H-4_c), 5.38 (dd, 1-H, $J_{2,3}$ 10.9, $J_{3,4}$ 3.5 Hz, H-3_c), 5.33–5.15 (m, 4-H, H-4_{cer}, H-3_{cer}, H-3_a, H-2_c), 5.09 (dd, 1-H, $J_{1,2}$ 7.8, $J_{2,3}$ 10.5 Hz, H-2_b), 4.97 (d, 1-H, $J_{1,2}$ 3.6 Hz, H-1_c), 4.88 (dd, 1-H, $J_{1,2}$ 7.9, $J_{2,3}$ 9.2 Hz, H-2_a), 4.72 (dd, 1-H, $J_{2,3}$ 10.7, $J_{3,4}$ 2.2 Hz, H-3_b), 4.52–4.39 (m, 5-H, H-1_b, H-5_c, H-1_a, H-$\check{6}_b$, H-$\check{6}_a$), 4.30 (m, 1-H, H-2_{cer}), 4.16-4.02 (m, 4-H, H-$\check{6}_c$, H-6_c, H-6_a, H-6_b), 4.01 (d, 1-H, $J_{3,4} \approx J_{4,5} < 1.0$ Hz, H-4_b), 3.90 (dd, 1-H, $J_{1',2}$ 3.4, J_{gem} 10.3 Hz, H-$1'_{cer}$), 3.82-3.73 (m, 2-H, H-4_a, H-5_b), 3.60 (m, 1-H, H-5_a), 3.53 (dd, 1-H, $J_{1,2}$ 4.5, J_{gem} 10.3 Hz, H-1_{cer}), 2.12–2.01 (m, 37-H, 11 CH_3CO, CH_2CO, CH_2-6_{cer}), 1.57–1.24 [m, 40-H, 20-(CH_2)-], 0.87 (t, 6-H, *J* 6.2, 6.9 Hz, 2 CH_3); ^{13}C NMR ($CDCl_3$): δ 172.66–168.86 (12 C = O), 137.32 (C-5_{cer}), 124.62 (C-4_{cer}), 101.09 (C-1_b), 100.28 (C-1_a), 99.65 (C-1_c), 76.20–60.25 (17-C), 50.44 (C-2_{cer}), 36.81–22.68 [21-(CH_2)-], 21.07–20.52 (11 CH_3CO, $COCH_2$), 14.12 (2-CH_3); (HRMS, FAB$^+$) *m/z* calculated for $C_{70}H_{111}NO_{29}$, [M + Na]$^+$ 1,452.7139. Found: 1,452.7104.

***(2S,3R,4E)* 3-Acetoxy-2-(docosanoylamino)-1-[(2,3,4,6-tetra-*O*-acetyl-*α*-D-galactopyranosyl)-(1→4)-(2,3,6-tri-*O*-acetyl-*β*-D-galactopyranosyl)-(1→4)-(2,3,6-tri-*O*-acetyl-*β*-D-glucopyranoside)]-4-octadecene (15).** A mixture of **9** (54 mg, 42 μmol), **12** (60 mg, 0.13 mmol), PPh_3 (45 mg, 0.17 mmol) in pyridine:H_2O (4.5:0.5, 2.5 ml) was stirred at room temperature for 24 h then

©2007 Nature Publishing Group

co-evaporated with toluene *in vacuo*. The residue was purified by flash chromatography (1.5:1 ethyl acetate:petroleum ether) to yield **15** (11.6 mg, 17%) as colourless residue. $[\alpha]_D$ +27.5 (*c* 0.12, $CHCl_3$); R_f 0.25 (1.5:1 ethyl acetate:petroleum ether); ^{1}H NMR ($CDCl_3$): δ 5.76 (m, 1-H, H-5$_{cer}$), 5.64 (d, 1-H, *J* 9.2 Hz, NH), 5.57 (d, 1-H, $J_{3,4}$ 3.5, $J_{4,5}$ 2.2 Hz, H-4$_c$), 5.38 (dd, 1-H, $J_{2,3}$ 10.7, $J_{3,4}$ 3.5 Hz, H-3$_c$), 5.33–5.16 (m, 4-H, H-4$_{cer}$, H-3$_{cer}$, H-3$_a$, H-2$_c$), 5.09 (dd, 1-H, $J_{1,2}$ 8.5, $J_{2,3}$ 11.0 Hz, H-2$_b$), 4.97 (d, 1-H, $J_{1,2}$ 3.1 Hz, H-1$_c$), 4.87 (dd, 1-H, $J_{1,2}$ 7.5, $J_{2,3}$ 9.7 Hz, H-2$_a$), 4.72 (dd, 1-H, $J_{2,3}$ 11.3, $J_{3,4}$ 2.5 Hz, H-3$_b$), 4.52–4.40 (m, 5-H, H-1$_b$, H-5$_c$, H-1$_a$, H-6̀$_b$, H-6̀$_a$), 4.31 (m, 1-H, H-2$_{cer}$), 4.20–4.04 (m, 4-H, H-6̀$_c$, H-6$_c$, H-6$_a$, H-6$_b$), 4.00 (d, 1-H, $J_{3,4} \approx J_{4,5} < 1.0$ Hz, H-4$_b$), 3.90 (dd, 1-H, $J_{1',2}$ 3.5, J_{gem} 9.9 Hz, H-1′$_{cer}$), 3.83–3.60 (m, 2-H, H-4$_a$, H-5$_b$), 3.61 (m, 1-H, H-5$_a$), 3.51 (dd, 1-H, $J_{1,2}$ 3.5, J_{gem} 9.9 Hz, H-1$_{cer}$), 2.21–1.98 (m, 35-H, 11 CH_3CO, CH_2CO, CH_2-6$_{cer}$), 1.42–1.25 [m, 60-H, 30-(CH_2)-], 0.87 (t, 6-H, *J* 5.4, 7.0 Hz, 2-CH_3); ^{13}C NMR ($CDCl_3$): δ 172.71–168.86 (12 C = O), 137.34 (C-5$_{cer}$), 124.61 (C-4$_{cer}$), 101.01, 100.27 (C-1$_a$, C-1$_b$), 99.65 (C-1$_c$), 76.59–60.25 (17-C), 50.44 (C-2$_{cer}$), 36.82–22.68 [31-(CH_2)-], 21.08–20.53 (11 CH_3CO, $COCH_2$), 14.13 (2-CH_3); (HRMS, FAB$^+$) *m/z* calculated for $C_{80}H_{131}NO_{29}$, $[M + Na]^+$ 1,592.8704. Found: 1,592.8737.

(2S,3R,4E) **3-Acetoxy-2-(docosenoylamino)-1-[(2,3,4,6-tetra-*O*-acetyl-*α*-D-galactopyranosyl)-(1→4)-(2,3,6-tri-*O*-acetyl-*β*-D-galactopyranosyl)-(1→4)-(2,3,6-tri-*O*-acetyl-*β*-D-glucopyranoside)]-4-octadecene (16)**. A mixture of **9** (54 mg, 42 μmol), **13** (41 mg, 32 μmol), PPh_3 (34 mg, 129 μmol) in pyridine: H_2O (0.5:0.02, 0.52 ml) was stirred at room temperature for 24 h then coevaporated with toluene *in vacuo*. The residue was purified by flash chromatography (1.5:1 ethyl acetate:petroleum ether) to yield **16** (8 mg, 16%) as colourless residue. $[\alpha]_D$ +16.0 (*c* 0.2, $CHCl_3$); R_f 0.20 (1.5:1 ethyl acetate:petroleum ether); ^{1}H NMR ($CDCl_3$): δ 5.76 (m, 1-H, H-5$_{cer}$), 5.65 (d, 1-H, *J* 9.4 Hz, NH), 5.57 (d, 1-H, $J_{3,4} \approx J_{4,5}$ 2.6 Hz, H-4$_c$), 5.41–5.16 (m, 7-H, H-4$_{cer}$, H-3$_c$, H-3$_{cer}$, H-3$_a$, H-2$_c$, CH = CH), 5.10 (dd, 1-H, $J_{1,2}$ 7.9, $J_{2,3}$ 10.5 Hz, H-2$_b$), 4.97 (d, 1-H, $J_{1,2}$ 3.5 Hz, H-1$_c$), 4.87 (dd, 1-H, $J_{1,2}$ 8.0, $J_{2,3}$ 9.2 Hz, H-2$_a$), 4.72 (dd, 1-H, $J_{2,3}$ 10.5, $J_{3,4}$ 1.5 Hz, H-3$_b$), 4.52–4.40 (m, 5-H, H-1$_b$, H-5$_c$, H-1$_a$, H-6̀$_b$, H-6̀$_a$), 4.30 (m, 1-H, H-2$_{cer}$), 4.20-4.04 (m, 4-H, H-6̀$_c$, H-6$_c$, H-6$_a$, H-6$_b$), 4.01 (d, 1-H, $J_{3,4}$ 1.5, $J_{4,5} < 1.0$ Hz, H-4$_b$), 3.90 (dd, 1-H, $J_{1',2}$ 3.2, J_{gem} 10.0 Hz, H-1′$_{cer}$), 3.82-3.74 (m, 2-H, H-4$_a$, H-5$_b$), 3.61 (m, 1-H, H-5$_a$), 3.51 (dd, 1-H, $J_{1,2}$ 4.3, J_{gem} 10.0 Hz, H-1$_{cer}$), 2.24–2.01 (m, 41-H, 11 CH_3CO, CH_2CO, $CH_2CH = CHCH_2$, CH_2-6$_{cer}$), 1.71–1.25 [m, 52-H, 26-(CH_2)-], 0.87 (t, 6-H, *J* 6.8 Hz, 2-CH_3); ^{13}C NMR ($CDCl_3$): δ 172.69–168.86 (12 C = O), 137.35 (C-5$_{cer}$), 129.91 (CH = CH-$_{acyl}$), 124.61 (C-4$_{cer}$), 101.10 (C-1$_b$), 100.27 (C-1$_a$), 99.65 (C-1$_c$), 76.60–60.24 (17-C), 50.43 (C-2$_{cer}$), 36.82–20.53 (30-CH_2, 11 CH_3CO), 14.14 (2-CH_3); (HRMS, FAB$^+$) *m/z* calculated for $C_{80}H_{129}NO_{29}$, $[M + Na]^+$. 1,590.8548. Found: 1,590.8579.

General procedure for the synthesis of compounds (2–4). A mixture of the ester derivative **14–16** (5 μmol) in NaOMe:MeOH (0.05 M, 2 ml) was stirred at room temperature overnight, then neutralized with Amberlite IR-120 resin (H^+-form) and evaporated *in vacuo* to afford the corresponding Gb3 derivative.

(2S,3R,4E) **2-(Dodecanoylamino)-3-hydroxy-1-[(*α*-D-galactopyranosyl)-(1→4)-(*β*-D-galactopyranosyl)-(1→4)-(-*β*-D-glucopyranoside)]-4-octadecene (2)**. (97%) As amorphous yellowish mass. $[\alpha]_D$ +8.0 (*c* 0.05, $CHCl_3$); R_f 0.6 (5:3:1 ethyl acetate:^{iso}PrOH:H_2O); ^{1}H NMR (MeOH-d_4): δ 5.68 (m, 1-H, H-5), 5.44 (dd, 1-H, *J* 7.5, 15.0 Hz, H-4), 4.94 (d, 1-H, $J_{1,2}$ 3.2 Hz, H-1$_c$), 4.14–3.2 (m, 24-H), 2.19 (m, 2-H, $COCH_2$), 2.01 (m, 2-H, = $CHCH_2$), 1.51 (m, 2-H, $COCH_2CH_2$), 1.28 (m, 38-H, 19 CH_2;), 0.90 (t, 6-H, *J* 6.3, 6.9 Hz, 2 CH_3); (MS, MALDITOF) *m/z* calculated for $C_{48}H_{89}NO_{18}$, $[M + Na]^+$ 990.5977; $[M + K]^+$ 1,006.5716. Found: 990.6281; 1,006.5989.

(2S,3R,4E) **2-(Docosanoylamino)-3-hydroxy-1-[(*α*-D-galactopyranosyl)-(1→4)-(*β*-D-galactopyranosyl)-(1→4)-(*β*-D-glucopyranoside)]-4-octadecene (3)**. The residue was purified by flash chromatography (5:3:1 ethyl acetate:^{iso}PrOH:H_2O) to afford **3** (97%) as amorphous yellowish mass. $[\alpha]_D$ +19.5 (*c* 0.24, $CHCl_3$); R_f 0.19 (10:6:1 ethyl acetate:^{iso}PrOH:H_2O); ^{1}H NMR (MeOH-d_4): δ 5.66 (m, 1-H, H-5$_{cer}$), 5.41 (m, 1-H, H-4$_{cer}$), 4.94–3.20 (m, 25-H), 2.21 (m, 2-H, $COCH_2$), 2.01 (m, 2-H, = $CHCH_2$), 1.57 (m, 2-H, $COCH_2CH_2$), 1.39–1.25 (m, 58-H, 29 CH_2), 0.90 (m, 6-H, 2 CH_3); (MS, MALDITOF) *m/z* Calculated for $C_{58}H_{109}NO_{18}$, $[M + Na]^+$ 1,130.7542; $[M + K]^+$ 1146.7281. Found: 1,130.7555; 1146.7381.

(2S,3R,4E) **2-(docosenoylamino)-3-hydroxy-1-[(*α*-D-galactopyranosyl)-(1→4)-(*β*-D-galactopyranosyl)-(1→4)-(*β*-D-glucopyranoside)]-4-octadecene (4)**. The residue was purified by flash chromatography (6:3:2 ethyl acetate:^{iso}PrOH:H_2O) to afford 4 (97%) as amorphous yellowish mass. $[\alpha]_D$ +13.5 (*c* 0.14, $CHCl_3$); R_f 0.51 (6:3:2 ethyl acetate:^{iso}PrOH:H_2O); ^{1}H NMR (MeOH-d_4): δ 5.68 (m, 1-H, H-5$_{cer}$), 5.41 (m, 1-H, H-4$_{cer}$), 5.34 (t, 2-H, *J* 4.7 Hz, CH = CH-$_{acyl}$), 4.94 (d, 1-H, $J_{1,2}$ 3.0 Hz, H-1$_c$), 4.41–3.34 (m, 24-H), 2.17 (t, 2-H, *J* 7.4 Hz, $COCH_2$), 2.03–2.02 (m, 6-H, 3 = $CHCH_2$), 1.58–1.21 (m, 52-H, 26 CH_2), 0.90 (m, 6-H, *J* 6.8 Hz, 2 CH_3); (MS, MALDITOF) *m/z* calculated for $C_{58}H_{107}NO_{18}$, $[M + Na]^+$ 1,128.7385. Found: 1,129.1710.

46. Schmidt, R. R. & Zimmermann, P. Synthesis of D-erythro-sphingosines. *Tetrahedr. Lett.* **27**, 481–484 (1986).
47. Manzoni, L., Lay, L. & Schmidt, R. R. Synthesis of Lewis A and Lewis X pentasaccharides based on *N*-trichloroethoxycarbonyl protection. *J. Carbohydr. Chem.* **17**, 739–758 (1998).
48. Aly, M. R., Rochaix, P., Amessou, M., Johannes, L. & Florent, J. C. Efficient Synthesis of globo- and isoglobotriosides bearing a cinnamoylphenyl tag as novel electrophilic thiol-specific carbohydrate reagents. *Carbohydr. Res.* **341**, 2026–2036 (2006).
49. Qiu, D. & Schmidt, R. R. Glycosyl imidates. LII. Synthesis of globotriaosylceramide (Gb_3) and isoglobotriaosylceramide (isoGb$_3$). *Liebigs Ann. Chem.* **3**, 217–224 (1992).
50. Figueroa-Perez, S. & Schmidt, R. R. Total synthesis of α-galactosyl cerebroside. *Carbohydr. Res.* **328**, 95–102 (2000).

©2007 Nature Publishing Group

ARTICLES

A dual-Ca^{2+}-sensor model for neurotransmitter release in a central synapse

Jianyuan Sun[1,2], Zhiping P. Pang[1], Dengkui Qin[1], Abigail T. Fahim[4], Roberto Adachi[4] & Thomas C. Südhof[1,2,3]

Ca^{2+}-triggered synchronous neurotransmitter release is well described, but asynchronous release—in fact, its very existence—remains enigmatic. Here we report a quantitative description of asynchronous neurotransmitter release in calyx-of-Held synapses. We show that deletion of synaptotagmin 2 (*Syt2*) in mice selectively abolishes synchronous release, allowing us to study pure asynchronous release in isolation. Using photolysis experiments of caged Ca^{2+}, we demonstrate that asynchronous release displays a Ca^{2+} cooperativity of ~2 with a Ca^{2+} affinity of ~44 μM, in contrast to synchronous release, which exhibits a Ca^{2+} cooperativity of ~5 with a Ca^{2+} affinity of ~38 μM. Our results reveal that release triggered in wild-type synapses at low Ca^{2+} concentrations is physiologically asynchronous, and that asynchronous release completely empties the readily releasable pool of vesicles during sustained elevations of Ca^{2+}. We propose a dual-Ca^{2+}-sensor model of release that quantitatively describes the contributions of synchronous and asynchronous release under conditions of different presynaptic Ca^{2+} dynamics.

Two modes of Ca^{2+}-triggered neurotransmitter release have been described: fast synchronous release predominates in all synapses during low-frequency action-potential firing[1,2]; slower asynchronous release mediates synaptic transmission in some synapses during high-frequency action-potential trains[3–7], but remains a minor component in other synapses[1,2]. Precise measurements of Ca^{2+} triggering of synchronous release were obtained in the calyx-of-Held synapse, which allows simultaneous patching of pre- and postsynaptic neurons, and enables monitoring of Ca^{2+} currents and capacitance of nerve terminals[8–12]. Such measurements revealed that the Ca^{2+} sensor for synchronous release exhibits an apparent cooperativity of ~5, and an apparent K_d of ~10 or ~105 μM Ca^{2+} (refs 11 and 12). In vertebrate synapses, synaptotagmin 1, 2 and 9 (Syt1, Syt2 and Syt9) function as Ca^{2+} sensors for fast synchronous release[13–18], exhibit a binding stoichiometry of 5 Ca^{2+} ions per molecule, and a micromolar Ca^{2+} affinity (note that 12 other synaptotagmins are expressed in the brain that do not function as Ca^{2+} sensors for fast release). Of Syt1, Syt2 and Syt9, Syt2 is a likely Ca^{2+} sensor for synchronous release at the calyx synapse because a mutation that decreases Syt2 levels impairs synchronous release from calyx terminals[19].

In synapses with predominantly asynchronous release during high-frequency action-potential trains, asynchronous release outcompetes synchronous release during the action-potential train[3–7]. Measurements of asynchronous release suggested a higher apparent Ca^{2+} affinity, but the same Ca^{2+} cooperativity as synchronous release[20,21]. These properties would explain the ability of asynchronous release to out-compete synchronous release during high-frequency action-potential trains, because the accumulating residual Ca^{2+} in the train would trigger asynchronous release in the intervals between action potentials at Ca^{2+} levels at which synchronous release cannot be induced. Consistent with this notion, Ca^{2+} chelators such as EGTA block asynchronous release during high-frequency action-potential trains but have less effect on synchronous release[3–6,22]. According to this view, synapses without predominantly asynchronous release during high-frequency trains, such as the calyx synapse, either lack asynchronous release, or have efficient Ca^{2+} buffering and Ca^{2+} extrusion mechanisms that prevent accumulation of residual Ca^{2+} (refs 23 and 24). Moreover, the similar apparent Ca^{2+} cooperativity of synchronous and asynchronous release suggested that asynchronous release could be a modification of synchronous release[25], indicating that there are no separate universal synchronous and asynchronous release pathways[24]. Indeed, quantitative models of release that are based on the assumption of a single Ca^{2+} sensor are widely applicable[11,12]. These models, however, are inaccurate at low Ca^{2+} concentrations; even a recently refined model that proposes allosterically modulated release machinery with a single Ca^{2+} sensor is unable to accurately predict release rates at low Ca^{2+} concentrations[25]. Thus, at present, no accurate quantitative model for neurotransmitter release is available. Evaluating asynchronous release and generating an accurate quantitative model for release is difficult because asynchronous release has primarily been characterized in synapses that have a strong synchronous component, which could have contaminated the measurements[20,21], and because the biophysical properties of asynchronous release remain unknown. However, these questions have wide implications for our understanding of synaptic transmission because an accurate description of asynchronous release is required for any quantitative model of synaptic transmission. We now provide such a description in the calyx-of-Held synapse, and demonstrate that asynchronous release represents a separate and distinct release pathway that is likely to be universally present in synapses.

Syt2 knockout blocks synchronous release

Mice lacking Syt2 initially develop normally, but perish after postnatal day 21 (P21). In the calyx-of-Held synapse, we detected no co-expression of Syt1 with Syt2 at any time between P5 and P14, or of Syt9 at P11 (Fig. 1a, and Supplementary Fig. 2). Moreover, we observed no compensatory increases of Syt1 in Syt2-deficient calyces, and found no overall change in protein composition (Fig. 1b, and Supplementary Figs 2–4). Thus, calyx terminals express only one (Syt2) of the three synaptotagmins (Syt1, Syt2 and Syt9) that serve as Ca^{2+} sensors for synchronous release[18,19], presumably because Syt2

[1]Departments of Neuroscience and [2]Molecular Genetics, and [3]Howard Hughes Medical Institute, The University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA.
[4]Department of Pulmonary Medicine, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, USA.

©2007 Nature Publishing Group

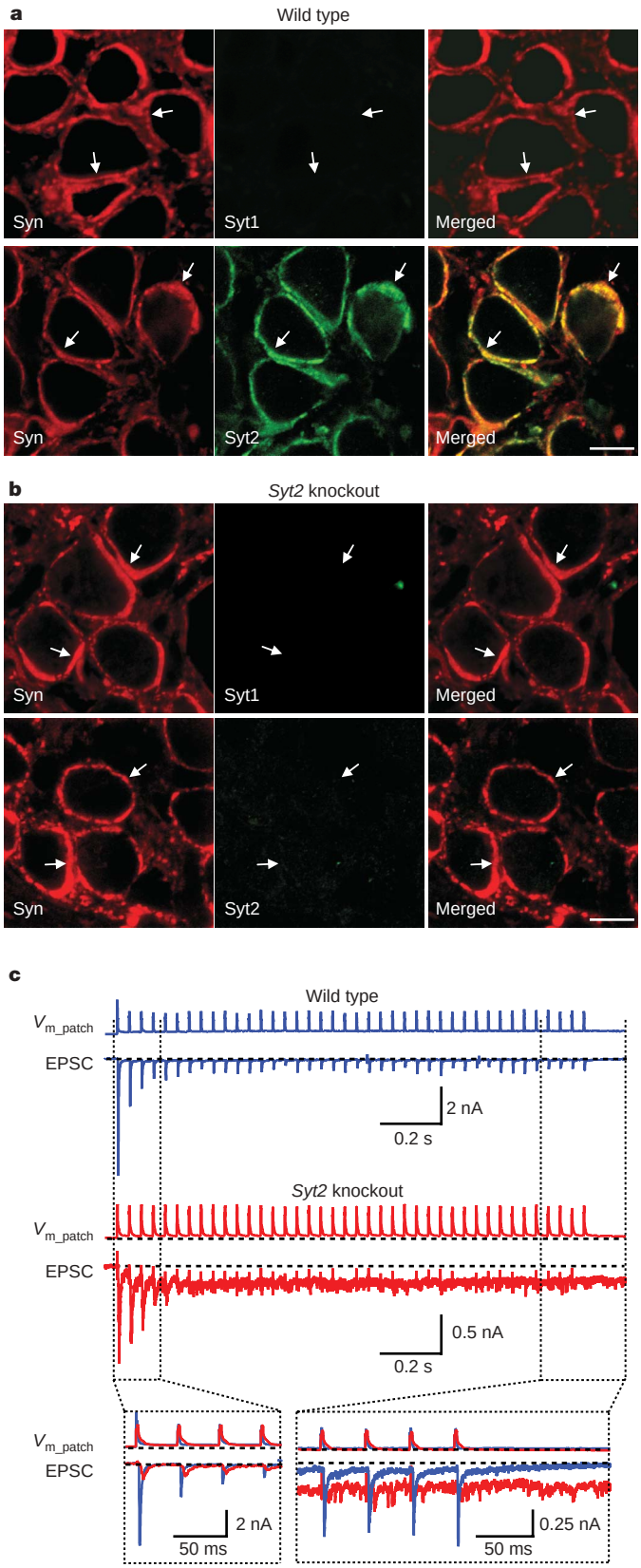


Figure 1 | Calyx synapses in Syt2-deficient mice. **a**, **b**, Immunofluorescence analysis of brainstem sections from wild-type (**a**) and Syt2-deficient mice (**b**) at P11 with antibodies to synapsins 1–3 (Syn; red; left panels) and Syt1 or Syt2 (green; middle panels; Supplementary Fig. 2). Merged images are shown on the right with coincident staining in yellow (scale bar, 10 μm, applies to all panels; arrows, synapses). **c**, EPSCs recorded in response to 25 Hz action-potential stimulation induced by a cell-attached presynaptic pipette (V_{m_patch}, extracellular voltage in the patched area). The inset at the bottom shows a superposition of wild-type and mutant traces.

is the fastest Ca^{2+} sensor[18] and the calyx synapse is specialized for precise fast responses[8,9].

To determine the effect of the Syt2 deletion on release, we examined brainstem slices from 7–9-day-old mice. Using cell-attached patches on presynaptic calyx terminals, we induced a train of presynaptic action potentials and measured postsynaptic responses by whole-cell recordings. Strikingly, evoked synaptic responses in Syt2-deficient terminals were small and delayed, with release being increasingly triggered in the inter-stimulus intervals during the stimulus train (Fig. 1c). These responses are consistent with the

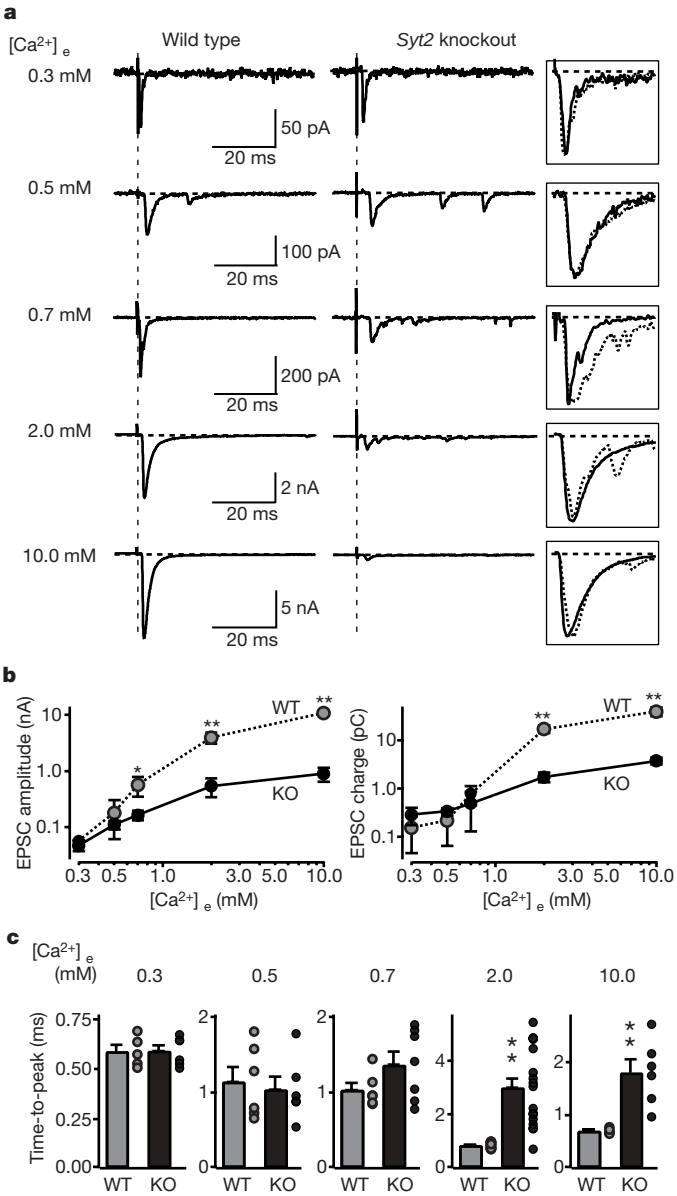


Figure 2 | Synaptic transmission evoked by isolated action potentials. Postsynaptic voltage-clamp recordings of ESPCs evoked by afferent fibre stimulation in wild-type (left) and Syt2-deficient calyx synapses (right) at the indicated $[Ca^{2+}]_e$ in the presence of 50 μM AP-5. **a**, Representative EPSCs. Boxed traces on right display scaled superimposed EPSCs to illustrate EPSC kinetics. **b**, Double-logarithmic plots of EPSC amplitudes (left panel) or EPSC charge transfer (right panel) as a function of $[Ca^{2+}]_e$ (mean ± s.e.m.; number of recordings/$[Ca^{2+}]_e$: wild-type (WT), 5/0.3 mM, 6/0.5 mM, 6/0.7 mM, 14/2 mM, 5/10 mM; KO, 6/0.3 mM, 6/0.5 mM, 7/0.7 mM, 17/2 mM, 6/10 mM). **c**, Summary graphs of the rise times of EPSCs recorded for the indicated $[Ca^{2+}]_e$. For each Ca^{2+} concentration, the bar depicts mean ± s.e.m.; dots next to the bar exhibit individual data points (*, $P < 0.05$; **, $P < 0.01$; Student's *t*-test).

©2007 Nature Publishing Group

notion that, in the absence of Syt2 as the Ca^{2+} sensor for synchronous release, accumulating Ca^{2+} during the stimulus train triggers asynchronous release.

To characterize this asynchronous release, we induced presynaptic action potentials by afferent fibre stimulations, and recorded excitatory postsynaptic currents (EPSCs) at different extracellular Ca^{2+} concentrations ($[Ca^{2+}]_e = 0.3$–10 mM). At 0.3–0.7 mM $[Ca^{2+}]_e$, wild-type and mutant terminals exhibited similar amounts of release (Fig. 2a, b). At 2 mM and 10 mM $[Ca^{2+}]_e$, however, Syt2-deficient terminals displayed >10-fold smaller EPSC amplitudes and charge transfers than wild-type terminals (Fig. 2a, b). Moreover, at 2 mM $[Ca^{2+}]_e$ (a nearly physiological concentration), Syt2-deficient synapses were ~3-times slower in reaching the EPSC maximum (Fig. 2c), and exhibited ~5-fold slower release kinetics (measured as the time to achieve 50% synaptic charge transfer; wild-type (WT) = 4.9 ± 0.7 ms; *Syt2* knockout (KO) = 26.8 ± 1.7 ms (means ± s.e.m.); $P < 0.01$; Supplementary Fig. 5). We conclude that deletion of Syt2 severely impairs fast synchronous release in the calyx-of-Held synapse, but leaves a slower, asynchronous form of Ca^{2+}-triggered release intact. In addition, the deletion of Syt2, like that of Syt1 in forebrain[19], increased the frequency of miniature spontaneous release events but did not alter the size and kinetics of these events (Supplementary Fig. 6).

A potential concern is that, because the *Syt2* knockout mice die at P21–P24, the knockout neurons may not be healthy. However, at P7–P14, when we analyse the *Syt2* knockout mice, the knockout mice are not visibly distressed, the protein composition of their brainstems is not detectably altered (Supplementary Fig. 4), and the basic electrical properties of the postsynaptic calyx neurons are not impaired (Supplementary Table 1). Thus, the changes we observe in the knockout mice are probably specifically caused by the deletion of Syt2.

Ca^{2+} currents and readily releasable pool size

In presynaptic terminals of the calyx of Held, action potentials gate Ca^{2+} influx via P/Q- and N-type Ca^{2+} channels[26,27]. Because Syt1 interacts with Ca^{2+} channels[28–30], we tested whether deletion of the highly homologous Syt2 impairs Ca^{2+} channel function. We simultaneously patched presynaptic calyx terminals and postsynaptic neurons in the presence of drugs that block action-potential propagation, AMPA(α-amino-3-hydroxy-5-methylisoxazole-propionic acid)-receptor desensitization, and NMDA (*N*-methyl-D-aspartic acid)-receptor activation. We then recorded presynaptic Ca^{2+} currents and postsynaptic EPSCs in response to prolonged depolarization (4 ms predepolarization from −80 to +80 mV followed by 50 ms depolarization to +20 mV)[31]. Deletion of Syt2 caused no detectable change in the amplitudes of presynaptic Ca^{2+} currents, electrical charge transfer mediated by the Ca^{2+} channels (integrated over 100 ms), or the *I*/*V* relationship, suggesting that Syt2 is not involved in regulating Ca^{2+} channels (Figs 3a–3c, and Supplementary Figs 7 and 8).

The 50 ms depolarization in Fig. 3a depletes the readily releasable pool (RRP) of vesicles by inducing a prolonged increase in intracellular $[Ca^{2+}]_i$ (ref. 31). Postsynaptic recordings of synaptic responses showed that deletion of Syt2 depressed the peak amplitude of the depolarization-induced EPSC ~2-fold, slowed its rise time ~3-fold, and increased its latency ~5-fold (Fig. 3d–f). However, the Syt2 deletion did not alter the total synaptic charge transfer induced by the 50 ms presynaptic depolarization (Fig. 3g; integrated over 2 s). Thus, deletion of Syt2 did not affect the size of the RRP, and asynchronous release induces exocytosis of the entire RRP of vesicles in the absence of Syt2, albeit with a slower time course.

To compare the size of the RRP in wild-type and *Syt2* knockout mice by an independent approach, we puffed a 2 M sucrose solution onto the terminal for 1 s, and integrated the synaptic charge transfer of the induced EPSCs over 2 s (Figs 3h, i, and Supplementary Fig. 9). Again, no difference between wild-type and Syt2-deficient terminals was detected. The absolute size of the RRP determined after sustained depolarization was larger than the RRP size measured by puffing of hypertonic sucrose (Fig. 3). This difference may be due to the distinct measurement conditions used (ESPCs induced by depolarization but not by sucrose were monitored in the presence of cyclothiazide as a blocker of receptor desensitization); in addition, Ca^{2+}-dependent mobilization of the RRP during the depolarization may have increased its size during the monitoring period[32].

Ca^{2+} dependence of release

To characterize the Ca^{2+} dependence of transmitter release at wild-type and Syt2-deficient synapses, we determined the peak release rates in calyx terminals as a function of the presynaptic intracellular Ca^{2+} concentration $[Ca^{2+}]_i$. For Ca^{2+} concentrations of <1 μM, we employed two methods: (1) injection of terminals via the patch pipette with defined concentrations of Ca^{2+} and Ca^{2+} buffers, or (2) with caged Ca^{2+} (9 mM DM-nitrophen, 8.6 mM $CaCl_2$, and various Ca^{2+} buffers) that is released by flash photolysis with a weak laser pulse (Supplementary Fig. 10)[11,12,33,34]. For Ca^{2+} concentrations of >1 μM, we employed only the flash photolysis method with a

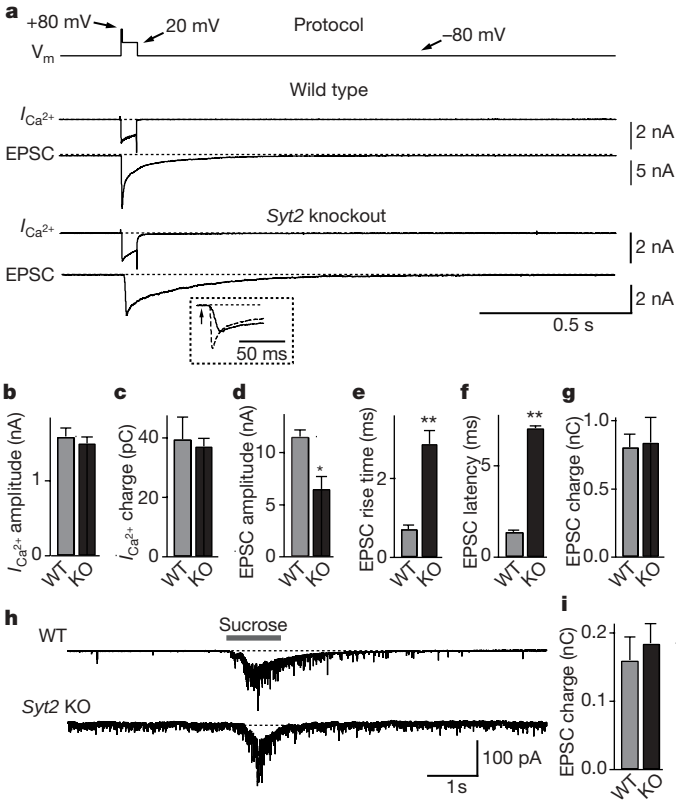


Figure 3 | RRP size, Ca^{2+} currents and release kinetics in Syt2-deficient calyx synapses. **a**, Experimental protocol involving a 4-ms predepolarization followed by a 50-ms depolarization (upper line), and representative traces of presynaptic Ca^{2+} currents ($I_{Ca^{2+}}$) and postsynaptic EPSCs (lower traces). The inset displays an expansion of the initial phase of the EPSCs from wild-type (dotted line) and Syt2-deficient terminals (continuous line; arrow = start of Ca^{2+} current). Experiments were performed by simultaneous pre- and postsynaptic voltage-clamp recordings in calyx terminals at P7–P9 in 0.1 μM tetrodotoxin, 0.1 mM cyclothiazide, 1 mM kynurenic acid and 50 μM D-AP5. **b**–**g**, Quantifications of Ca^{2+} current amplitudes (**b**), Ca^{2+}-current electrical charge transfer (**c**, integrated over 100 ms), EPSC amplitudes (**d**), EPSC rise times (**e**; 20–80%), EPSC latencies (**f**; from onset of Ca^{2+} current to 10% of the EPSC) and EPSC charge transfer (**g**; integrated over 2 s) induced by sustained presynaptic depolarization. Data shown are mean ± s.e.m. (WT, $n = 12$; KO, $n = 14$). **h**, **i**, Representative traces (**h**) and summary graphs (**i**; integrated over 5 s) of the electric charge transfer of synaptic responses induced by 1-s applications of 2 M sucrose via a glass pipette positioned ~5 μm from the calyx (mean ± s.e.m.; WT, $n = 10$; KO, $n = 11$).

©2007 Nature Publishing Group

stronger laser pulse. Flash photolysis of caged Ca^{2+} produces a rapid and spatially uniform, defined rise in $[Ca^{2+}]_i$, which we monitored *in situ* in the terminals, using co-injected Ca^{2+}-indicator dyes. To cover the entire range of Ca^{2+} concentrations examined (0.1–15 μM), we employed three different Ca^{2+}-indicator dyes (Fura-2, Fura-4F and Fura-6F), and calibrated the Ca^{2+} signals directly in the calyx terminals (Supplementary Fig. 10). Finally, we quantified vesicle exocytosis by deconvolution of evoked EPSCs, using a measured miniature EPSC waveform to calculate the release rates[31] (see Methods and Supplementary Materials).

We first elevated $[Ca^{2+}]_i$ to concentrations of 0.05–1.0 μM. Increases of $[Ca^{2+}]_i$ to <0.2 μM induced an enhancement in miniature EPSC frequency but no clearcut evoked EPSCs (Fig. 4a). At these Ca^{2+} concentrations, release was slightly higher in Syt2-deficient than wild-type synapses (Fig. 4e), presumably because deletion of Syt2 increases the resting frequency of miniature EPSCs (Supplementary Fig. 6). At $[Ca^{2+}]_i$ of 0.2–1.0 μM, wild-type and Syt2-deficient synapses exhibited indistinguishable release rates, independent of whether the $[Ca^{2+}]_i$ was constantly clamped with Ca^{2+} buffers, or acutely raised by photolysis of caged Ca^{2+} with a weak flash (Fig. 4, and Supplementary Fig. 11). Thus, as observed for action-potential-induced release at low $[Ca^{2+}]_e$ (Fig. 2), deletion of Syt2 does not impair release evoked at low $[Ca^{2+}]_i$. This result suggests that, even in wild-type synapses, release at low $[Ca^{2+}]_i$ is physiologically mediated by the asynchronous Ca^{2+} sensor, a conclusion that is consistent with previously described properties of release induced by low $[Ca^{2+}]_i$ (ref. 35).

We next examined larger increases in $[Ca^{2+}]_i$ produced by flash photolysis of caged Ca^{2+}. In wild-type neurons, a peak release rate of ~24 vesicles ms^{-1} was evoked when $[Ca^{2+}]_i$ was elevated to 2 μM (Fig. 4c). The peak release rate steeply increased to ~1,750 vesicles ms^{-1} when $[Ca^{2+}]_i$ was raised to 9.2 μM, demonstrating a very high Ca^{2+} cooperativity (Fig. 4d). In Syt2-deficient synapses, conversely, a 2 μM $[Ca^{2+}]_i$ rise caused release with a peak rate of 3.7 vesicle ms^{-1}. Elevation of $[Ca^{2+}]_i$ to 10.5 μM only increased the release rate to 31 vesicles ms^{-1}, revealing a ~50-fold lower release rate and a much lower Ca^{2+} cooperativity in Syt2-deficient synapses than in wild-type synapses.

Figure 4e displays the Ca^{2+} dependence of the peak release rate obtained from 78 wild-type (open symbols) and 106 Syt2-deficient terminals (filled symbols), plotted on logarithmic coordinates (see Supplementary Table 2 for numerical values). Whereas at $[Ca^{2+}]_i$ of <1 μM, the vesicular release rates are not decreased in Syt2-deficient terminals, at $[Ca^{2+}]_i > 2$ μM, deletion of Syt2 reduced the peak transmitter release rate 10–50-fold compared to wild-type synapses. In the most dynamic range of 0.7–5.0 μM $[Ca^{2+}]_i$, the relation of the peak release rate to $[Ca^{2+}]_i$ follows a fifth power function in wild-type terminals, but only a second power function in Syt2-deficient terminals (Fig. 4e). At higher $[Ca^{2+}]_i$, release saturates in wild-type synapses, but not in knockout synapses. Thus, whereas release triggered at $[Ca^{2+}]_i = 0.2$–1.0 μM exhibits a similarly low Ca^{2+} cooperativity and magnitude in wild-type and Syt2-deficient synapses, release triggered at $[Ca^{2+}]_i > 1$ μM exhibits a dramatically different Ca^{2+} cooperativity and magnitude in wild-type and mutant synapses.

A potential concern with a genetic study is that a compensatory developmental change could lead to the expression of a non-physiological Ca^{2+} sensor in mutant synapses. However, release at low $[Ca^{2+}]_i$, measured with $[Ca^{2+}]_i$ raised by two different methods, is indistinguishable between wild-type and Syt2-deficient synapses. This fact argues strongly against such a developmental change in the mutant terminals because such a change should have affected the behaviour of Ca^{2+}-triggered release at all Ca^{2+} concentrations. Moreover, the good prediction of release in both wild-type and mutant synapses by our quantitative model described below is inconsistent with such a change.

A dual-Ca^{2+}-sensor model of release

Previous Ca^{2+} photolysis studies on calyx synapses prompted formulation of a quantitative model of neurotransmitter release with a single Ca^{2+} sensor for release (referred to here as the 'conventional

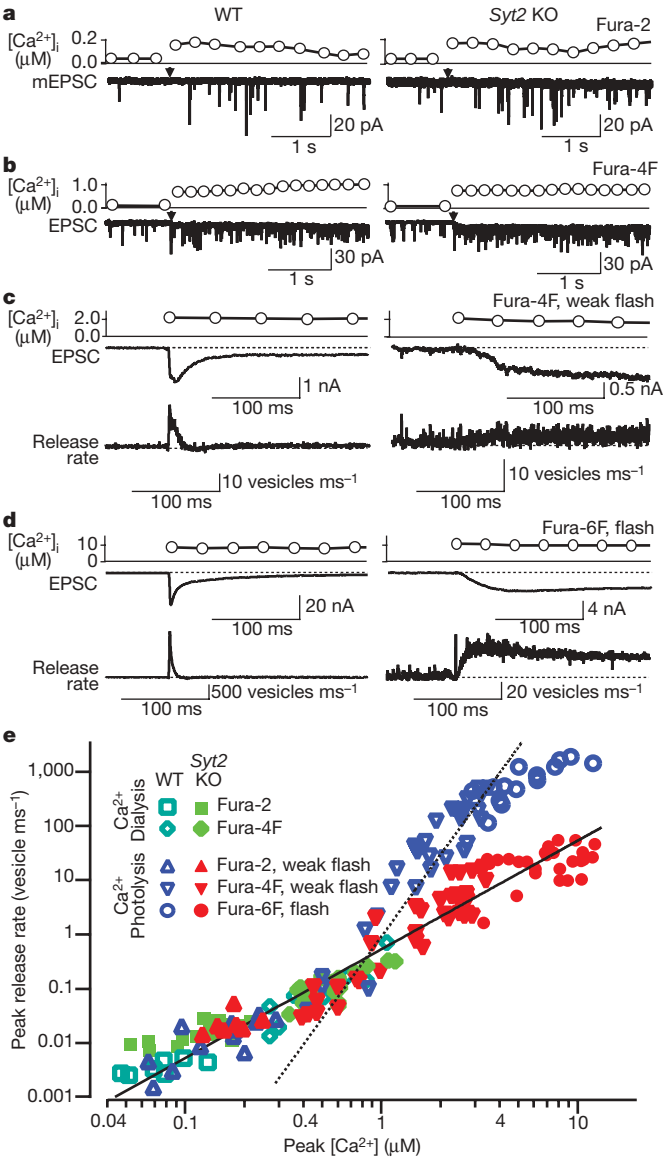


Figure 4 | Relationship between peak vesicular release rates and $[Ca^{2+}]_i$ in calyx terminals. EPSCs were recorded in calyx synapses in the double-patch configuration at P7–P9 in the presence of 50 μM D-AP5 in all experiments, and of 0.1 mM cyclothiazide in experiments in which release rates were quantified by deconvolution (for example, panels **c** and **d**), and of 2 mM γ-DGG in experiments in which $[Ca^{2+}]_i$ was raised to >3 μM (for example, panel **d**). Presynaptic $[Ca^{2+}]_i$ increases were achieved by photolysis of caged Ca^{2+} dialysed into the terminal, or by dialysis of buffered Ca^{2+} at low $[Ca^{2+}]$. $[Ca^{2+}]_i$ was monitored optically using three different Ca^{2+}-sensitive dyes as indicated (for *in situ* calibration of $[Ca^{2+}]_i$, see Supplementary Materials). **a**–**d**, Representative experiments in wild-type (left panels) and Syt2-deficient calyces (right panels) at four characteristic Ca^{2+} concentrations; the Ca^{2+}-indicator dyes used are shown on the right. Panels **c** and **d** also display the vesicle release rate as calculated by EPSC deconvolutions. **e**, Summary graph of peak release rates and $[Ca^{2+}]_i$ ($n = 78$ for wild-type (open symbols), 106 for Syt2-deficient terminals (filled symbols); green symbols represent the data obtained by dialysis of Ca^{2+} buffers into the terminals via the patch pipette). The dashed line represents a fit of a fifth power function to the data from wild-type terminals at $[Ca^{2+}]_i = >1$ μM ($y = 0.90x^5$); the solid line represents the fit of a second power function to the data from mutant terminals ($y = 0.54x^2$); note that the solid line also fits the wild-type responses at low $[Ca^{2+}]_i$.

©2007 Nature Publishing Group

one-sensor model')[11,12]. These models, however, were unable to predict release at low $[Ca^{2+}]_i$ (Fig. 5a), presumably because this release—as depicted in Figs 2 and 4—is primarily carried by the asynchronous Ca^{2+} sensor. Partly to remedy this problem, and partly to account for the change in release induced by phorbol esters, an allosteric model of release was proposed[25] that improves prediction of the Ca^{2+} dependence of peak release rates (Fig. 5a). The allosteric model, however, also postulates a single Ca^{2+} sensor, and underestimates the time-to-peak rate at lower $[Ca^{2+}]_i$ (Fig. 5b). To formulate a more accurate model, we developed a dual-Ca^{2+}-sensor kinetic model that is based on the current information (Fig. 5c). This model postulates that (1) synchronous release is triggered by Ca^{2+} binding to a Ca^{2+} sensor with a Ca^{2+} cooperativity of ~5, consistent with the Ca^{2+}-binding properties of synaptotagmins; (2) asynchronous release is triggered by Ca^{2+} binding to an unidentified Ca^{2+} sensor with a Ca^{2+} cooperativity of ~2 (Fig. 4); and (3) spontaneous release occurs with an effectivity factor C to account for the change in spontaneous release in synaptotagmin-deficient synapses[17,19] (Supplementary Fig. 6). The dual-Ca^{2+}-sensor model assumes that these three release pathways compete with each other, and operate on the same vesicle pools.

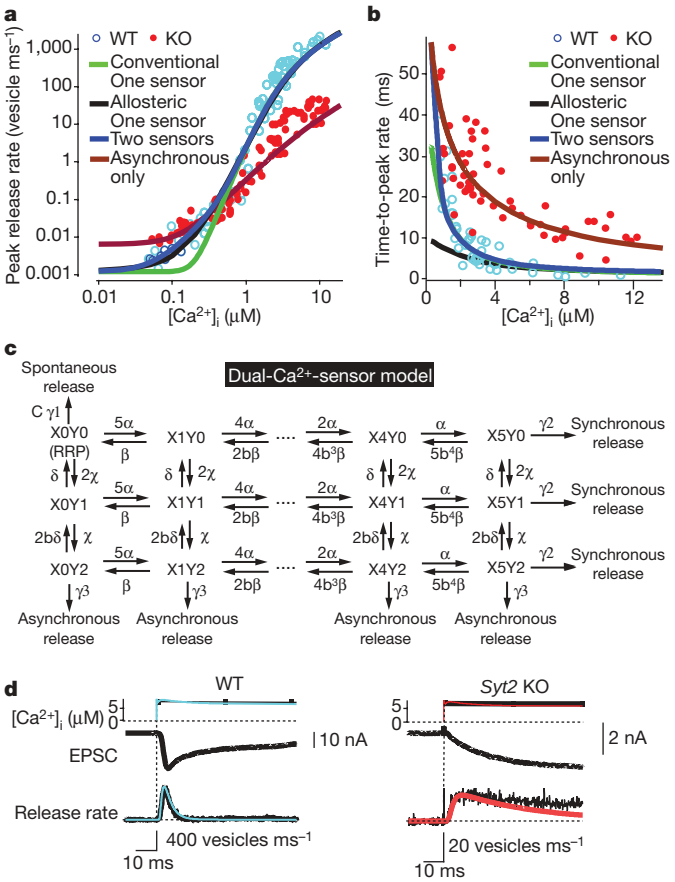


Figure 5 | A dual-Ca^{2+}-sensor model for neurotransmitter release. a, b, Fits of the Ca^{2+}-photolysis data from wild-type and Syt2-deficient calyces for the peak release rate (**a**) and time-to-peak rate (**b**) to the conventional 1-sensor model[11,12], the allosteric 1-sensor model[25], the 2-sensor model, and the 2-sensor model with inactivation of the synchronous Ca^{2+} sensor (see Supplementary Materials and Supplementary Figs 10–13 for details; RRP is set to 3,000 vesicles[11,12]). **c,** Reaction scheme. $\gamma1$, $\gamma2$, $\gamma3$, rates of spontaneous, synchronous and asynchronous release, respectively (defined as the fraction of the RRP released per second); α and β, and χ and $\delta = k_{on}$ and k_{off} for Ca^{2+} action for synchronous and asynchronous release, respectively; X0–X5 and Y0–Y2, Ca^{2+}-binding states of the Ca^{2+} sensor for synchronous (X) and asynchronous release (Y), respectively (note that X0Y0 = RRP); b, cooperativity factor[11,12,34]. The curve-fitting parameters were (only $\gamma1$ differs between wild-type and knockout): $\alpha = 1.53 \times 10^8\ M^{-1}s^{-1}$, $\beta = 5{,}800\ s^{-1}$, b = 0.25; $\chi = 2.94 \times 10^6\ M^{-1}s^{-1}$, $\delta = 130\ s^{-1}$, $\gamma1 = 0.417 \times 10^{-3}\ s^{-1}$ in wild type and $2.23 \times 10^{-3}\ s^{-1}$ in knockout; $\gamma2 = 6{,}000\ s^{-1}$; $\gamma3 = 6{,}000\ s^{-1}$. $\gamma1$ was measured experimentally, b and $\gamma2$ were from ref. 11, and $\gamma3$ was assumed to be equal to $\gamma2$ on the basis of the postulate that Ca^{2+} binding to the asynchronous and synchronous release Ca^{2+} sensors will trigger the same release rates because both empty the entire RRP. **d,** Local $[Ca^{2+}]_i$ signal predicted by the Ca^{2+}-relaxation model (Supplementary Fig. 12) and transmitter release rates predicted by the dual-Ca^{2+}-sensor model in wild-type (left) and *Syt2*-knockout calyces (right). Uppermost panels, measured and predicted $[Ca^{2+}]_i$ (black, and blue or red, respectively); middle panels, recorded EPSCs; lowest panels, vesicular release rates deconvoluted from EPSCs (black), and predicted by the model (blue or red).

The dual-Ca^{2+}-sensor model allows a satisfactory description of all of our Ca^{2+} photolysis data, both of the Ca^{2+} dependence of the peak release rates and of the time-to-peak rate for wild-type and Syt2-deficient synapses (Fig. 5a, b). Moreover, the dual-Ca^{2+}-sensor model accurately predicts the kinetics of Ca^{2+}-dependent vesicle release in wild-type and Syt2-deficient calyx synapses (Fig. 5d, e, and Supplementary Fig. 13). The dual-Ca^{2+}-sensor model calculates that synchronous release exhibits an apparent Ca^{2+} affinity of ~38 μM, a Ca^{2+} cooperativity of ~5, and an apparent k_{on} rate of $\alpha = 1.53\ \ 10^8\ M^{-1}s^{-1}$, and that asynchronous release exhibits a similar apparent Ca^{2+} affinity of ~44 μM, but a much lower Ca^{2+} cooperativity of ~2, and a ~50-fold slower k_{on} rate of $\chi = 2.94\ 10^6\ M^{-1}s^{-1}$ (Fig. 5).

Our parameters for synchronous release are squarely in the middle between the two previous estimates in calyx synapses[11,12], but our parameters for asynchronous release differ from previous suggestions[20,21], probably because previous suggestions were based on extrapolations of asynchronous release in the continued presence of synchronous release, which may have contaminated the estimates. Although the dual-Ca^{2+}-sensor model thus reveals that asynchronous release exhibits a relatively low apparent Ca^{2+} affinity and release rate, the model still predicts that asynchronous release is the major physiological mode of synaptic transmission at low $[Ca^{2+}]_i$ because its lower degree of Ca^{2+} cooperativity renders asynchronous release more effective at low $[Ca^{2+}]_i$. At high $[Ca^{2+}]_i$, conversely, synchronous release dominates because the predicted k_{on} for synchronous release is ~100-fold higher than that for asynchronous release (Fig. 5).

Summary

Here we show that among synaptotagmin isoforms that act as Ca^{2+} sensors for synchronous release[18], calyx terminals only express Syt2 (Fig. 1, and Supplementary Fig. 2). We demonstrate that in calyx terminals, Syt2 is essential for Ca^{2+} triggering of fast synchronous release (Fig. 2), but not for Ca^{2+} influx or vesicle priming (Fig. 3). Thus, Syt2—as suggested for Syt1 in forebrain synapses[13]—is selectively required for Ca^{2+} triggering of fast release in calyx-of-Held synapses. We show that although asynchronous release contributes little to action-potential-induced vesicle exocytosis in wild-type calyx synapses at physiological $[Ca^{2+}]_i$, asynchronous release triggers exocytosis of the entire RRP on prolonged increases in $[Ca^{2+}]_i$ in *Syt2*-knockout calyx synapses (Fig. 3). Therefore, synchronous and asynchronous release act on the same vesicle pools. Moreover, release at low $[Ca^{2+}]_i$ exhibits a similar magnitude and Ca^{2+} dependence in wild-type and Syt2-deficient synapses, suggesting that release at low $[Ca^{2+}]_i$ is normally asynchronous, and that asynchronous release does not change in the knockout mice.

Traditionally, asynchronous release is explained by three competing hypotheses: (1) asynchronous and synchronous release share the same Ca^{2+} sensor, but differ in the coupling of vesicles to Ca^{2+} channels, the state of the vesicles, and/or the Ca^{2+}-buffering mechanisms involved[24,25,36,37]; (2) synchronous and asynchronous release are mediated by the same Ca^{2+} sensor but use different, allosterically regulated vesicle pools[38]; or (3) different Ca^{2+} sensors with distinct properties mediate synchronous and asynchronous release[13]. Clearly the distance between synaptic vesicles and Ca^{2+} channels is a major determinant of the vesicular release probability, and

©2007 Nature Publishing Group

differences in the Ca^{2+} channel proximity of vesicles probably contribute to the vesicles' heterogeneity of release probabilities[39]. Moreover, undoubtedly, different pools of vesicles exist and contribute to the heterogeneity of release probabilities at a synapse. Nevertheless, these differences are probably unrelated to the differences between synchronous and asynchronous release, but are rather involved in regulating all release. Indeed, the following findings suggest that synchronous and asynchronous release are caused by independent mechanisms.

- Synchronous and asynchronous release exhibit qualitatively different properties, as shown by the unexpectedly low Ca^{2+} affinity and distinctly low Ca^{2+} cooperativity of asynchronous release, and by its persistence after synchronous release was selectively eliminated on deletion of Syt2 (Fig. 4).
- Asynchronous release is independent of the proximity of synaptic vesicles to Ca^{2+} channels, because our flash photolysis experiments produce a uniform increase in $[Ca^{2+}]_i$ that bypasses Ca^{2+}-channel activation (Fig. 4).
- Single-Ca^{2+}-sensor models[11,12] adequately describe release at higher $[Ca^{2+}]_i$ but are unable to account for release at low $[Ca^{2+}]_i$. Even the sophisticated allosteric model[25] underestimates the time-to-peak release rate at low $[Ca^{2+}]_i$ (Fig. 5b).
- Release in wild-type and Syt2-deficient terminals is identical at low $[Ca^{2+}]_i$ (Figs 2 and 4), suggesting that this type of release is physiologically asynchronous. Moreover, asynchronous release becomes apparent during high-frequency action-potential trains in wild-type calyx synapses (Supplementary Fig. 14), consistent with studies of the precise kinetics of individual EPSCs demonstrating that asynchronous release in the calyx normally occurs even during isolated action protentials[40].

The definition of the biophysical properties of asynchronous, Ca^{2+}-triggered release made it possible to formulate a quantitative model for neurotransmitter release that probably applies to all synapses (Fig. 5). Our model indicates that the Ca^{2+} sensors for synchronous and asynchronous release operate in competition with each other, with the asynchronous Ca^{2+} sensor being slower but able to bind Ca^{2+} at lower concentrations, whereas the synchronous Ca^{2+} sensor is faster with a higher Ca^{2+} cooperativity. As a result, in this competition, the synchronous Ca^{2+} sensor 'wins' during pulses of high Ca^{2+} concentrations, whereas the asynchronous Ca^{2+} sensor prevails during sustained phases of lower Ca^{2+} concentrations (Fig. 5). Within this framework, differences between synapses are primarily determined by which synaptotagmin isoform is being used as the Ca^{2+} sensor for synchronous release[18], and by the accumulation of residual Ca^{2+} in the intervals between action potentials during stimulus trains (that is, by the Ca^{2+}-buffering properties of a nerve terminal)[34–36,41]. It is of interest here that, fittingly, the calyx terminals with their exquisitely fast-release properties use as the synchronous Ca^{2+} sensor only Syt2, the fastest of the three Ca^{2+} sensors[18].

METHODS SUMMARY

Syt2-knockout mice were bred and genotyped as described[17]. Brain slices (200 μm) containing MNTB were prepared in a parasagital orientation from P7–P9 mice (for double-patch or presynaptic cell-attached recording), or in a transverse orientation from P10–P14 mice (for fibre stimulation and purely postsynaptic recordings), and employed for single and/or double whole-cell recordings of nerve terminals and MNTB neurons with presynaptic electrical stimulation or stimulation by Ca^{2+} uncaging or Ca^{2+} dialysis, largely as described[11,12,31,38,42]. Ca^{2+} uncaging was achieved with an intense ultraviolet radiation pulse from a frequency-tripled YAG-ND laser, and the resulting Ca^{2+} concentrations were measured *in situ* by ratiometric fluorescence imaging of nerve terminals filled with Fura-2, Fura-4F, or Fura-6F (Supplementary Fig. 10)[34,43]. Ca^{2+} dyes were calibrated *in situ*. Release rates were calculated using the Neher deconvolution program (http://www.mpibpc.mpg.de/groups/neher/software/index.html) with a miniature EPSC size of 30 pA and a measured waveform[31]. We applied the different kinetic models to fit the data in our experiments (Fig. 5a, b, and Supplementary Figs 11 and 13), with the conventional and allosteric single-Ca^{2+}-sensor kinetic models simulated as described[11,12,25] (Supplementary Fig. 13). In our dual-Ca^{2+}-sensor model, each vesicle in the RRP can be released via three independent pathways (Fig. 5): (1) Ca^{2+}-independent fusion in the spontaneous mode, that is, direct exocytosis of vesicles from the RRP with a release rate of $\gamma 1$. An effectivity factor (C) to account for the change in spontaneous release rate in Syt1- and Syt2-deficient synapses was included; (2) synchronous Ca^{2+}-evoked fusion mode triggered by full occupancy of the five binding sites of the synchronous release Ca^{2+} sensor[12,25]; or (3) asynchronous Ca^{2+}-evoked release triggered by occupancy of two Ca^{2+}-binding sites of an unidentified Ca^{2+} sensor.

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

Received 20 August; accepted 24 September 2007.

1. Meinrenken, C. J., Borst, J. G. & Sakmann, B. Local routes revisited: the space and time dependence of the Ca^{2+} signal for phasic transmitter release at the rat calyx of Held. *J. Physiol. (Lond.)* **547**, 665–689 (2003).
2. Schneggenburger, R. & Neher, E. Presynaptic calcium and control of vesicle fusion. *Curr. Opin. Neurobiol.* **15**, 266–274 (2005).
3. Atluri, P. P. & Regehr, W. G. Delayed release of neurotransmitter from cerebellar granule cells. *J. Neurosci.* **18**, 8214–8227 (1998).
4. Lu, T. & Trussell, L. O. Inhibitory transmission mediated by asynchronous transmitter release. *Neuron* **26**, 683–694 (2000).
5. Hagler, D. J. Jr & Goda, Y. Properties of synchronous and asynchronous release during pulse train depression in cultured hippocampal neurons. *J. Neurophysiol.* **85**, 2324–2334 (2001).
6. Otsu, Y. *et al.* Competition between phasic and asynchronous release for recovered synaptic vesicles at developing hippocampal autaptic synapses. *J. Neurosci.* **24**, 420–433 (2004).
7. Hefft, S. & Jonas, P. Asynchronous GABA release generates long-lasting inhibition at a hippocampal interneuron–principal neuron synapse. *Nature Neurosci.* **8**, 1319–1328 (2005).
8. Forsythe, I. D. Direct patch recording from identified presynaptic terminals mediating glutamatergic EPSCs in the rat CNS, *in vitro*. *J. Physiol. (Lond.)* **479**, 381–387 (1994).
9. Borst, J. G. & Sakmann, B. Calcium influx and transmitter release in a fast CNS synapse. *Nature* **383**, 431–434 (1996).
10. Sun, J. Y. & Wu, L. G. Fast kinetics of exocytosis revealed by simultaneous measurements of presynaptic capacitance and postsynaptic currents at a central synapse. *Neuron* **30**, 171–182 (2001).
11. Schneggenburger, R. & Neher, E. Intracellular calcium dependence of transmitter release rates at a fast central synapse. *Nature* **406**, 889–893 (2000).
12. Bollmann, J. H., Sakmann, B. & Borst, J. G. Calcium sensitivity of glutamate release in a calyx-type terminal. *Science* **289**, 953–957 (2000).
13. Geppert, M. *et al.* Synaptotagmin I: a major Ca^{2+} sensor for transmitter release at a central synapse. *Cell* **79**, 717–727 (1994).
14. Fernandez-Chacon, R. *et al.* Synaptotagmin I functions as a Ca^{2+}-regulator of release probability. *Nature* **410**, 41–49 (2001).
15. Stevens, C. F. & Sullivan, J. M. The synaptotagmin C_2A domain is part of the calcium sensor controlling fast synaptic transmission. *Neuron* **39**, 299–308 (2003).
16. Nagy, G. *et al.* Different effects on fast exocytosis induced by synaptotagmin 1 and 2 isoforms and abundance, but not by phosphorylation. *J. Neurosci.* **26**, 632–643 (2006).
17. Pang, Z. P. *et al.* Synaptotagmin-2 is essential for survival and contributes to Ca^{2+} triggering of neurotransmitter release in central and neuromuscular synapses. *J. Neurosci.* **26**, 13493–13504 (2006).
18. Xu, J., Mashimo, T. & Südhof, T. C. Synaptotagmin-1, -2, and -9: Ca^{2+}-sensors for fast release that specify distinct presynaptic properties in subsets of neurons. *Neuron* **54**, 801–812 (2007).
19. Pang, Z. P., Sun, J., Rizo, J., Maximov, A. & Südhof, T. C. Genetic analysis of Syt2 in spontaneous and Ca^{2+}-triggered neurotransmitter release. *EMBO J.* **25**, 2039–2050 (2006).
20. Goda, Y. & Stevens, C. F. Two components of transmitter release at a central synapse. *Proc. Natl Acad. Sci. USA* **91**, 12942–12946 (1994).
21. Ravin, R., Spira, M. E., Parnas, H. & Parnas, I. Simultaneous measurement of intracellular Ca^{2+} and asynchronous transmitter release from the same crayfish bouton. *J. Physiol. (Lond.)* **501**, 251–262 (1997).
22. Maximov, A. & Südhof, T. C. Autonomous function of synaptotagmin 1 in triggering asynchronous release independent of asynchronous release. *Neuron* **48**, 547–554 (2005).
23. Chuhma, N. & Ohmori, H. Role of Ca^{2+} in the synchronization of transmitter release at calyceal synapses in the auditory system of rat. *J. Neurophysiol.* **87**, 222–228 (2002).
24. Schneggenburger, R. & Forsythe, I. D. The calyx of Held. *Cell Tissue Res.* **326**, 311–337 (2006).
25. Lou, X., Scheuss, V. & Schneggenburger, R. Allosteric modulation of the presynaptic Ca^{2+} sensor for vesicle fusion. *Nature* **435**, 497–501 (2005).
26. Iwasaki, S. & Takahashi, T. Developmental changes in calcium channel types mediating synaptic transmission in rat auditory brainstem. *J. Physiol. (Lond.)* **509**, 419–423 (1998).

©2007 Nature Publishing Group

27. Wu, L. G., Westenbroek, R. E., Borst, J. G., Catterall, W. A. & Sakmann, B. Calcium channel types with distinct presynaptic localization couple differentially to transmitter release in single calyx-type synapses. *J. Neurosci.* **19,** 726–736 (1999).
28. Leveque, C. *et al.* Purification of the N-type calcium channel associated with syntaxin and synaptotagmin. A complex implicated in synaptic vesicle exocytosis. *J. Biol. Chem.* **269,** 6306–6312 (1994).
29. Charvin, N. *et al.* Direct interaction of the calcium sensor protein synaptotagmin I with a cytoplasmic domain of the α1A subunit of the P/Q-type calcium channel. *EMBO J.* **16,** 4591–4596 (1997).
30. Zhong, H., Yokoyama, C. T., Scheuer, T. & Catterall, W. A. Reciprocal regulation of P/Q-type Ca^{2+} channels by SNAP-25, syntaxin and synaptotagmin. *Nature Neurosci.* **2,** 939–941 (1999).
31. Sakaba, T. & Neher, E. Quantitative relationship between transmitter release and calcium current at the calyx of Held synapse. *J. Neurosci.* **21,** 462–476 (2001).
32. Stevens, C. F. & Wesseling, J. F. Activity-dependent modulation of the rate at which synaptic vesicles become available to undergo exocytosis. *Neuron* **21,** 415–424 (1998).
33. Mulkey, R. M. & Zucker, R. S. Action potentials must admit calcium to evoke transmitter release. *Nature* **350,** 153–155 (1991).
34. Heidelberger, R., Heinemann, C., Neher, E. & Matthews, G. Calcium dependence of the rate of exocytosis in a synaptic terminal. *Nature* **371,** 513–515 (1994).
35. Awatramani, G. B., Price, G. D. & Trussell, L. O. Modulation of transmitter release by presynaptic resting potential and background calcium levels. *Neuron* **48,** 109–121 (2005).
36. Taschenberger, H., Scheuss, V. & Neher, E. Release kinetics, quantal parameters and their modulation during short-term depression at a developing synapse in the rat CNS. *J. Physiol. (Lond.)* **568,** 513–537 (2005).
37. Wadel, K., Neher, E. & Sakaba, T. The coupling between synaptic vesicles and Ca^{2+} channels determines fast neurontransmitter release. *Neuron* **53,** 563–575 (2007).
38. Wölfel, M., Lou, X. & Schneggenburger, R. A mechanism intrinsic to the vesicle fusion machinery determines fast and slow transmitter release at a large CAN synapse. *J. Neurosci.* **27,** 3198–3210 (2007).
39. Trommershauser, J., Schneggenburger, R., Zippelius, A. & Neher, E. Heterogeneous presynaptic release probabilities: functional relevance for short-term plasticity. *Biophys. J.* **84,** 1563–1579 (2003).
40. Kushmerick, C., Renden, R. & von Gersdorff, H. Physiological temperatures reduce the rate of vesicle pool depletion and short-term depression via an acceleration of vesicle recruitment. *J. Neurosci.* **26,** 1366–1377 (2006).
41. Bollmann, J. H. & Sakmann, B. Control of synaptic strength and timing by the release-site Ca^{2+} signal. *Nature Neurosci.* **8,** 426–434 (2005).
42. Sun, J., Bronk, P., Liu, X., Han, W. & Südhof, T. C. Synapsins regulate use-dependent synaptic plasticity in the calyx of Held by a Ca^{2+}/calmodulin-dependent pathway. *Proc. Natl Acad. Sci. USA* **103,** 2880–2885 (2006).
43. Grynkiewicz, G., Poenie, M. & Tsien, R. Y. A new generation of Ca^{2+} indicators with greatly improved fluorescence properties. *J. Biol. Chem.* **260,** 3440–3450 (1985).

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

Acknowledgements We thank I. Kornblum, A. Roth, L. Fan and J. Mitchell for excellent technical assistance, and J. Bollmann, X. Lou and R. Schneggenburger for advice. This study was supported by a grant from NARSAD (to J.S.) and by The University of Texas M. D. Anderson Cancer Center Physician Scientist Program (R.A.).

Author Contributions J.S. performed the electrophysiology and photolysis experiments and modelling. Z.P.P. carried out the biochemical, immunohistochemical, and mouse genetics experiments; D.Q. participated in the electrophysiology and photolysis experiments; A.T. F. and R.A. generated the *Syt2* knockout mice, and T.C.S. and J.S. designed the experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.S. (Jianyuan.Sun@UTSouthwestern.edu) or T.C.S. (Thomas.Sudhof@UTSouthwestern.edu).

©2007 Nature Publishing Group

METHODS

***Syt2*-knockout mice.** Syt2-deficient mice were bred and genotyped as described[17] (see Supplementary Materials for detailed methods). All analyses were performed on littermate offspring from heterozygous matings ('wild-type mice' refers to those that are homo- or heterozygous for the wild-type allele).

Slice electrophysiology. All experiments involved postsynaptic whole-cell recordings with an Axopatch 200B amplifier (Axon Instruments). Presynaptic whole-cell recordings were obtained with an EPC-9 amplifier (HEKA). The pre- and postsynaptic series resistances (<15 MΩ and 7 MΩ) were compensated by 60% and 98% (lag 10 μs), respectively. Both pre- and postsynaptic currents were low-pass-filtered at 5 kHz and digitized at 20 kHz. Six recording configurations were employed (see Supplementary Materials for details): (1) Presynaptic cell-attached current injections to induce presynaptic action-potential trains with postsynaptic whole-cell recordings that monitor the evoked EPSCs (Fig. 1c). Stimulations were applied as 40 presynaptic current injections of 1 nA for 3 ms at 50 Hz. (2) Presynaptic afferent fibre stimulations with postsynaptic whole-cell voltage-clamp recordings (Fig. 2). Stimuli were applied with a bipolar electrode delivering 3–30 V for 0.1 ms. (3) Double-patch recordings by simultaneous pre- and postsynaptic whole-cell voltage-clamp recordings to measure the presynaptic RRP and Ca^{2+} currents (Figs 3a–g). Stimulations consist of a presynaptic 4 ms predepolarization to 70–80 mV, followed by 50 ms depolarization to 20 mV. (4) Sucrose stimulation with postsynaptic whole-cell voltage-clamp recordings to measure the RRP (Fig. 3h, i). Stimulation involves puffing 2 M sucrose in bath solution onto the target terminal with a pipette that is located about 5 μm from the calyx. (5) Double-patch experiments for simultaneous measurements of presynaptic $[Ca^{2+}]_i$ and postsynaptic ESPCs with manipulation of the presynaptic $[Ca^{2+}]_i$ (Fig. 4). Stimulations were effected either by dialysis of Ca^{2+}-containing solutions into the terminal via the presynaptic pipette (Fig. 4e, and Supplementary Fig. 11a, b), or by flash photolysis of DM-nitrophen/Ca^{2+} (Fig. 4, and Supplementary Fig. 11c). Release rate was estimated by deconvolution[31]. (6) Mini recordings (Supplementary Fig. 6 and Supplementary Methods).

All recordings were performed in the presence of 50 μM D-AP5 in the bath; in addition, for the double-patch experiments in (3) and (4), we added 0.1 mM cyclothiazide and 1 mM kynurenic acid or 2 mM γ-DGG when strong flash photolysis was given, resulting in $[Ca^{2+}]_i$ of >3 μM, in which case the obtained EPSCs were multiplied by 2 because control experiments determined that 2 mM γ-DGG decreased the EPSC amplitude twofold.

Ca^{2+} uncaging and Ca^{2+} imaging. The Ca^{2+}-uncaging and Ca^{2+}-imaging setup (Supplementary Fig. 5a) used an intense ultraviolet radiation pulse from a frequency-tripled YAG-ND laser (355 nm, Surelite I, Continuum) for Ca^{2+} uncaging. Ca^{2+} concentrations were measured *in situ* by ratiometric fluorescence imaging of nerve terminals filled with Fura-2, Fura-4F or Fura-6F (refs 34, 43). Ca^{2+}-indicator dyes were excited with an ultraviolet light source at 340 nm and 380 nm (energy ≤175 W) using a monochromator (DG-4, Sutter Instrument). The laser pulse was coupled into the epifluorescence port of an Axioskop and combined with the ultraviolet light using a beam-splitter (customized 90%T/10%R for 355 nm with a bandwith of <10 nm, Chroma Tech). Both ultraviolet beams were collimated to optimize the intensity on the targeted terminal. A charge-coupled device camera (ORCA-ER) with on-chip binning was used to capture infrared images (300 × 300 pixels) and Ca^{2+} images (19 × 19 pixels) of the terminal (Supplmentary Fig. 9b, c). The fluorescence in the measuring area with background fluorescence subtraction (off-line) was used to calculate the $[Ca^{2+}]_i$. Images were captured using MetaFluor software and analysed by IgorPro (Wavemetrics). For *in vivo* calibration of Ca^{2+}-indicator signals, we introduced Ca^{2+} indicators with an intracellular K-gluconate pipette solution into the terminal (see Supplementary Materials). For Fura-2 imaging, we used exposure times of 30 ms (in photolysis experiments) and 100 ms (for defined $[Ca^{2+}]$ solution injection experiment) with a 2 Hz capture rate. For Fura-4F and Fura-6F imaging, we used 10 ms exposure times with 2 Hz capture rates before the flash, and 10–30 Hz capture rates after the flash. Ca^{2+} relaxation rates were modulated by the ultraviolet-light-illumination during the ratiometric Ca^{2+}-imaging procedure, which was thus adjusted to maintain stable Ca^{2+} levels (see Supplementary Materials).

Data processing and modelling. Release rates were calculated using the Neher deconvolution program (http://www.mpibpc.mpg.de/groups/neher/software/index.html) with a mEPSC size of 30 pA and a measured waveform[31].

Modelling. We applied different kinetic models to fit the data in our experiments (Fig. 5a, b). The conventional one-Ca^{2+}-sensor kinetic model and the allosteric one-Ca^{2+}-sensor kinetic model were simulated as described[11,12,25] (see Supplementary Fig. 13 legend). In our dual-Ca^{2+}-sensor model, each vesicle in the RRP can be released via three independent pathways: (1) Ca^{2+}-independent fusion in the spontaneous mode, that is, direct exocytosis of vesicles from the RRP with a release rate of γ1. An effectivity factor (C) to account for the change in spontaneous release rate in Syt1 and Syt2-deficient synapses was included. (2) Synchronous Ca^{2+}-evoked fusion mode triggered by full occupancy of the five binding sites of the synchronous release Ca^{2+} sensor[12,25]. (3) Asynchronous Ca^{2+}-evoked release triggered by occupancy of two Ca^{2+}-binding sites of an unidentified Ca^{2+} sensor. The Ca^{2+}-binding states can be defined by X*n*Y*m*(*t*), where X*n* represents the state in which *n* binding sites of the Ca^{2+} sensor for synchronous release have been occupied ($n = 0$–5), and Y*m* the state in which *m* binding sites of the Ca^{2+} sensor of asynchronous release have been occupied ($m = 0$–2). α and β represent the binding and dissociation constants, respectively, of the Ca^{2+} sensor for synchronous release; χ and δ the binding constants for asynchronous release, respectively, b is the cooperativity factor, and *t* is the time[34]. Note that X0Y0(0) = RRP. The kinetics of Ca^{2+}-binding states can be described as (Fig. 5c):

when $0 < n < 5, 0 < m < 2$,

$$\mathrm{d}(\mathrm{X}n\mathrm{Y}m)/\mathrm{dt} = \alpha \times (5-n+1) \times \mathrm{X}n-1\mathrm{Y}m \times [\mathrm{Ca}^{2+}]_\mathrm{i} + \beta \times \mathrm{b}^n \times (n+1) \times \mathrm{X}n+1\mathrm{Y}m - \alpha \times (5-n) \times \mathrm{X}n\mathrm{Y}m \times [\mathrm{Ca}^{2+}]_\mathrm{i} - \beta \times \mathrm{b}^{n-1} \times n \times \mathrm{X}n\mathrm{Y}m + \chi \times (2-m+1) \times \mathrm{X}n\mathrm{Y}m-1 \times [\mathrm{Ca}^{2+}]_\mathrm{i} + \delta \times \mathrm{b}^m \times (m+1) \times \mathrm{X}n\mathrm{Y}m+1 - \chi \times (2-m) \times \mathrm{X}n\mathrm{Y}m \times [\mathrm{Ca}^{2+}]_\mathrm{i} - \delta \times \mathrm{b}^{m-1} \times m \times \mathrm{X}n\mathrm{Y}m$$

when $n = 0$, $m = 0$,

$$\mathrm{d}(\mathrm{X0Y0})/\mathrm{dt} = \beta \times \mathrm{X1Y0} - 5 \times \alpha \times \mathrm{X0Y0} \times [\mathrm{Ca}^{2+}]_\mathrm{i} + \delta \times \mathrm{X0Y1} - 2 \times \chi \times \mathrm{X0Y0} \times [\mathrm{Ca}^{2+}]_\mathrm{i} - \gamma 1 \times \mathrm{X0Y0}$$

when $n = 5$,

$$\mathrm{d}(\mathrm{X}n\mathrm{Y}m)/\mathrm{dt} = \alpha \times (5-n+1) \times \mathrm{X}n-1\mathrm{Y}m \times [\mathrm{Ca}^{2+}]_\mathrm{i} + \beta \times \mathrm{b}^n \times (n+1) \times \mathrm{X}n+1\mathrm{Y}m - \alpha \times (5-n) \times \mathrm{X}n\mathrm{Y}m \times [\mathrm{Ca}^{2+}]_\mathrm{i} - \beta \times \mathrm{b}^{n-1} \times n \times \mathrm{X}n\mathrm{Y}m + \chi \times (2-m+1) \times \mathrm{X}n\mathrm{Y}m-1 \times [\mathrm{Ca}^{2+}]_\mathrm{i} + \delta \times \mathrm{b}^m \times (m+1) \times \mathrm{X}n\mathrm{Y}m+1 - \chi \times (2-m) \times \mathrm{X}n\mathrm{Y}m \times [\mathrm{Ca}^{2+}]_\mathrm{i} - \delta \times \mathrm{b}^{m-1} \times m \times \mathrm{X}n\mathrm{Y}m - \gamma 2 \times (\mathrm{X}n\mathrm{Y}m)$$

when $m = 2$

$$\mathrm{d}(\mathrm{X}n\mathrm{Y}m)/\mathrm{dt} = \alpha \times (5-n+1) \times \mathrm{X}n-1\mathrm{Y}m \times [\mathrm{Ca}^{2+}]_\mathrm{i} + \beta \times \mathrm{b}^n \times (n+1) \times \mathrm{X}n+1\mathrm{Y}m - \alpha \times (5-n) \times \mathrm{X}n\mathrm{Y}m \times [\mathrm{Ca}^{2+}]_\mathrm{i} - \beta \times \mathrm{b}^{n-1} \times n \times \mathrm{X}n\mathrm{Y}m + \chi \times (2-m+1) \times \mathrm{X}n\mathrm{Y}m-1 \times [\mathrm{Ca}^{2+}]_\mathrm{i} + \delta \times \mathrm{b}^m \times (m+1) \times \mathrm{X}n\mathrm{Y}m+1 - \chi \times (2-m) \times \mathrm{X}n\mathrm{Y}m \times [\mathrm{Ca}^{2+}]_\mathrm{i} - \delta \times \mathrm{b}^{m-1} \times m \times \mathrm{X}n\mathrm{Y}m - \gamma 3 \times (\mathrm{X}n\mathrm{Y}m)$$

The total release within Δ*t*:

$$\mathrm{fuse}(t,\Delta t) = [\gamma 1 \times \mathrm{X0Y0} + \gamma 2 \times (\mathrm{X5Y0} + \mathrm{X5Y1} + \mathrm{X5Y2}) + \gamma 3 \times (\mathrm{X0Y2} + \mathrm{X1Y2} + \mathrm{X2Y2} + \mathrm{X3Y2} + \mathrm{X4Y2} + \mathrm{X5Y2})] \times \Delta t$$

where,

$$\text{spontaneous release} = \gamma 1 \times \mathrm{X0Y0} \times \Delta t$$

$$\text{synchronous release} = \gamma 2 \times (\mathrm{X5Y0} + \mathrm{X5Y1} + \mathrm{X5Y2}) \times \Delta t$$

$$\text{asynchronous release} = \gamma 3 \times (\mathrm{X0Y2} + \mathrm{X1Y2} + \mathrm{X2Y2} + \mathrm{X3Y2} + \mathrm{X4Y2} + \mathrm{X5Y2}) \times \Delta t$$

Miscellaneous. Immunofluorescence labelling and immunoblotting experiments were performed essentially as described[19]. All statistical analyses were performed using Student's *t*-test.

©2007 Nature Publishing Group

ARTICLES

Determining the architectures of macromolecular assemblies

Frank Alber[1]*, Svetlana Dokudovskaya[2]*†, Liesbeth M. Veenhoff[2]*†, Wenzhu Zhang[3], Julia Kipper[2]†, Damien Devos[1]†, Adisetyantari Suprapto[2]†, Orit Karni-Schmidt[2]†, Rosemary Williams[2], Brian T. Chait[3], Michael P. Rout[2] & Andrej Sali[1]

To understand the workings of a living cell, we need to know the architectures of its macromolecular assemblies. Here we show how proteomic data can be used to determine such structures. The process involves the collection of sufficient and diverse high-quality data, translation of these data into spatial restraints, and an optimization that uses the restraints to generate an ensemble of structures consistent with the data. Analysis of the ensemble produces a detailed architectural map of the assembly. We developed our approach on a challenging model system, the nuclear pore complex (NPC). The NPC acts as a dynamic barrier, controlling access to and from the nucleus, and in yeast is a 50 MDa assembly of 456 proteins. The resulting structure, presented in an accompanying paper, reveals the configuration of the proteins in the NPC, providing insights into its evolution and architectural principles. The present approach should be applicable to many other macromolecular assemblies.

A mechanistic understanding of the cell requires the structural characterization of the thousands of its constituent biological assemblies[1]. So far, conventional approaches have provided a valuable but limited window into the structures of these assemblies. For example, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy can resolve the atomic details of individual proteins and small complexes, whereas electron microscopy produces morphological maps but can lack the ability to identify and detail specific components in the map of the whole assembly. As a result, we do not yet have atomic-resolution structures, or even low-resolution representations, for the vast majority of complexes in the cell. How, then, are we to resolve the molecular architectures of these assemblies?

In an attempt to address this problem, we have taken the yeast (*Saccharomyces cerevisiae*) nuclear pore complex (NPC) as a case in point. The NPC is among the largest macromolecular assemblies in the cell, mediating the exchange of molecules that pass between the nuclear and cytoplasmic compartments. Yeast NPCs are ~50 MDa structures built of multiple copies of some 30 different proteins (nucleoporins), totalling at least 456 protein molecules[2]. Each NPC is a plastic structure embedded in the nuclear envelope and is composed of eight morphologically similar 'spokes' surrounding a central tube[3–6]. Filling this tube and projecting into both the cytoplasmic and nuclear sides are flexible filamentous domains from proteins termed FG (phenylalanine-glycine) repeat nucleoporins; these domains form the docking sites for transport factors that carry macromolecular cargoes through the NPC.

The NPC represents a significant challenge for conventional structure determination approaches owing to its large size and the high degree of flexibility of the complex and its components. Thus, although electron microscopy has provided valuable insights into the overall shape of the NPC, its molecular architecture (that is, the spatial configuration of its component proteins) has yet to be revealed, and atomic structures have only been solved for domains covering ~5% of its component protein sequences[7]. The NPC therefore encapsulates many of the obstacles that will be encountered in the detailed structural examination of other macromolecular assemblies.

We describe here a set of proteomics experiments and a computational platform for converting the resulting data into the structures of macromolecular assemblies. Central to this approach is the realization that many kinds of biophysical and proteomic data contain valuable structural information about assemblies.

Overview of integrative structure determination

Our approach to structure determination can be seen as an iterative series of four steps: data generation by experiment, translation of the data into spatial restraints, calculation of an ensemble of structures by satisfaction of these restraints, and an analysis of the ensemble to produce the final structure (Fig. 1). The structure calculation part of this process is expressed as an optimization problem, a solution of which requires three main components: (1) a representation of the assembly in terms of its constituent parts; (2) a scoring function, consisting of individual spatial restraints that encode all the data; and (3) an optimization of the scoring function, which aims to yield structures that satisfy the restraints.

Formally, our approach is similar to the determination of protein structures by NMR spectroscopy, in which the folding of the polypeptide chain is determined by satisfying distance restraints between pairs of atoms[8]. As with NMR spectroscopy, a structure is computationally determined from experimental data. Here, atoms

[1]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, Byers Hall, Suite 503B, 1700 4th Street, University of California at San Francisco, San Francisco, California 94158-2330, USA. [2]Laboratory of Cellular and Structural Biology, and [3]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. †Present addresses: Laboratory of Nucleocytoplasmic Transport, Institut Jacques Monod, 2 place Jussieu, Tour 43, Paris 75251, France (S.D.); Department of Biochemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands (L.M.V.); German Aerospace Center (PT-DLR), Heinrich-Konen-Strasse 1, D-53227 Bonn, Germany (J.K.); Structural Bioinformatics, EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany (D.D.); Office of Technology Transfer, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA (A.S.); Herbert Irving Comprehensive Cancer Center, Columbia University, 1130 St Nicholas Avenue, New York, New York 10032, USA (O.K.-S.).
*These authors contributed equally to this work.

©2007 Nature Publishing Group

are replaced by proteins, and their positions and relative proximities are restrained on the basis of data from a variety of proteomics and other experiments, including affinity purification, ultracentrifugation, electron microscopy and immuno-electron microscopy (immuno-EM).

Data generation. The most important aspect of our approach is its potential to use simultaneously almost any conceivable type of information to determine assembly structures. For example, sedimentation analysis of the isolated proteins can be used to infer their shapes; immuno-EM can give an approximate localization of each protein in the assembly; and affinity purification of tagged proteins and protein complexes can yield information about the arrangement and interactions of proteins within the assembly. These data can be of a kind not normally used for structure determination (for example, complexes identified by affinity purification), can refer to different levels in the structural hierarchy (for example, a protein domain, a whole protein, or a protein complex), and can be ambiguous in terms of their structural interpretation (for example, the uncertainty as to which copy of the protein is involved in an interaction, when multiple copies exist).

The use of such data for structure determination presented us with four major challenges. First, large amounts of suitable data must be collected to give sufficient spatial information to define structures; fortunately, the proteomic revolution has provided methodologies that allow us to garner enough information. Second, much of the data can be of relatively low precision; thus, to avoid over-interpretation, appropriate tolerances must be used in its structural interpretation. Third, the possibility of false-positive data must be minimized and taken into consideration. Fourth, ambiguity of the data in terms of its structural interpretation must be treated when multiple copies of the same protein are present in an assembly and the experiment does not determine which specific instance of a protein is detected. All of these challenges can be addressed by an integrative approach that incorporates information varying greatly in terms of its accuracy and precision; limitations of any particular type of data can be overcome by the use of large and diverse data sets derived from synergistic experimental methods[1,9].

Data translation into spatial restraints. The data can be used to restrain many different features of the assembly, such as the positions of proteins, protein contacts, proximity between proteins, and the shape and symmetry of the whole assembly. A 'restraint' specifies values of the restrained feature that are consistent with the experimental information about it; a perfectly satisfied restraint is indicated here by 0, whereas values larger than 0 correspond to a violated restraint. Thus, a restraint encodes our uncertainty in the restrained feature. In essence, restraints can be thought of as generating a 'force' on each component in the assembly, to mould them into a configuration that satisfies the data used to define the restraints.

Optimization. All the restraints are summed to obtain a scoring function, which determines the degree of consistency between the restrained spatial features in a structure and the experimental information; a perfect structure is indicated by 0, reflecting the summed values of perfectly satisfied restraints, whereas values larger than 0 correspond to a structure that increasingly violates restraints. The scoring function is then optimized to calculate a structure that minimizes violations of the restraints. It is necessary to generate many such structures to provide a good sampling of structures that are consistent with the data (that is, the 'ensemble').

Ensemble analysis. All of the structures that satisfy the input restraints are clustered into distinct sets, on the basis of their similarities. There are three possible outcomes of such clustering. First, if only a single cluster of structures satisfies all the input information,

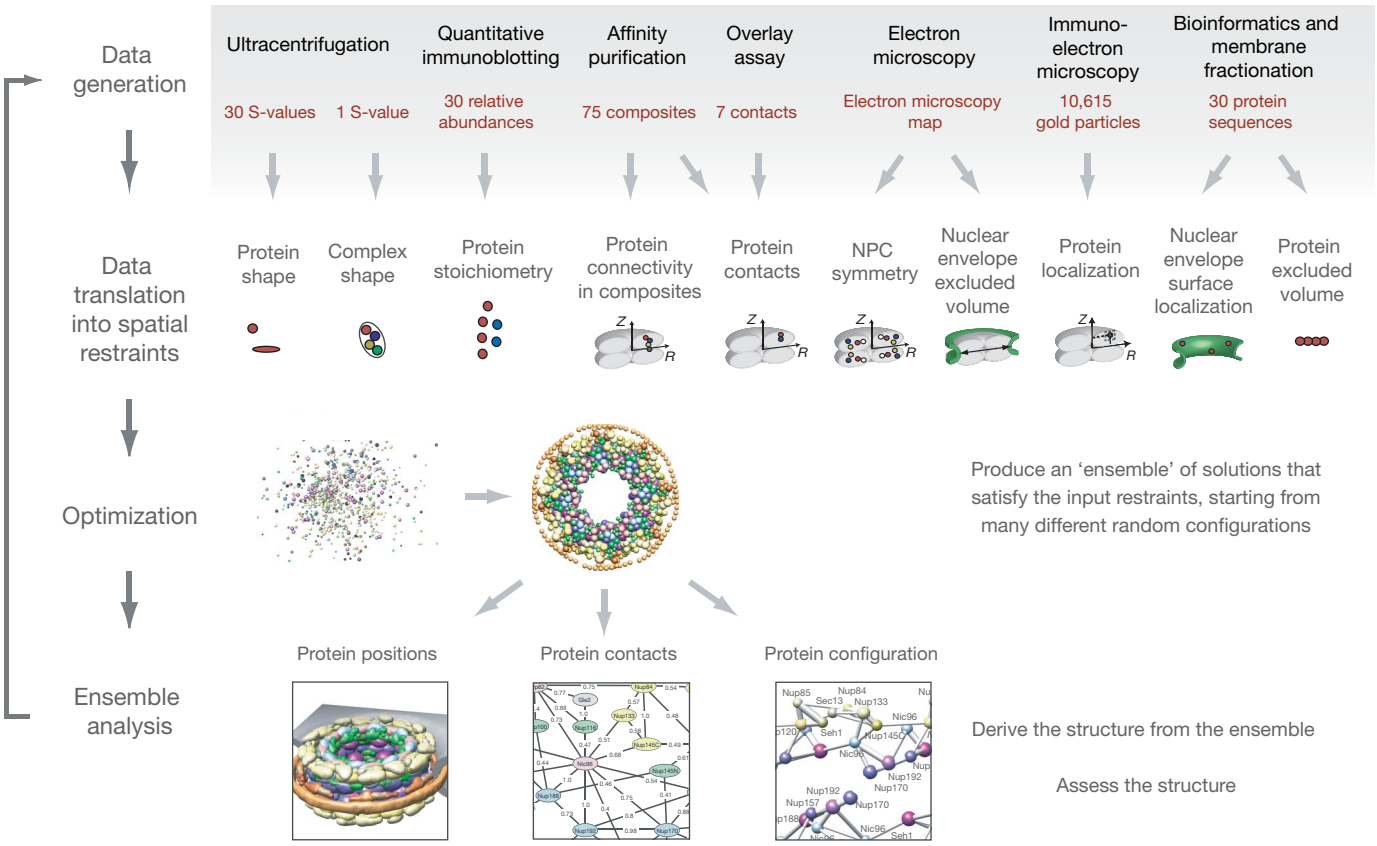


Figure 1 | Determining the architecture of the NPC by integrating spatial restraints from proteomic data. First, structural data (red) are generated by various experiments (black). Second, the data are translated into spatial restraints. Third, an ensemble of structural solutions that satisfy the data are obtained by minimizing the violations of the spatial restraints, starting from many different random configurations. Fourth, the ensemble is clustered into distinct sets of structures on the basis of their similarities, and analysed in terms of protein positions, contacts and configuration.

©2007 Nature Publishing Group

there is probably sufficient data for determining the unique native state. Second, if different clusters are consistent with the input information, either the data are insufficient to define the single native state or there are multiple native structures. If the number of clusters is small, the structural differences between them may suggest additional experiments so as to narrow down the possible solutions. Third, if no structures satisfy all input information, either the data or their interpretation in terms of the restraints is incorrect. Given the first two outcomes, the ensemble can be analysed to determine different aspects of the native state, such as protein positions, contacts and configuration. The variability of the ensemble provides an estimate of the precision of the structure determination.

We illustrate our approach by determining the configuration of the protein components in the NPC from the yeast *S. cerevisiae* (Fig. 1).

Data generation

As no single experimental technique has been sufficient to solve the molecular architecture of the NPC, we used a variety of techniques, each of which gave different and synergistic information about the structure; the techniques were chosen to generate the needed structural information with a defined level of accuracy.

An NPC component list. To determine any structure, we must first define its parts (Fig. 2). In the case of the NPC, we have already determined that some 30 nucleoporins constitute the assembly[2]. Although the exact composition is still uncertain because some proteins interact relatively transiently with the NPC, potential omission of a small fraction of such transient components is unlikely to interfere with structure determination.

The stoichiometry of each component in the NPC. The stoichiometry of each nucleoporin in each half-spoke has been previously established[2]. However, having found the stoichiometry of Nup82 to be ambiguous, we re-examined it with new strains and found that Nup82 is present in two copies per spoke (Fig. 3 and Supplementary Fig. 7).

The shape and size of each component. Next, we must represent the structures of the constituent nucleoporins. Because atomic structures have not yet been solved for most nucleoporins, we estimated their shapes based primarily on their sedimentation coefficients determined by ultracentrifugation of the purified proteins (Fig. 3 and Supplementary Information). The sedimentation behaviour of most FG nucleoporins agrees with their predicted filamentous, native disordered structure[10,11]. Pom152, an integral membrane component, appeared to be a highly elongated structure, consistent with its multiple domains modelled as β-cadherin-like folds[7]. Most of the other nucleoporins appear to have a relatively compact tertiary structure that is again in agreement with their predicted fold assignments[7,12]. The seven-protein Nup84 complex[13] could be separated into two smaller complexes on sedimentation: an elongated tetramer (composite 30, see below) and an elongated hexamer (composite 45, see below), consistent with their elongated appearance when visualized by electron microscopy[14].

The size, shape and symmetry of the NPC. It is also helpful to have some information on the overall shape and symmetry of the NPC. The position of the nuclear envelope membrane relative to the NPC and the NPC's symmetry are based on our electron microscopy and cryo-electron microscopy (cryo-EM) data[5]. These studies have revealed an eight-fold rotational symmetry of the yeast NPC and an approximate two-fold rotational symmetry between the nucleoplasmic and cytosolic halves of the NPC, defining the 'half-spoke' as a 16-fold pseudo-symmetry unit of the NPC (Fig. 2). We have also previously shown that heparin treatment of isolated NPCs produced a ring-like substructure ('Pom rings'), which is associated with the pore membrane and perinuclear space in the intact NPC[15]. We isolated and examined these rings (Supplementary Information), and found that they had a maximum diameter of ~106 nm, consistent with the measured maximum NPC diameter of ~97 nm[5].

The localization of each component in the NPC. We have previously obtained the coarse localization of most nucleoporins within the NPC by immuno-EM, relying on a gold-labelled antibody that specifically interacted with the localized protein through its carboxy-terminal PrA tag (Fig. 4a)[2]. We have now generated a more accurate and complete immunolocalization map of the NPC, in which its constituent proteins, except Sec13, have been localized using a larger data set and improved analysis (Fig. 4b and Supplementary Information).

Inherent limitations in the immuno-EM method allow it to provide only a broad range of allowed axial and radial values for each nucleoporin. Nevertheless, these ranges are smaller than the dimensions of the half-spoke and so are still informative. Notably, most nucleoporins are found on both the nuclear and cytoplasmic sides of the NPC and are tightly packed within a region adjacent to the nuclear membrane (Fig. 4). Most of the FG nucleoporins are found on both sides of the NPC, with a small number found exclusively on the cytoplasmic or nuclear side; for simplicity, we consider Nup116 and Nup100 to be cytoplasmically disposed and Nup145N to be nucleoplasmically disposed, although ~20% of the signal of each is found on the opposite side. Most of the non-FG nucleoporins are also found on both sides. The membrane proteins are found close to the nuclear envelope membrane, and Pom152–PrA is localized to the lumen of the nuclear envelope. Our immuno-EM map agrees almost entirely with independent localizations performed by other groups. For example, Nup159 and Nup82 have previously been shown to be restricted to the peripheral cytoplasmic face[16]; Nup1 was found on the peripheral nuclear face[17]; and Nup157, Nup170, Nup53 and Nup59 were shown to localize proximally to both sides of the

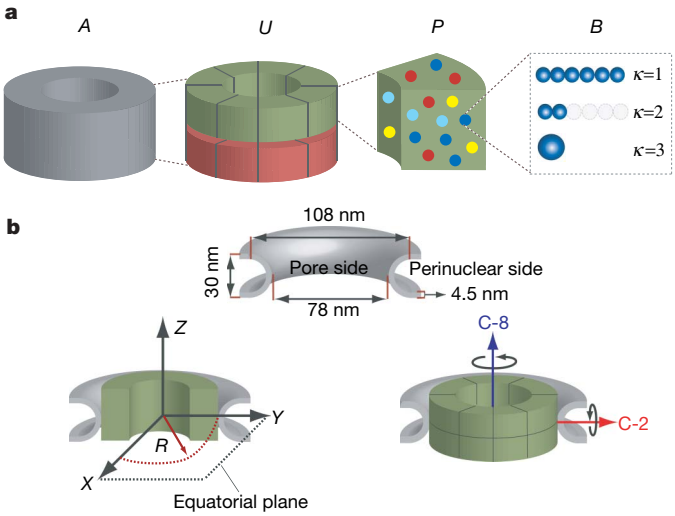


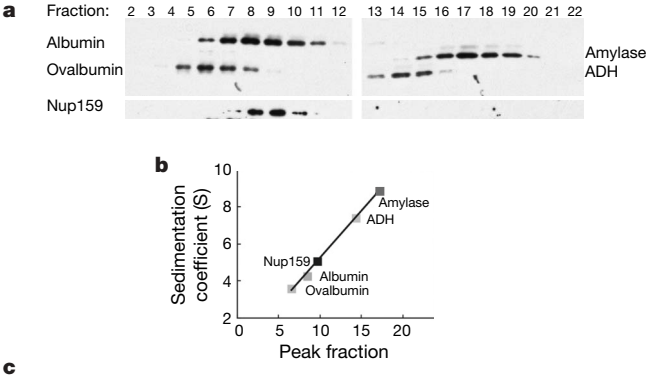
Figure 2 | Structural representation of the NPC. a, Hierarchical representation of the NPC that facilitates the expression of the experimental data in terms of spatial restraints. Formally, we define the whole NPC assembly *A* as a set of symmetry units *U* of two different types with eight instances each, referred to as half-spokes. Half-spokes of the first type (green) reside at the cytoplasmic side and half-spokes of the second type (red) reside at the nucleoplasmic side of the nuclear envelope. Two adjacent half-spokes, one of each type, form a spoke. Each of the 16 NPC half-spokes consists of a set of proteins *P* that are described by their type and index. Each protein is represented by a flexible string of beads *B* in the root representation $\kappa = 1$. Additional representations $\kappa > 1$ can be derived from the root representation (for example, by omitting some beads as in $\kappa = 2$ or by combining beads as in $\kappa = 3$). For the NPC, each protein is described with up to nine different representations. **b**, Top panel: the dimensions of the nuclear envelope, as taken from cryo-EM images (ref. 5). Bottom-left panel: the coordinate system we use has the origin at the centre of the nuclear envelope pore. The nuclear envelope is indicated in grey. Bottom-right panel: the eight-fold (C-8) and two-fold (C-2) symmetry axes of the NPC, as revealed primarily by cryo-EM[5]. We apply the two-fold symmetry only to proteins that appear with identical stoichiometry in both the nucleoplasmic and cytoplasmic half-spokes.

©2007 Nature Publishing Group

NPC[18] (other independent localizations are listed in Supplementary Table 10).

How the NPC components fit together. The coarse shape, approximate position and stoichiometry of each nucleoporin are not enough to build an accurate picture of the NPC: rather like the pieces in a jigsaw puzzle, we also need information on the interactions between nucleoporins. We obtained this information from a large number of overlay assays and affinity purification experiments, as well as from the composition of the Pom rings (consisting of Pom34 and Pom152). An overlay assay identifies a pair of proteins that interact with each other, whereas an affinity purification identifies one or more proteins that interact directly or indirectly with the bait protein (Figs 5 and 6 and Supplementary Information). An affinity purification produces a distinctive set of co-isolating proteins, which we term a composite. A composite may represent a single complex of physically interacting proteins or a mixture of such complexes overlapping at least at the tagged protein. We only used overlay and affinity purification data with a signal-to-noise ratio above a demanding threshold (Supplementary Information).

We designed several affinity purification methods to obtain a large and diverse set of composites (Supplementary Information). PrA was used as a high-affinity C-terminal purification tag on each nucleoporin. Different cell fractions from the tagged strains served as starting materials, although most fractions were produced by whole-cell cryolysis, which proved to be rapid and convenient, yielding high amounts of each complex with minimal losses and proteolytic damage. We generated ~20 variants of extraction buffers with diverse properties to release different kinds of complexes from the fractions. Complexes were isolated via the tagged nucleoporins using antibody conjugated to either Sepharose or magnetic beads, although we preferred magnetic beads as it permitted rapid, high-yield isolations, and eliminated an upper size limit on the purified complexes (Supplementary Information). We also performed affinity purifications from diploid compared with haploid strains to detect a potential second, untagged copy of a given nucleoporin in the complex—a strong indication of a homotypic interaction for that nucleoporin; Pom152–PrA and Nup82–PrA were the only two nucleoporins giving composites containing a second untagged copy. Although



c

Protein	Molecular mass (kDa) (protein +PrA)	S_{obs}	S_{max}/S_{obs}	Stoichiometry (cyt.)	Stoichiometry (nucl.)	Bead number	Bead representation	Bead radius (nm)
Nup192	217	9.1	1.4	1	1	2		3.0
Nup188	214	9.3	1.4	1	1	2		3.0
Nup170	195	9.0	1.4	1	1	2		2.9
Nup159	185	5.1	2.3	1	0	11		1.6
Nup157	183	7.1	1.6	1	1	3		2.5
Pom152	178	-	-	1	1	10		1.6
Pom152*	155	4.8	2.2					
Nup133	159	7.6	1.4	1	1	2		2.7
Nup120	146	8.0	1.3	1	1	2		2.6
Nup116*	100	3.6	2.2	-	-	13		1.3
Nup1	140	-	-	0	1	9		1.5
Nup1*	120	3.7	2.4					
Nup100	126	4.4	2.0	1	0	13		1.3
Nic96	122	6.3	1.4	2	2	2		2.4
Nsp1	112	3.5	2.4	2	2	12		1.3
Nup85/Seh1	150	6.8	1.5	1	1	3		2.0
Nup85	111	-	-	1	1			
Nup84	110	4.9	1.7	1	1	3		2.0
Nup82	108	5.3	1.5	2	0	2		2.3
Nup145C	107	6.7	1.2	1	1	2		2.3
Ndc1	100	-	-	1	1	2		2.2
Gle1	88	5.9	1.2	1	0	2		2.1
Nup60	85	3.8	1.8	0	1	4		1.6
Nup59	85	4.2	1.7	1	1	4		1.6
Nup57	83	4.1	1.7	1	1	3		1.8
Nup53	79	4.1	1.6	1	1	3		1.7
Nup145N	86	3.7	1.9	0	2	6		1.5
Nup49	75	3.9	1.6	1	1	3		1.7
Nup42	69	3.0	2.0	1	0	5		1.4
Gle2	66	4.6	1.3	1	1	1		2.3
Seh1	65	3.4	1.7	1	1	1		2.2
Pom34	60	-	-	1	1	3		1.5
Sec13	59	4.2	1.3	1	1	1		2.1
Complex 30	357	6.7	1.2	-	-	-	-	-
Complex 45	468	10	2.2	-	-	-	-	-

Figure 3 | Protein shape and stoichiometry information. **a**, Protein shape from hydrodynamic experiments. Purified native PrA-tagged nucleoporins were sedimented on sucrose gradients, together with a set of biotin-labelled marker proteins. Fractions were collected and analysed by immunoblotting of the biotin and PrA tags. An immunoblot of fractions from a typical sedimentation analysis is shown, indicating the position of the tagged protein (Nup159–PrA) together with the markers ovalbumin (3.6 S), bovine serum albumin (4.3 S), alcohol dehydrogenase (ADH, 7.4 S) and β-amylase (8.9 S). **b**, Peak positions for the sedimenting proteins were determined and linear regression was used to calibrate the sedimentation coefficients of the PrA-tagged nucleoporin. **c**, Bead representations $\kappa = 1$ of the NPC proteins and their stoichiometries per half-spoke. The stoichiometry of a protein in the cytoplasmic (cyt.) and nucleoplasmic (nucl.) half-spoke, as measured by quantitative immunoblotting[2], is shown. S_{max} values were calculated based on the molecular mass (kDa) of each protein; $S_{max}/S_{obs} < 1.4$ indicates a globular protein; 1.6–1.9, moderately elongated; >2, highly elongated[45]. An asterisk indicates that C-terminal fragments were measured. Also shown is a visualization of the protein as a flexible bead chain (shown here in its most extended configuration), which is based on sedimentation analysis, identification of domains by sequence comparison and secondary structure prediction.

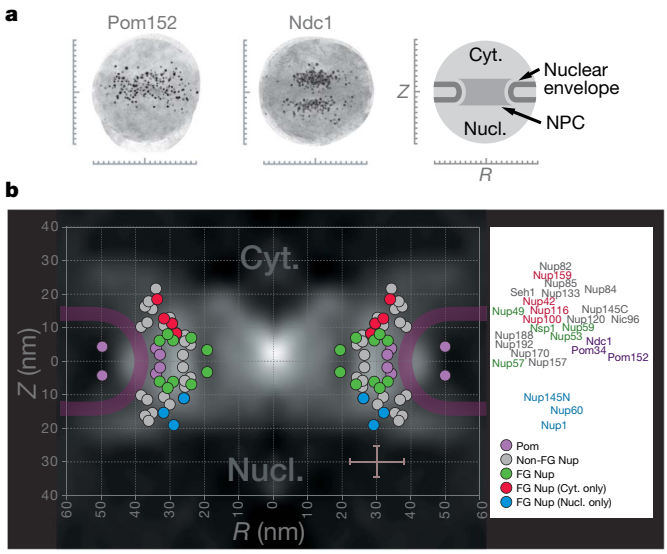


Figure 4 | Localization of proteins by immuno-EM. **a**, Immuno-EM montages for Pom152–PrA nuclei and Ndc1–PrA nuclear envelopes. Scale bars are graduated in 10-nm intervals using the coordinate system defined in Fig. 2b. The major features in each montage are shown schematically at the right, showing how the position of every gold particle in each montage was measured from both the central *Z*-axis of the NPC (*R*) and from the equatorial plane of the nuclear envelope (*Z*). **b**, Estimated position of the C terminus of each protein in the NPC relative to the central *Z*-axis of the NPC (*R*) and the equatorial plane (*Z*) superimposed on the protein density map of a cross-section of the yeast NPC obtained by cryo-EM[5]. The average allowed ranges along the *R* and *Z* coordinates (±8 nm and ±4.5 nm, respectively) are indicated by the brown bars in the bottom right corner.

©2007 Nature Publishing Group

we originally designed our approaches for the purification of NPC complexes, they have proved to be useful for the isolation of many types of complexes from different cells[7,12,19–24].

Identification of proteins was performed by mass spectrometry[25,26]. Generally, the most vicinal associates of the tagged protein should be approaching stochiometric amounts in the purified complexes; conversely, distally associating proteins may be less abundant. By concentrating on only Coomassie-stainable SDS–polyacrylamide gel electrophoresis (PAGE) bands, we ensured that we identified only the more abundant proteins in any given affinity purification and avoided trace residuals (Fig. 5a). Polypeptides below ~20 kDa were excluded from this analysis for technical reasons[27]; however, due to their small volume, their exclusion is not likely to significantly affect structure determination.

Affinity purifications of tagged versions of all yeast nucleoporins, as well as the NPC-associated messenger RNA transport factors Gle1 and Gle2 (refs 28, 29), yielded 73 distinct composites; together with overlay assays and Pom ring data, we have defined a total of 82 composites (Fig. 6a and Supplementary Information). The composites varied in complexity from dimers to those containing 20 proteins (composite 82) and, importantly, shared significant overlap in composition (Fig. 6b). Therefore, we expect considerable synergy among the composites when used to map the architecture of the whole assembly.

A good example of the compositional overlap is the Nup84 complex (Fig. 5a, b)[13,14,30]. The smallest building blocks of this complex are heterodimers (Fig. 5, composites 7, 14, 15). Under different isolation conditions, these dimers can be purified with an increasing number of additional proteins, such as trimers (25, 20), a tetramer (33), a pentamer (39), hexamers (44, 45, 51), and the full septameric Nup84 complex (53, 54, 57). This full complex interacts with Nup157 (63, 66) and Nup145N (60). Finally, the entire Nup84 complex co-precipitates together with the Nup170 complex and an Nsp1-containing complex (79). Our data also agree with composites generated by other groups. For example, the Nup84 composites[13,14,30], a

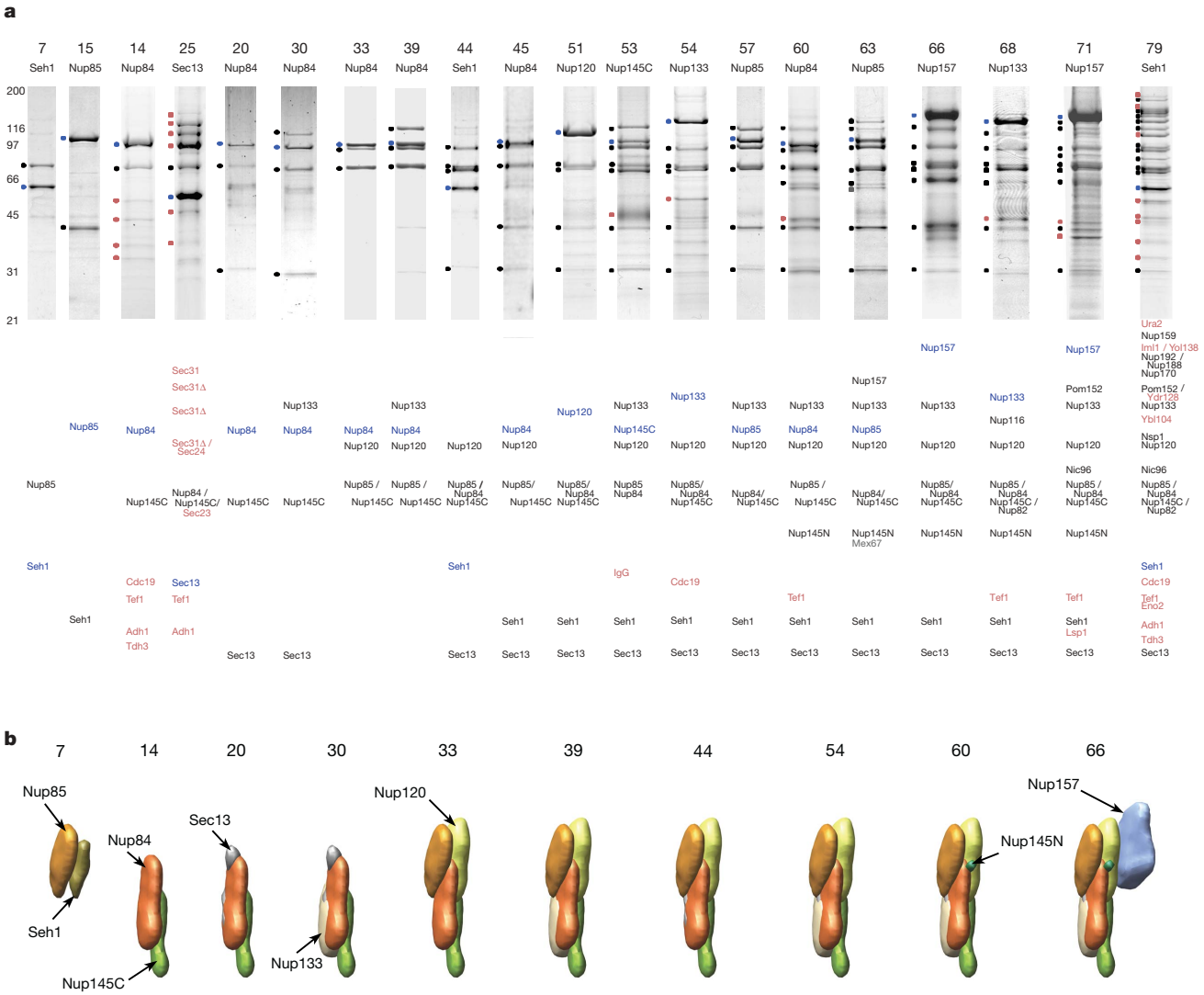


Figure 5 | Protein interactions of the Nup84 complex. **a**, A sample of affinity purifications containing Nup84 complex proteins. Affinity-purified PrA-tagged proteins and interacting proteins were resolved by SDS–PAGE and visualized with Coomassie blue. The name of the PrA-tagged protein together with a corresponding identification number for the composite is indicated above each lane (Supplementary Information). Molecular mass standards (kDa) are indicated to the left of the panel. The bands marked by filled circles at the left of the gel lanes were identified by mass spectrometry (either of the example shown here or of a parallel version; Supplementary Information). The identity of the co-purifying proteins is indicated in order below each lane; PrA-tagged proteins are indicated in blue, co-purifying nucleoporins in black, NPC-associated proteins in grey, and other proteins (including contaminants) in red. Each individual gel image was differentially scaled along its length so that its molecular mass standards aligned to a single reference set of molecular mass standards, and contrast-adjusted to improve visibility. **b**, The mutual arrangement of the Nup84-complex-associated proteins as visualized by their localization volumes. The localization volumes, obtained from the final NPC structure (Fig. 9), allow a visual interpretation of the relative proximities of the proteins.

©2007 Nature Publishing Group

Nup116 composite[31], a Nup170 composite[18], a Nup42–Gle1 dimer[29], a Nic96 composite[32] and others (Supplementary Table 9) have been previously described, and are completely consistent with the composites identified here.

Data translation into spatial restraints

The next step is to translate the experimental data about the NPC structure into spatial restraints (Fig. 1). These restraints were numerous, overlapping and varied in type, and thus were expected to be sufficient for defining the architecture of the NPC.

Restraints and the scoring function. Structure determination is enabled by expressing information as a scoring function, the global optimum of which corresponds to the structure of the native assembly[33]. One such function is a joint probability density function (PDF) of protein positions, given the available information I about the system, $p(C/I)$, where $C = (c_1, c_2, \ldots, c_n)$ is the list of the cartesian coordinates (c_i) of the n component proteins in the assembly (that is, the configuration of the proteins). This joint PDF gives the probability density that a component i of the native configuration is positioned very close to c_i, given the information I we wish to consider in the calculation. In general, I may include any structural information from experiments, physical theories, or statistical preferences. The complete joint PDF is generally unknown, but can be approximated as a product of PDFs p_f that describe individual assembly features (for example, distances or relative orientations of proteins):

$$p(C/I) = \prod_f p_f(C/I_f)$$

The scoring function $F(C)$ is then defined as the logarithm of the joint PDF:

$$F(C) = -\ln \prod_f p_f(C/I_f) = \sum_f r_f(C)$$

For convenience, we refer to the logarithm of a feature PDF as a restraint r_f and the scoring function is therefore the sum of the individual restraints.

Setting up the representation of the NPC. To define restraints on the components of an assembly, we must first specify the symmetry unit of the assembly (that is, the half-spoke in the case of the NPC) (Fig. 2a) and the stoichiometry of its components (Fig. 3). In addition, we must define the representations of the components. Each nucleoporin was represented by a flexible chain consisting of a small number of connected beads (Figs 2a and 3). The number and radii of the beads were chosen to reproduce the protein masses and the sedimentation coefficients[34]. The flexibility of the representation and the low granularity of the NPC structure are sufficient to accommodate uncertainties in the measured S-values and their interpretation. For the FG nucleoporins, no restraints other than the chain connectivity and excluded volume were imposed on the beads representing the FG-repeat regions.

Given the symmetry unit and the protein representations, we can formally represent the NPC with a four-level hierarchy corresponding to the whole NPC, the half-spokes, proteins and beads representing each protein (Fig. 2a). In addition, the nuclear envelope was represented as a rigid surface of many small beads, providing a mould in which the NPC forms (Fig. 2b).

Symmetry of the NPC. The eight-fold and approximate two-fold rotational symmetries of the NPC (Fig. 2b) were imposed by requiring essentially identical configurations of the proteins in common within each half-spoke; the corresponding restraint is formally the root-mean-square of the differences between equivalent intra-half-spoke distances. Although any individual NPC assembly may be perturbed from this perfect symmetry at any given point in time, restraints on the symmetry are nevertheless justified by the relatively low-resolution structure reported here, our intent to characterize the average structure, and exclusion of the FG-repeat regions from the symmetry restraints.

Protein positions from immuno-EM. To reflect the uncertainty in the immuno-EM data, we do not restrain a protein to a specific position. Instead, the C-terminal bead of each protein, corresponding to the tag position, was restrained by imposing lower and upper harmonic bounds on its Z and R coordinates (Fig. 2b), corresponding to the ranges allowed by the immuno-EM data. On average, the allowed area spans 16 and 9 nm along the R and Z coordinate, respectively (Fig. 4 Supplementary Tables 2 and 7, and Supplementary Fig. 8). With such large allowed ranges, the immuno-EM data provide little more information to the structure calculation than which side of the nuclear envelope each nucleoporin is on, and whether it is close to or distal from the NPC equatorial plane and the NPC axis.

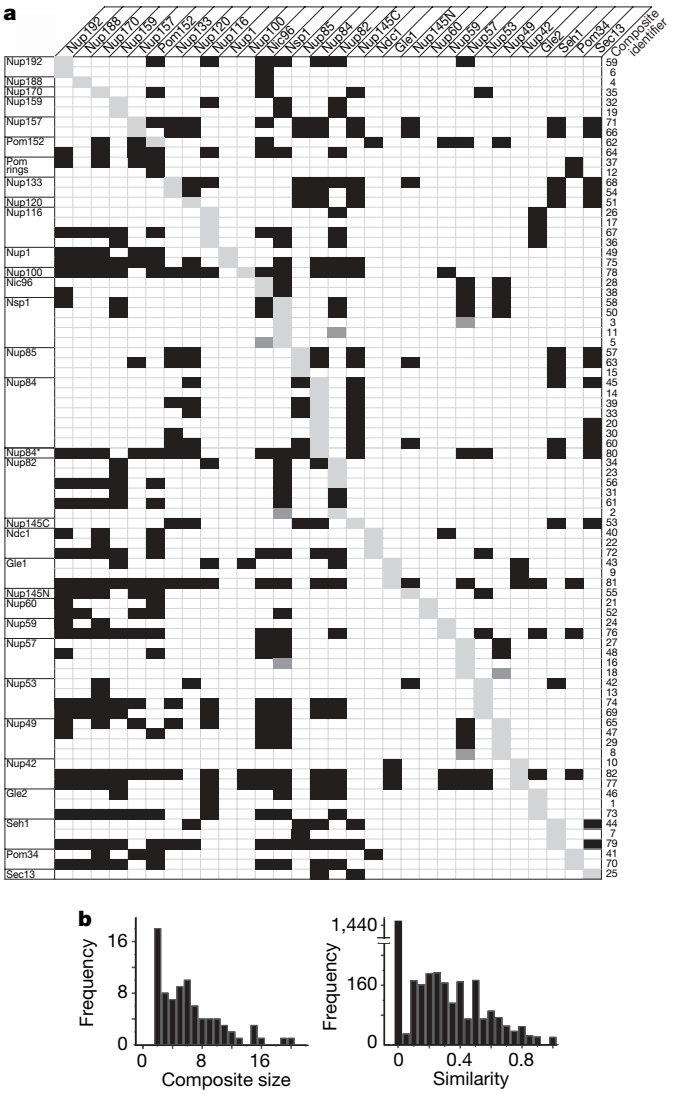


Figure 6 | Protein proximity by affinity purification. a, Composites determined by affinity purification. The affinity-purified nucleoporin–PrA is indicated on the vertical axis, and the corresponding nucleoporins in each composite are shown on the horizontal axis. Composite identifiers are indicated to the right. Presence of a nucleoporin in a composite is indicated by a black box, and the tagged nucleoporin is indicated by a light grey box. In composite 64 (Pom152) and in composites 31 and 61 (Nup82), a second untagged copy of a corresponding protein is present, indicated by a black box. A direct interaction determined by overlay assay is indicated by a dark grey box. The asterisk for Nup84 indicates that the data were obtained with GFP-tagged Nup84. **b**, Distributions of composite size (left) and composite similarity (right). The similarity between two composites is defined by $2a/(2a + b + c)$, where a is the number of proteins that occur in both composites, b is the number of proteins present only in the first composite, and c is the number of proteins present only in the second composite.

©2007 Nature Publishing Group

Protein positions using the nuclear envelope as a mould. The transmembrane-spanning helices of the three membrane proteins Pom152, Ndc1 and Pom34 were predicted by the program TMHH[35]. The corresponding beads were then restrained to the surface of the nuclear envelope by harmonic positional restraints. In addition, the terminal regions of each protein were restrained either to the pore or perinuclear sides of the nuclear envelope, on the basis of the immuno-EM data and the number of predicted transmembrane helices[2].

Protein proximities from overlay assays and affinity purifications. The overlay assays and affinity purifications carry information about protein proximities, and so are encoded by the same type of spatial restraint. These data provide the richest set of restraints for our NPC structure.

To interpret each composite in terms of a spatial restraint, we must consider three ambiguities. First, there is an ambiguity as to what contacts are present in a composite when it contains more than two proteins. A composite implies only that a copy of each protein in the composite must directly interact with at least one copy of another protein in the composite; any structure that satisfies this condition is consistent with the observed composite. In other words, a composite of n proteins implies at least $n-1$ such interactions between proteins of all types in the composite. Thus, each allowed combination of protein interactions corresponds to a 'spanning tree' of a 'composite graph' (as explained in Fig. 7b). Second, when there are multiple copies of the same protein in the assembly, there is an ambiguity as to which copy is involved in a given type of interaction (Fig. 7a). A measured interaction implies only that at least one copy of the protein is involved in that interaction. Third, when multiple beads are used to represent a protein, there is an ambiguity as to which bead is involved in the interaction (Fig. 2a). A measured interaction implies only that at least one bead of the protein is involved in that interaction. As a result of these three ambiguities, we need to encode a composite by a 'conditional restraint', ensuring that all allowed combinations of alternative assignments of interacting bead pairs are considered (Fig. 7b). Finding the assignment of interactions to specific beads that satisfies the data becomes part of the optimization process (see below). Other minor restraints were also derived from

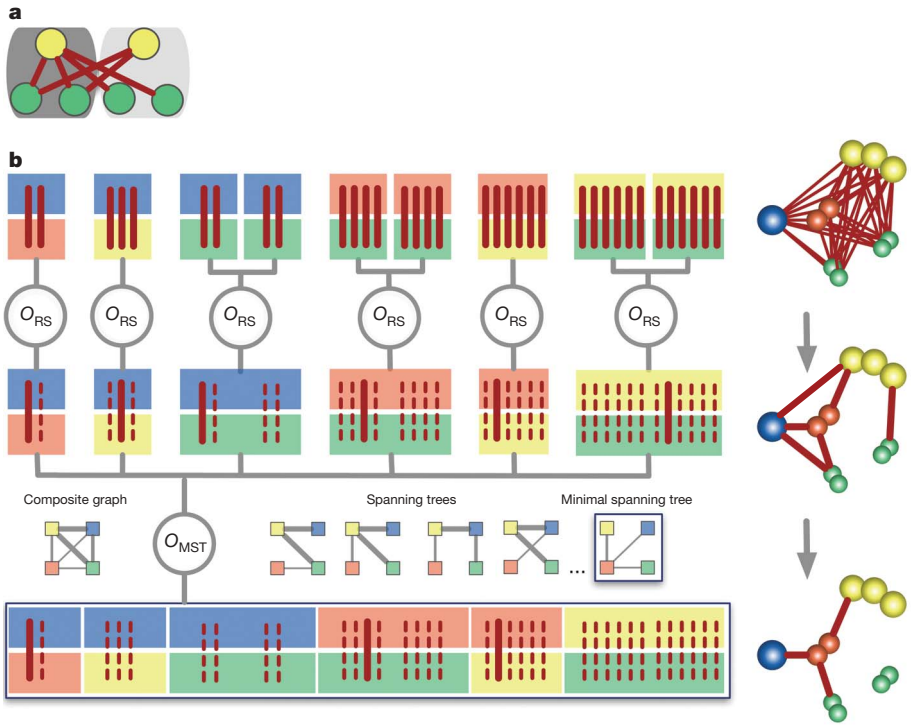


Figure 7 | Ambiguity in data interpretation and conditional restraints. **a**, The ambiguity for a protein interaction between proteins of green and yellow types is illustrated. The ambiguity results from the presence of multiple copies of the same protein in the same or neighbouring symmetry unit. In our NPC calculations, both neighbouring half-spokes on the cytoplasmic and nucleoplasmic sides are considered, for a total of four neighbouring half-spokes (not shown). **b**, The conditional restraint is illustrated by an example of a composite of four protein types (yellow, blue, red, green), derived from an assembly containing a single copy of the yellow, blue, and red protein and two copies of the green protein; proteins are represented by a single bead (blue protein), a pair of beads (green and red proteins), and a string of three beads (yellow protein) (right panel). This composite implies that at least three of the following six possible types of interaction must occur: blue–red, blue–yellow, blue–green, red–green, red–yellow and yellow–green. In addition, (1) the three selected interactions must form a 'spanning tree' of the 'composite graph' (defined below); (2) each type of interaction can involve either copy of the green protein (in general, all alternatives must be considered as illustrated in **a**); and (3) each protein can interact through any of its beads. These considerations can be encoded through a tree-like evaluation of the conditional restraint. At the top level, all optional bead–bead interactions between all protein copies are clustered by protein types. Each alternative bead interaction is restrained by a harmonic upper bound on the distance between the beads; these are 'optional restraints', because only a subset is selected for contribution to the final value of the conditional restraint. Next, a 'rank-and-select' operator (O_{RS}) selects only the least violated optional restraint from each interaction type, resulting in six restraints (thick red line) at the middle level of the tree. Finally, the minimal spanning tree operator (O_{MST}) finds the combination of three restraints that are most consistent with the composite data (thick red line); here the edge weights in the minimal spanning tree (defined below) correspond to the restraint values given the current assembly structure. The column on the right shows a structural interpretation of the composite with proteins represented by their coloured beads and alternative interactions indicated by edges between them. The composite graph (shown on the left) is a fully connected graph that consists of nodes for all identified protein types and edges for all pairwise interactions between protein types; in the context of the conditional restraint, the edge weights correspond to the restraint values. Five of the sixteen possible spanning trees are also shown. A spanning tree is a graph with the smallest possible number of edges that connect all nodes. The minimal spanning tree is the spanning tree with the minimal sum of edge weights. This restraint evaluation process is executed at each optimization step based on the current configuration, thus resulting in possibly different subsets of selected optional restraints at each step.

©2007 Nature Publishing Group

the overlay assay and affinity purification data (Supplementary Information).

Optimization

With the scoring function in hand, the positions of the proteins are determined by optimization of the scoring function (Supplementary Information), resulting in structures that are consistent with the data (Fig. 1). The optimization starts with a random configuration of the constituent proteins' beads, and then iteratively moves them so as to minimize violations of the restraints (Fig. 8). In essence, the restraints cooperate to slowly 'pull together' the proteins into a good-scoring configuration. We use standard methods of conjugate gradients and molecular dynamics with simulated annealing (Supplementary Information). These methods allow the evolving structure some 'breathing room' to explore the scoring function landscape, minimizing the likelihood of getting caught in local scoring function minima (Fig. 8a). To comprehensively sample structures consistent with the data, independent optimizations of randomly generated initial configurations were performed until an ensemble of 1,000 structures satisfying the input restraints was obtained (approximately 200,000 trials were required, running for approximately 30 days on 200 CPUs) (Fig. 8b).

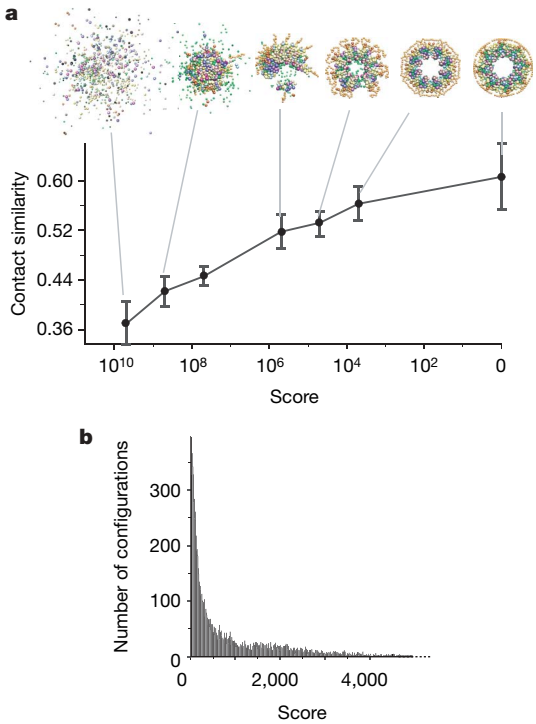


Figure 8 | Calculation of the NPC bead structure by satisfaction of spatial restraints. a, Representation of the optimization process as it progresses from an initial random configuration to an optimal structure. The graph shows the relationship between the score (a measure of the consistency between the configuration and the input data) and the average contact similarity. The contact similarity quantifies how similar two configurations are in terms of the number and types of their protein contacts; a contact between two proteins occurs if the distance between their closest beads is less than 1.4 times the sum of the bead radii (Supplementary Information). The average contact similarity at a given score is determined from the contact similarities between the lowest scoring configuration and a sample of 100 configurations with the given score. Error bars indicate standard deviation. Representative configurations at various stages of the optimization process from left (very large scores) to right (with a score of 0) are shown above the graph; a score of 0 indicates that all input restraints have been satisfied. As the score approaches zero, the contact similarity increases, showing that there is only a single cluster of closely related configurations that satisfy the input data. **b**, Distribution of configuration scores. The presence of configurations with the score close to 0 demonstrates that our sampling procedure finds configurations consistent with the input data. These configurations satisfy all the input restraints within the experimental error.

Ensemble interpretation

We analysed the ensemble of 1,000 structures that satisfy the input data (Fig. 8b) in terms of protein positions, contacts and configuration (Figs 9 and 10).

Protein positions. These 1,000 structures were first superposed (Fig. 9a) (Supplementary Information). Next, the superposed structures were converted into the probability of any volume element being occupied by a given protein (that is, the 'localization probability') (Fig. 9b). The spread around the maximum localization probability of each protein describes how precisely its position was defined by the input data. The positions that have a single narrow maximum in their probability distribution in the ensemble are determined most precisely. When multiple maxima are present in the distribution at the precision of interest, the input restraints are insufficient to define the single native state of that protein (or there are multiple native states).

The actual localization probabilities yielded single pronounced maxima for almost all proteins, demonstrating that the input restraints define one predominant structure. The average standard deviation for the distance between neighbouring protein centroids is 5 nm; the precision of the larger, centrally positioned proteins seems to be higher than that of the anchor domains of some FG nucleoporins. This level of precision defines a region smaller than the diameters of many nucleoporins. Thus, our map is sufficient to determine the relative positions of proteins in the NPC; we do not interpret features smaller than this precision. On the basis of the localization probabilities (Fig. 9b), we also define the volume most likely occupied by each protein, termed the 'localization volume' (Figs 9c and 10a). The localization volumes of the proteins overlap only to a small degree, such that only 10% of the NPC volume is assigned to two or more proteins, again underscoring how well the position of each nucleoporin is resolved. On the basis of our current data, we are not able to distinguish between the two possible mirror-symmetric structures; here, we present one of them.

Protein contacts. The proximities of any two proteins in the structure can be measured by their relative 'contact frequency', which is defined by how often the two proteins contact each other in the ensemble (Fig. 10b). Contacts are highly conserved among the ensemble structures, despite some variability; 32 protein pairs have a contact frequency higher than 65%. Of all the 435 contact frequencies, 7% are high (65–100%) and 73% are low (0–25%); this again demonstrates that the structure is well defined, as an ensemble of varied structures would yield mainly medium contact frequencies. Notably, few high-contact frequencies are seen between proteins of the same type, indicating that the NPC is held together primarily by heterotypic interactions.

We can improve our determination of contacts by considering not only the contact frequencies but also the composite data (Fig. 10c). More specifically, we define two proteins to be 'adjacent' if their relative contact frequency is larger than 65% or if they appear in the maximal spanning tree of any composite graph whose edge weights correspond to contact frequencies (as explained in Fig. 10c). If two proteins are adjacent, they are more likely to interact with each other in the native NPC structure than when they are not adjacent[36]. In total, 51 types of adjacencies were found (Fig. 10d). A particularly large number of adjacencies are observed for Nic96 and Nup82, which both appear in two copies per symmetry unit, as well as for the core proteins Nup192 and Nup188. Whereas the latter two proteins bridge the bulk of the NPC to the membrane proteins and also provide anchor sites for FG nucleoporins, Nic96 bridges major ring structures of the NPC and also serves as an anchor site for FG nucleoporins[37]. Most FG nucleoporins are peripherally located and therefore show only a few adjacencies.

©2007 Nature Publishing Group

Protein configuration. We can now combine the protein positions and adjacencies into a configuration of the NPC proteins (Fig. 10e, f). This representation allows us to deconvolute the composites into their constituent complexes (for example, see Figs 5b and 10g).

Synergy among restraints. How our data act synergistically is best demonstrated by the progressive increase in the certainty about the protein positions, as a result of an incremental addition of information (Fig. 11a). Hence, the variability among the 1,000 structures is significantly smaller than the uncertainties in any of the original data. For example, the allowed ranges for protein localization by immuno-EM are reduced from ±4.5 and ±8 nm along the *Z*-axis and the radial coordinate, respectively, to ±2 and ±3 nm in the ensemble, as a direct result of data integration. Similarly, data integration also improves the prediction of protein interactions (Fig. 11b).

Assessment of precision and accuracy

The accuracy of a model is defined as the difference between the model and the native structure. Therefore, it is currently impossible to know with certainty the accuracy of the determined NPC structure. Nevertheless, five lines of evidence indicate that the accuracy of our structure is similar to its precision, and thus representative of the true configuration of the NPC.

Self-consistency of the experimental data. Inconsistencies in the experimental data or its interpretation can be identified when the optimization generates only frustrated structures that do not satisfy the input restraints. This is not the case for our NPC calculations; we find only a single cluster of NPC structures that satisfy all the input restraints. To show that it is not trivial to find structures satisfying all restraints, we repeated the calculations with a comparable, but partly incorrect set of restraints (Supplementary Information). Specifically, all untagged proteins were randomly swapped between composites, leaving the number of composites, the number of proteins in each composite, and all other restraints unchanged. An optimization using this modified restraint set failed to produce any structures that satisfied all restraints.

Variability in the ensemble. We have confirmed that the ensemble of 1,000 structures is sufficiently large for the precision of the NPC architecture to be determined reliably: the reproducibility of contact frequencies calculated from random subsets of the ensemble was plotted as a function of the subset size (Supplementary Information). The similarity between two sets of contact frequencies converges for random subsets of ~100 structures.

The ability of a restraint set to define a native state. We have previously described an approach to test whether or not a given restraint set is sufficient to reconstruct a known native state[36]. In this approach, a native structure is assumed, the restraints to be tested are simulated from this structure, the structure is then reconstructed based only on these restraints, and finally the reconstruction is compared to the original assumed structure. Using this approach, we have simulated composite restraints based on our NPC structure, reproducing the number of composites and the distribution of their size in the original data set; all other restraints were kept the same as in the real application. The accuracy of the reconstructed model was comparable to the precision of the current NPC model.

Patterns unlikely to occur by chance. The distribution of nucleoporins in our structure is expected to reflect their functionality and evolution, and so should be decidedly nonrandom. Indeed, as discussed at length in the accompanying paper[37], there is a striking co-segregation of proteins by fold type to particular locations in the

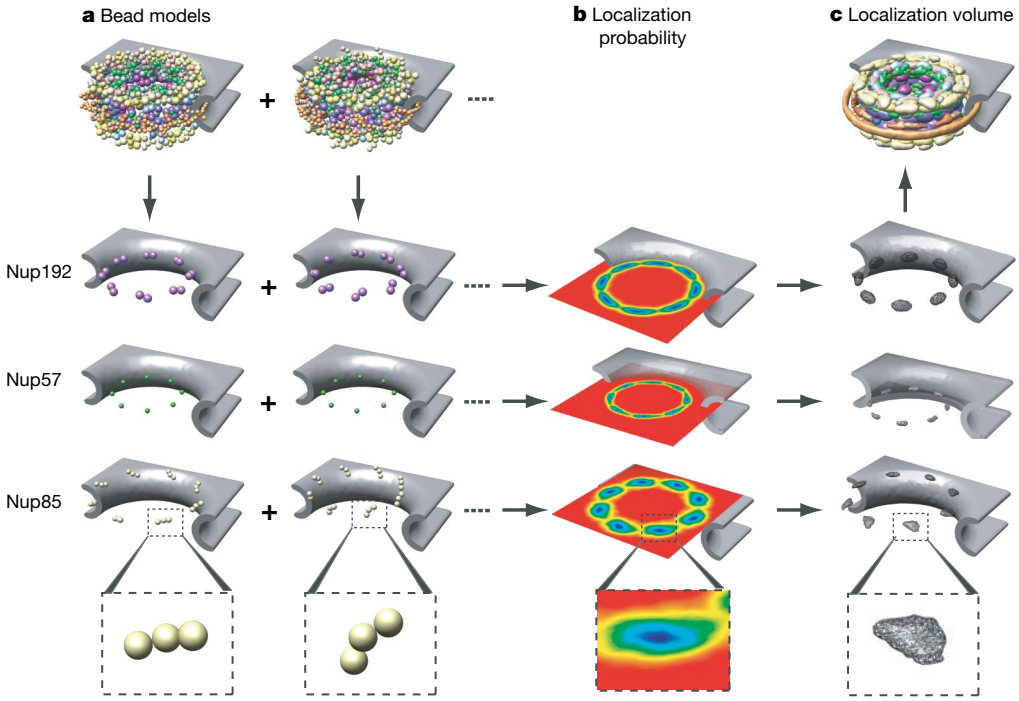


Figure 9 | Bead model, ensemble, localization probability and localization volume. **a**, Top: two representative bead models of the NPC (excluding the FG-repeat regions) from the ensemble of 1,000 superposed structures satisfying all restraints (Fig. 8b). The eight positions of three sample proteins (Nup192, Nup57 and Nup85) on the cytoplasmic side are shown, with a detailed view of the bead representation of one copy of Nup85 at the bottom. **b**, The localization probability for each protein type is obtained by converting the ensemble into the probability of any volume element being occupied by the protein. Shown are contour maps of the cross-sections in the plane parallel to the equatorial plane that contains the maximum value of the protein localization probability. **c**, The localization volume of the sample proteins, derived from the localization probability. The volume elements are first sorted by their localization probability values. The localization volume then corresponds to the top-ranked elements, the volume of which sums to the protein volume, estimated from its molecular mass. The localization volume of a protein reveals its most probable localization. Because of the limited precision of the information used here, the localization volume of a protein should not be mistaken for its density map, such as that derived by cryo-EM.

©2007 Nature Publishing Group

NPC, although no fold information (except for the transmembrane domains) was used in the generation of the structure.

Experimental data not used in the calculation of the model. Finally, our structure can be most directly tested by comparing it to experimentally determined data that were not included in the structure calculation. First, our structure is robust, in the sense that omission of a randomly chosen subset of 10% of the protein interaction data still results in structures with contact frequencies essentially identical to those derived from the complete data set. Second, the shape of our NPC structure[37] strongly resembles the published electron microscopy maps of the NPC[5,38–42], even though these data were not used here (Supplementary Fig. 22). Third, the diameter of the transport channel in our structure is ~38 nm (excluding the FG-repeat regions), in good agreement with the experimentally reported maximal diameter of transported particles[43]. Fourth, Nup133, which has been experimentally shown to interact with highly curved membranes via its ALPS-like motif, is adjacent to the nuclear envelope in our structure[44]. Moreover, three of the four additional scaffold nucleoporins that are predicted to contain the ALPS-like motif are also close to the nuclear envelope. Finally, perhaps the best example is that of the Nup84 complex. Our configuration for this complex (Fig. 5b)[37] is completely consistent with previous results[13,14,30]. Specifically, Nup85 and Seh1 form a dimer that together with Nup120 forms the trimeric 'head' of the complex, consistent with the top two arms of the 'Y'-shaped Nup84 complex (Fig. 5b)[14]. Similarly, Nup145C, Nup84, Sec13 and Nup133 form the 'tail' in

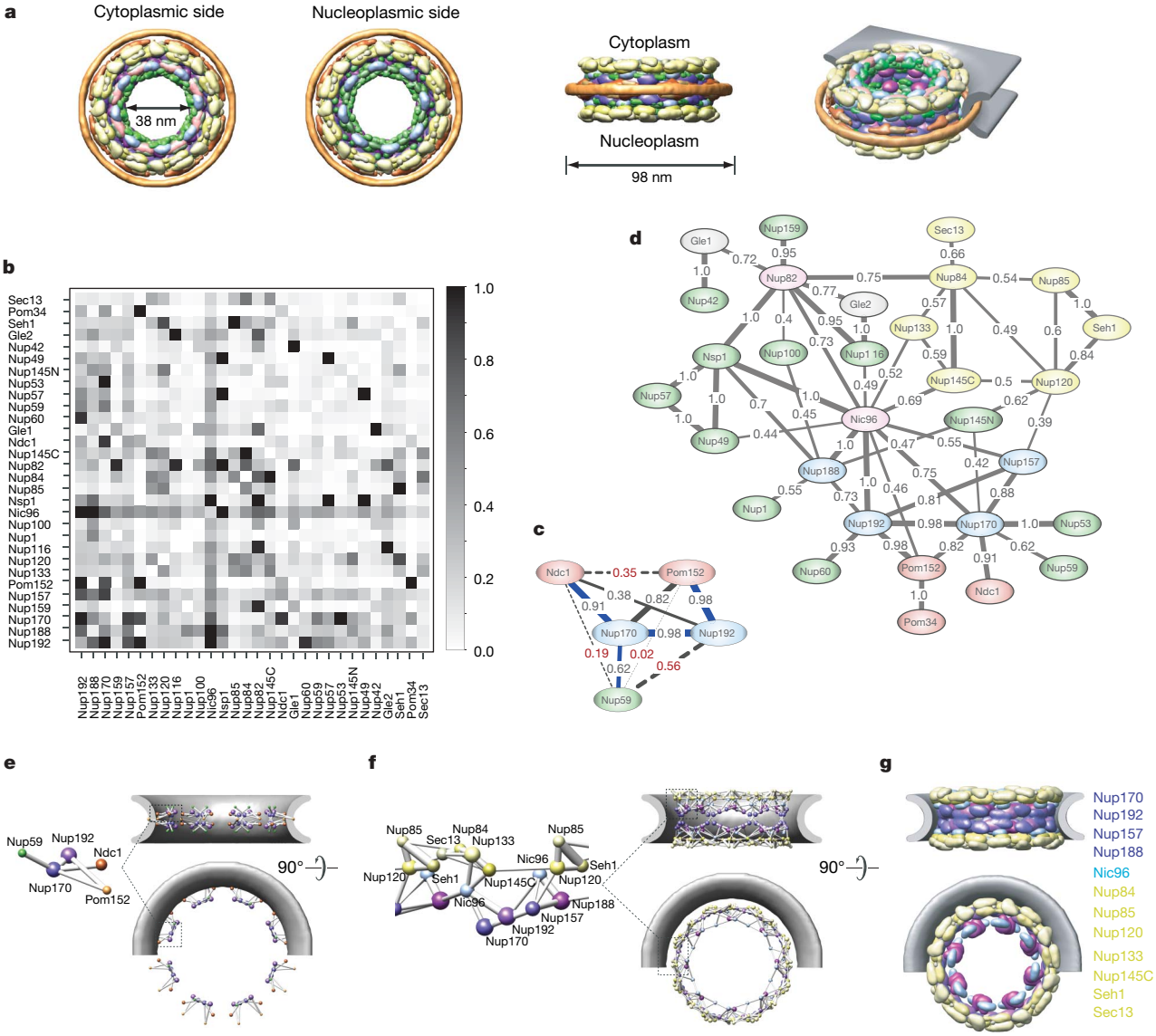


Figure 10 | Ensemble interpretation in terms of protein positions, contacts and configuration. **a**, Localization volumes of all 456 proteins in the NPC (excluding the FG-repeat regions) in four different views. The diameter of the transport channel and the NPC are also indicated. The proteins are colour-coded according to their assignment to the six NPC modules[37]. **b**, Contact frequencies for all pairs of proteins. The contact frequency of a pair of protein types is the fraction of structures in the ensemble that contains at least one protein contact between any protein instances of the two types. **c**, Contact frequencies between proteins in composite 40. Proteins are nodes connected by edges with the observed contact frequency as the edge weight (indicated by its thickness). Edges that are part of the maximal spanning tree are shown by thick blue lines; the maximal spanning tree is the spanning tree that maximizes the sum of the edge weights. All edges with a statistically significant reduction in contact frequency from their initial values implied by the composite data alone (P-value $< 10^{-3}$; Supplementary Information) are indicated by dotted lines with contact frequencies shown in red. **d**, Protein adjacencies for the whole NPC, with proteins as nodes and edges connecting proteins that are determined to be adjacent to each other. The edge weight is the observed contact frequency. **e**, Configuration of the proteins in composite 40. The location of a protein corresponds to the average position of the beads representing non-FG repeats of the protein. **f**, Configuration of Nic96 and the NPC scaffold proteins. **g**, Localization volume of Nic96 and the NPC scaffold proteins[37].

©2007 Nature Publishing Group

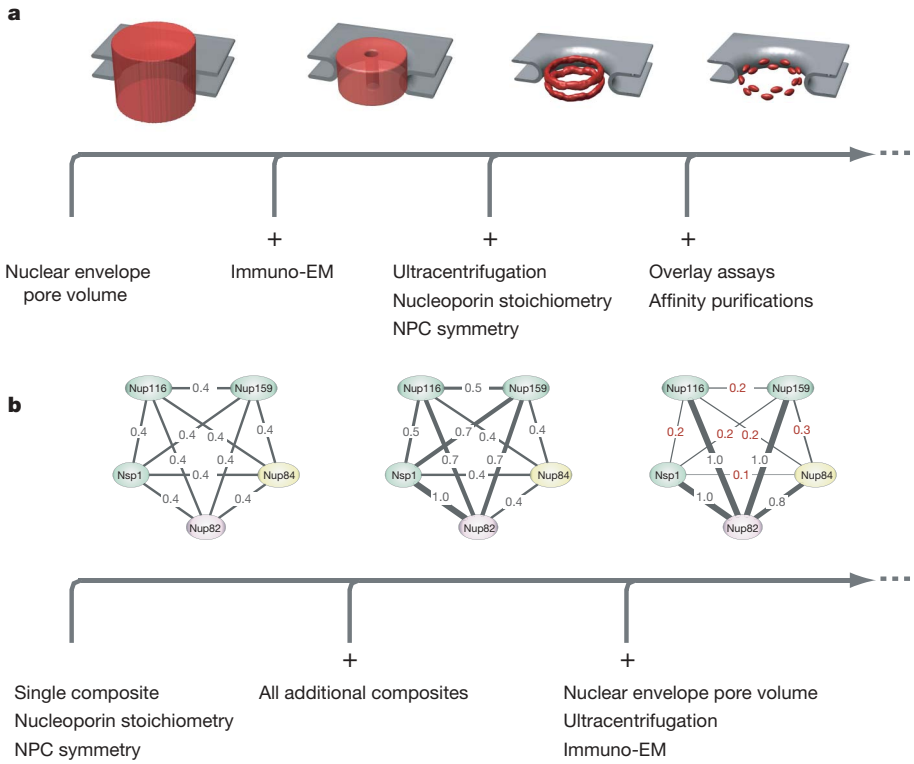


Figure 11 | The structure is increasingly specified by the addition of different types of synergistic experimental information. a, Protein positions. As an example, each panel illustrates the localization of 16 copies of Nup192 in the ensemble of NPC structures, generated using the data sets indicated below. The localization probability is contoured at 65% of its maximal value (red). The smaller the volume, the better localized are the proteins. The NPC structure is therefore essentially moulded into shape by the large amount of diverse experimental data. **b**, Protein contacts. Prediction of protein interactions from contact frequencies improves as more data are used. As an example, each panel illustrates the contact frequencies between proteins found in composite 34. Contact frequencies are shown as edge weights and indicated by the thickness of the lines connecting the proteins. Left: when only a single composite is used (together with stoichiometry and symmetry information), all interactions are equally likely (initial contact frequency, Supplementary Information). Middle: when the highest likelihood of interaction between a particular protein pair from all composites is used, the uncertainty about the interactions is reduced. Right: when all data are used, the contact frequencies are either very high (>0.65) or very low (<0.25), thus allowing a strong prediction of protein interactions. Contact frequencies reflect the likelihood that a protein interaction is formed given the data considered and are calculated from the ensemble of optimized structures. Numbers in red indicate final contact frequencies that significantly decreased (at a P-value $<10^{-3}$) from their initial values (Supplementary Information).

both our structure and the Y-shaped complex (Fig. 5b)[14]. Here, we resolve the relative positions of the proteins in this complex and show how the complex is integrated into the architecture of the entire NPC.

Together these assessments indicate that our data are sufficient to determine the configuration of the proteins comprising the NPC. Indeed, it is hard to conceive of any combination of errors that could have biased our structure towards a single solution that resembles known NPC features in so many ways.

Conclusions

We have devised an integrative approach to solve the structure of the NPC using diverse biophysical and proteomic data. This approach has several advantages. First, it benefits from the synergy among the input data. Data integration is in fact necessary for structure determination, because none of the individual data sets contains sufficient spatial information on its own. Despite the little structural information in each individual restraint, the concurrent satisfaction of all restraints derived from independent experiments markedly reduces the degeneracy of the final structures. Second, the integrative approach can potentially survey all the structures that are consistent with the data. Alternatively, if no structure is consistent with the data, then some experiments or their interpretations are incorrect. Third, this approach can make the process of structure determination more efficient, by indicating which measurements would be most informative. Fourth, the approach can, in principle, incorporate essentially any structural information about a given assembly. Thus, it is straightforward to adapt it for calculating higher resolution structures by including additional spatial restraints from higher resolution data sets, such as atomic structures of proteins, chemical cross-linking, footprinting, small angle X-ray scattering (SAXS) and cryo-EM. It is conceivable that these additional data sets might allow us to determine pseudo-atomic structures of assemblies as complex as the NPC. Furthermore, by obtaining detailed structural information concerning different stages of a dynamic process, our approach may animate the NPC's assembly and transport mechanisms[6].

The molecular architecture of many macromolecular complexes could, in principle, be resolved using a similar integrative approach. With regards to the NPC, the resulting structure has already provided abundant insights into the function and evolution of the cell[37].

METHODS SUMMARY

See Supplementary Information for a detailed description of our Methods. The experimental data, the Integrative Modelling Platform software and the NPC structural model are available at http://ncdir.org/npc.

Received 30 August; accepted 22 October 2007.

1. Sali, A., Glaeser, R., Earnest, T. & Baumeister, W. From words to literature in structural proteomics. *Nature* **422**, 216–225 (2003).
2. Rout, M. P. *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–651 (2000).
3. Macara, I. G. Transport into and out of the nucleus. *Microbiol. Mol. Biol. Rev.* **65**, 570–594 (2001).
4. Weis, K. Nucleocytoplasmic transport: cargo trafficking across the border. *Curr. Opin. Cell Biol.* **14**, 328–335 (2002).

©2007 Nature Publishing Group

5. Yang, Q., Rout, M. P. & Akey, C. W. Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. *Mol. Cell* **1**, 223–234 (1998).
6. Beck, M., Lucic, V., Förster, F., Baumeister, E. & Medalia, O. Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature* **449**, 611–615 (2007).
7. Devos, D. *et al.* Simple fold composition and modular architecture of the nuclear pore complex. *Proc. Natl Acad. Sci. USA* **103**, 2172–2177 (2006).
8. Havel, T. F. & Wüthrich, K. A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular 1H–1H proximities in solution. *Bull. Math. Biol.* **46**, 673–698 (1984).
9. Malhotra, A., Tan, R. K. & Harvey, S. C. Prediction of the three-dimensional structure of *Escherichia coli* 30S ribosomal subunit: a molecular mechanics approach. *Proc. Natl Acad. Sci. USA* **87**, 1950–1954 (1990).
10. Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L. & Rexach, M. Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl Acad. Sci. USA* **100**, 2450–2455 (2003).
11. Lim, R. Y. *et al.* Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc. Natl Acad. Sci. USA* **103**, 9512–9517 (2006).
12. Devos, D. *et al.* Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol.* **2**, e380 (2004).
13. Siniossoglou, S. *et al.* Structure and assembly of the Nup84p complex. *J. Cell Biol.* **149**, 41–54 (2000).
14. Lutzmann, M., Kunze, R., Buerer, A., Aebi, U. & Hurt, E. Modular self-assembly of a Y-shaped multiprotein complex from seven nucleoporins. *EMBO J.* **21**, 387–397 (2002).
15. Strambio-de-Castillia, C., Blobel, G. & Rout, M. P. Isolation and characterization of nuclear envelopes from the Yeast *Saccharomyces. J. Cell Biol.* **131**, 19–31 (1995).
16. Miller, A. L. *et al.* Cytoplasmic inositol hexakisphosphate production is sufficient for mediating the Gle1-mRNA export pathway. *J. Biol. Chem.* **279**, 51022–51032 (2004).
17. Solsbacher, J., Maurer, P., Vogel, F. & Schlenstedt, G. Nup2p, a yeast nucleoporin, functions in bidirectional transport of importin alpha. *Mol. Cell. Biol.* **20**, 8468–8479 (2000).
18. Marelli, M., Aitchison, J. D. & Wozniak, R. W. Specific binding of the karyopherin Kap121p to a subunit of the nuclear pore complex containing Nup53p, Nup59p, and Nup170p. *J. Cell Biol.* **143**, 1813–1830 (1998).
19. Archambault, V. *et al.* Genetic and biochemical evaluation of the importance of Cdc6 in regulating mitotic exit. *Mol. Biol. Cell* **14**, 4592–4604 (2003).
20. Archambault, V. *et al.* Targeted proteomic study of the cyclin-Cdk module. *Mol. Cell* **14**, 699–711 (2004).
21. Tackett, A. J. *et al.* I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions. *J. Proteome Res.* **4**, 1752–1756 (2005).
22. Cristea, I. M., Williams, R., Chait, B. T. & Rout, M. P. Fluorescent proteins as proteomic probes. *Mol. Cell. Proteomics* **4**, 1933–1941 (2005).
23. Niepel, M., Strambio-de-Castillia, C., Fasolo, J., Chait, B. T. & Rout, M. P. The nuclear pore complex-associated protein, Mlp2p, binds to the yeast spindle pole body and promotes its efficient assembly. *J. Cell Biol.* **170**, 225–235 (2005).
24. Cristea, I. M. *et al.* Tracking and elucidating alphavirus-host protein interactions. *J. Biol. Chem.* **281**, 30269–30278 (2006).
25. Zhang, W. & Chait, B. T. ProFound: an expert system for protein identification using mass spectrometric peptide mapping information. *Anal. Chem.* **72**, 2482–2489 (2000).
26. Krutchinsky, A. N., Kalkum, M. & Chait, B. T. Automatic identification of proteins with a MALDI-quadrupole ion trap mass spectrometer. *Anal. Chem.* **73**, 5066–5077 (2001).
27. Stelter, P. *et al.* Molecular basis for the functional interaction of dynein light chain with the nuclear-pore complex. *Nature Cell Biol.* **9**, 788–796 (2007).
28. Murphy, R., Watkins, J. L. & Wente, S. R. GLE2, a *Saccharomyces cerevisiae* homologue of the *Schizosaccharomyces pombe* export factor RAE1, is required for nuclear pore complex structure and function. *Mol. Biol. Cell* **7**, 1921–1937 (1996).
29. Murphy, R. & Wente, S. R. An RNA-export mediator with an essential nuclear export signal. *Nature* **383**, 357–360 (1996).
30. Lutzmann, M. *et al.* Reconstitution of Nup157 and Nup145N into the Nup84 complex. *J. Biol. Chem.* **280**, 18442–18451 (2005).
31. Bailer, S. M. *et al.* Nup116p associates with the Nup82p-Nsp1p-Nup159p nucleoporin complex. *J. Biol. Chem.* **275**, 2354–23548 (2000).
32. Grandi, P., Doye, V. & Hurt, E. C. Purification of NSP1 reveals complex formation with 'GLFG' nucleoporins and a novel nuclear pore protein NIC96. *EMBO. J.* **12**, 3061–3071 (1993).
33. Shen, M. Y. & Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–2524 (2006).
34. Harding, S. E. Determination of macromolecular homogeneity, shape, and interactions using sedimentation velocity analytical ultracentrifugation. *Methods Mol. Biol.* **22**, 61–73 (1994).
35. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
36. Alber, F., Kim, M. F. & Sali, A. Structural characterization of assemblies from overall shape and subcomplex compositions. *Structure* **13**, 435–445 (2005).
37. Alber, F. *et al.* The molecular architecture of the nuclear pore complex. *Nature* doi:10.1038/nature06405 (this issue).
38. Akey, C. W. & Radermacher, M. Architecture of the *Xenopus* nuclear pore complex revealed by three-dimensional cryo-electron microscopy. *J. Cell Biol.* **122**, 1–19 (1993).
39. Stoffler, D. *et al.* Cryo-electron tomography provides novel insights into nuclear pore architecture: implications for nucleocytoplasmic transport. *J. Mol. Biol.* **328**, 119–130 (2003).
40. Kiseleva, E. *et al.* Yeast nuclear pore complexes have a cytoplasmic ring and internal filaments. *J. Struct. Biol.* **145**, 272–288 (2004).
41. Hinshaw, J. E., Carragher, B. O. & Milligan, R. A. Architecture and design of the nuclear pore complex. *Cell* **69**, 1133–1141 (1992).
42. Beck, M. *et al.* Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* **306**, 1387–1390 (2004).
43. Pante, N. & Kann, M. Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm. *Mol. Biol. Cell* **13**, 425–434 (2002).
44. Drin, G. *et al.* A general amphipathic α-helical motif for sensing membrane curvature. *Nature Struct. Mol. Biol.* **14**, 138–146 (2007).
45. Schurmann, G., Haspel, J., Grumet, M. & Erickson, H. P. Cell adhesion molecule L1 in folded (horseshoe) and extended conformations. *Mol. Biol. Cell* **12**, 1765–1773 (2001).

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

Acknowledgements We thank H. Shio for performing the electron microscopic studies; J. Fanghänel, M. Niepel and C. Strambio-de-Castillia for help in developing the affinity purification techniques; M. Magnasco for discussions and advice; A. Kruchinsky for assistance with mass spectrometry; M. Topf, D. Korkin, F. Davis, M.-Y. Shen, F. Foerster, N. Eswar, M. Kim, D. Russel, B. Peterson and B. Webb for many discussions about structure characterization by satisfaction of spatial restraints; C. Johnson, S. G. Parker and C. Silva, T. Ferrin and T. Goddard for preparation of some figures; and S. Pulapura and X. J. Zhou for their help with the design of the conditional diameter restraint. We are grateful to J. Aitchison for discussion and insightful suggestions. We also thank all other members of the Chait, Rout and Sali laboratories for their assistance. We acknowledge support from an Irma T. Hirschl Career Scientist Award (M.P.R.), a Sinsheimer Scholar Award (M.P.R.), a grant from the Rita Allen Foundation (M.P.R.), a grant from the American Cancer Society (M.P.R.), the Sandler Family Supporting Foundation (A.S.), the Human Frontier Science Program (A.S., L.M.V.), NSF (A.S.), and grants from the National Institutes of Health (B.T.C., M.P.R., A.S.), as well as computer hardware gifts from R. Conway, M. Homer, Intel, Hewlett-Packard, IBM and Netapp (A.S.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.S. (sali@salilab.org), M.P.R. (rout@rockefeller.edu), or B.T.C. (chait@rockefeller.edu).

©2007 Nature Publishing Group

ARTICLES

The molecular architecture of the nuclear pore complex

Frank Alber[1]*, Svetlana Dokudovskaya[2]*†, Liesbeth M. Veenhoff[2]*†, Wenzhu Zhang[3], Julia Kipper[2]†, Damien Devos[1]†, Adisetyantari Suprapto[2]†, Orit Karni-Schmidt[2]†, Rosemary Williams[2], Brian T. Chait[3], Andrej Sali[1] & Michael P. Rout[2]

Nuclear pore complexes (NPCs) are proteinaceous assemblies of approximately 50 MDa that selectively transport cargoes across the nuclear envelope. To determine the molecular architecture of the yeast NPC, we collected a diverse set of biophysical and proteomic data, and developed a method for using these data to localize the NPC's 456 constituent proteins (see the accompanying paper). Our structure reveals that half of the NPC is made up of a core scaffold, which is structurally analogous to vesicle-coating complexes. This scaffold forms an interlaced network that coats the entire curved surface of the nuclear envelope membrane within which the NPC is embedded. The selective barrier for transport is formed by large numbers of proteins with disordered regions that line the inner face of the scaffold. The NPC consists of only a few structural modules that resemble each other in terms of the configuration of their homologous constituents, the most striking of these being a 16-fold repetition of 'columns'. These findings provide clues to the evolutionary origins of the NPC.

Nuclear pore complexes (NPCs) are large (~50 MDa) proteinaceous assemblies spanning the nuclear envelope, where they function as mediators of bidirectional exchange between the nucleoplasmic and cytoplasmic compartments[1]. Nucleocytoplasmic transport of macromolecular cargoes depends on their recognition by transport factors, which interact with the NPC to carry cargoes across the nuclear envelope[2,3].

NPCs show a broad degree of compositional and structural conservation among all eukaryotes studied[4,5]. Each NPC contains at least 456 individual protein molecules and is composed of ~30 distinct proteins (nucleoporins)[6,7]. Electron microscopy studies in several organisms have revealed that the general morphology of the NPC is conserved[8–13]. These studies show the NPC to be a doughnut-shaped structure, consisting of eight spokes arranged radially around a central channel that serves as the conduit for macromolecular transport. Each NPC spans the nuclear envelope through a pore formed by the fusion of the inner and outer nuclear envelope membranes. Numerous filamentous structures project from the NPC into the cytoplasm and nucleoplasm[1,5].

Using an approach described in the accompanying paper, we have determined the molecular architecture of the *Saccharomyces cerevisiae* NPC[14]. Here, we describe and discuss the major features revealed by this analysis.

Architectural overview of the NPC

Our NPC structure, when viewed as projections of the mass density of all the nucleoporins (Fig. 1a), closely resembles maps and images obtained using electron microscopy, although in the electron microscopy maps the positions of individual nucleoporins were not defined[8–13]. We have excluded from our map proteins not previously considered 'integral' components of the NPC, such as Mlp1, Mlp2, transport factors and Nup2 (ref. 6). There is no basket in our map, probably because Mlp1 and Mlp2, the probable basket-forming proteins[15], are absent. We do, however, observe a low-density cloud, corresponding to the unstructured regions of FG (phenylalanine-glycine) repeat nucleoporins (see below), filling the central channel and projecting into the nucleoplasm and cytoplasm. The diameter of the central channel in the absence of the unstructured components (Fig. 1b) is ~38 nm, matching the known maximum sizes of transported particles[16]. The diameter of the whole NPC is ~98 nm, and the height of the structured portion is ~37 nm (Fig. 1b), again close to the dimensions observed by electron microscopy[11,13].

A fundamental symmetry unit of the NPC is the spoke[13]. Each spoke can be divided into two almost identical nucleoplasmic and cytoplasmic halves, joined at the NPC's equator and related to each other by a dyad axis (Fig. 2 and Supplementary Movie). Together, eight spokes connect to form several co-axial rings (Fig. 2); we refer to these as the membrane rings (not to be confused with the nuclear envelope membrane itself), the two outer rings at the nucleoplasmic and cytoplasmic periphery, and the two adjacent inner rings. Groups of nucleoporins that we term linker nucleoporins are attached between both sets of outer and inner rings (Fig. 2). Another group of related proteins is collectively termed FG nucleoporins. Each FG nucleoporin contains a small structured domain that serves as the anchor site to the rest of the NPC, and a larger unstructured region that contains characteristic repetitive sequence motifs termed FG repeats. The FG nucleoporins are largely exposed on the inner surface of the spokes and are anchored either to the inner rings or to the

[1]Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Chemistry, and California Institute for Quantitative Biosciences, Mission Bay QB3, 1700 4th Street, Suite 503B, University of California at San Francisco, San Francisco, California 94158-2330, USA. [2]Laboratory of Cellular and Structural Biology, and [3]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA. †Present addresses: Laboratory of Nucleocytoplasmic Transport, Institut Jacques Monod, 2 place Jussieu, Tour 43, Paris 75251, France (S.D.); Department of Biochemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands (L.M.V.); German Aerospace Center (PT-DLR), Heinrich-Konen-Strasse 1, D-53227 Bonn, Germany (J.K.); Structural Bioinformatics, EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany (D.D.); Office of Technology Transfer, The Rockefeller University, 1230 York Avenue, New York, New York 10065, USA (A.S.); Herbert Irving Comprehensive Cancer Centre, Columbia University, 1130 St Nicholas Avenue, New York, New York 10032, USA (O.K.-S.).
*These authors contributed equally to this work.

©2007 Nature Publishing Group

linker nucleoporins; curiously, few make connections directly to the outer rings (Fig. 2). Among the most prominent features in electron micrographs of the NPC are the co-axial rings, the positions of which are similar to those of rings in our structure: the nucleoplasmic and cytoplasmic rings seen by electron microscopy coincide with the position of the two outer rings in our map, whereas the spoke ring complex and luminal ring coincide respectively with our map's inner rings and membrane rings (Figs 1 and 2, and Supplementary Fig. 1). Two copies per spoke of Nup82 (a linker nucleoporin) are found adjacent to the cytoplasmic outer ring. They add significant extra mass to the cytoplasmic side of the NPC compared with the nucleoplasmic side, probably accounting for much of the extra density of the cytoplasmic ring seen by electron microscopy (Figs 1 and 2)[8–13]. Details in some of these electron microscopy maps differ from our map, such as the presence of a 'central transporter' and weaker densities at the nucleoplasmic and cytoplasmic surfaces of the yeast NPC map[13]; such differences are probably due to the limits of our map's precision, as well as issues with NPC preservation, imaging and averaging for electron microscopy.

Figure 2 also illustrates how a series of discrete clusters of nucleoporins represents the 'building blocks' of the NPC. As described in the accompanying paper[14], these complexes can frequently be isolated from the NPC, demonstrating that they are indeed coherent NPC substructures. Most of the interactions holding the NPC together are heterotypic, with only a few homotypic interactions. Some nucleoporins act as 'keystones', forming significantly larger numbers of interactions than most other nucleoporins, bridging multiple NPC substructures (Fig. 10 in the accompanying paper)[14]. An example is Nic96, which bridges the inner and outer rings and also serves as an anchor site for FG nucleoporins. The FG nucleoporins themselves are mainly peripherally located and therefore each shows only a few interactions.

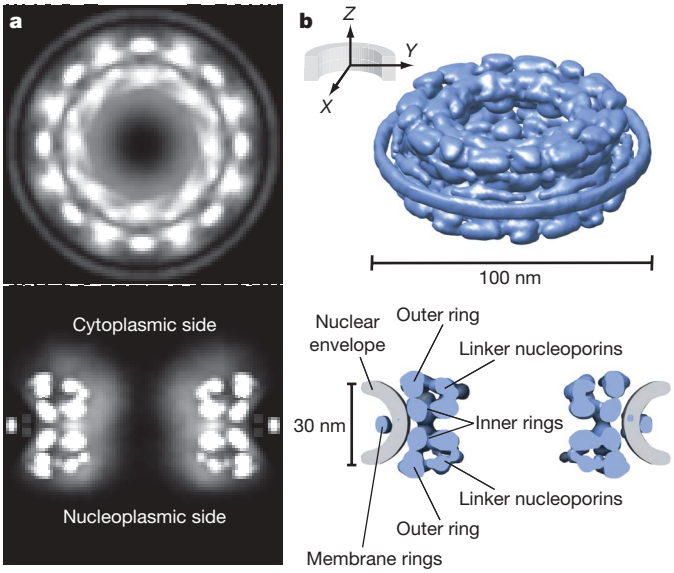


Figure 1 | Architectural overview of the NPC. a, Projections of nucleoporin mass density (white), derived from the combined localization volumes of all structured domains and the normalized localization probability of all unstructured regions[14]. Top, *en face* view showing a density projection along the Z-axis from $Z = -50$ nm to $Z = +50$ nm. As in electron microscopy maps of the NPC, radial arms of density correspond to spokes that interconnect to form two strong concentric rings encircling a central region containing low-density unstructured material and bounded by peripheral membrane rings, giving an overall diameter of ~98 nm. Bottom, a slice along the central z-axis showing a projection of density from $X = -5$ nm to $X = +5$ nm. More density can be seen on the cytoplasmic side of the NPC. The low-density unstructured material constricts the central channel to ~10 nm diameter. **b**, The structured nucleoporin domains of the NPC, represented by a density contour (blue) such that the volume of the contour corresponds approximately to the combined volume of the 456 nucleoporins comprising the NPC[14]. Top: view from a point ~30° from the equatorial plane of the NPC. Bottom: a slice along the central Z-axis between $X = -5$ nm and $X = 5$ nm, in which the nuclear envelope is also shown (in grey). Major features of the NPC are indicated.

A membrane-coating complex

We term the structural core of the NPC the core scaffold, formed from the outer and inner rings (Figs 2 and 3). Most linker nucleoporins and FG nucleoporin attachment sites are on the surface of the core scaffold that faces the central channel; on the outer surface are found the pore membrane (that is, that portion of the nuclear envelope into which the NPC is embedded) and the attachment sites for the membrane rings. The core scaffold probably has a key role in maintaining the stability of the nuclear envelope by ensuring coplanarity of the outer and inner nuclear envelope membranes as they exit the pore. We propose that a major function of the outer rings, by virtue of their location, is to facilitate the smooth transition of the pore membrane into the inner and outer nuclear envelope membranes (Fig. 3).

We have previously assigned fold types to domains in all the nucleoporins[17,18]. We are now in a position to determine the actual locations of these fold types in the NPC. Notably, we find that the core scaffold is made up of nucleoporins containing three distinct arrangements of only two fold types: proteins with a single α-solenoid domain (the α-solenoid fold is a superhelix of α-helices), proteins with a single β-propeller domain, and proteins composed of an amino-terminal β-propeller followed by a carboxy-terminal α-solenoid domain (Fig. 3). This latter pattern is unique to the clathrin-like proteins that coat transport vesicles, such as the clathrin heavy chain itself and Sec31, and all three fold arrangements are found in such coating complexes[17–23]. We have already proposed that the core structures of both NPCs and vesicle-coating complexes have a common evolutionary origin[17,18]. The resemblance of the core scaffold to vesicle-coating complexes does not end here; just like clathrin and the COPII coatomer, the scaffold nucleoporins form an interconnected network, such that the entire core scaffold faces and completely coats the pore membrane, with the clathrin/Sec31-like nucleoporins closer to the membrane than the α-solenoid domain proteins (Fig. 3). The vesicle-coating protein family can form different scaffolds, as demonstrated by the markedly different arrangements found in the COPII and clathrin coats[19,23]. This attribute may have made them well suited for accommodating the combination of negative and positive curvatures characteristic of the pore membrane[24].

Interestingly, five (Nup133, Nup120, Nup85, Nup170 and Nup188) of the scaffold nucleoporins contain an ALPS-like motif that has been experimentally verified to fold into an amphipathic α helix and target curved membrane regions[25]. Our structure is consistent with this observation, since at least four of these nucleoporins are found immediately adjacent to the pore membrane (Fig. 3).

Thus, not only do proteins of the core scaffold and vesicle-coating complexes share a similar fold composition, but they also appear to share a similar assembly architecture and function—namely, to form coats that curve membranes. In the case of the core scaffold, it is the nuclear envelope that is moulded into the sharply curved pore membrane (Fig. 3). The core scaffold is dominated by an evenly distributed meshwork of α-solenoid domains (Fig. 3), a fold type that is expected to be flexible[26], allowing large conformational changes without breaking protein–protein interactions (as is seen in vesicle-coating complexes). This fold flexibility could, in turn, explain the significant degree of flexibility reported for the whole NPC, necessary to accommodate the diverse sizes of nucleocytoplasmic transport cargoes and the malleability of the nuclear envelope[9,27,28].

©2007 Nature Publishing Group

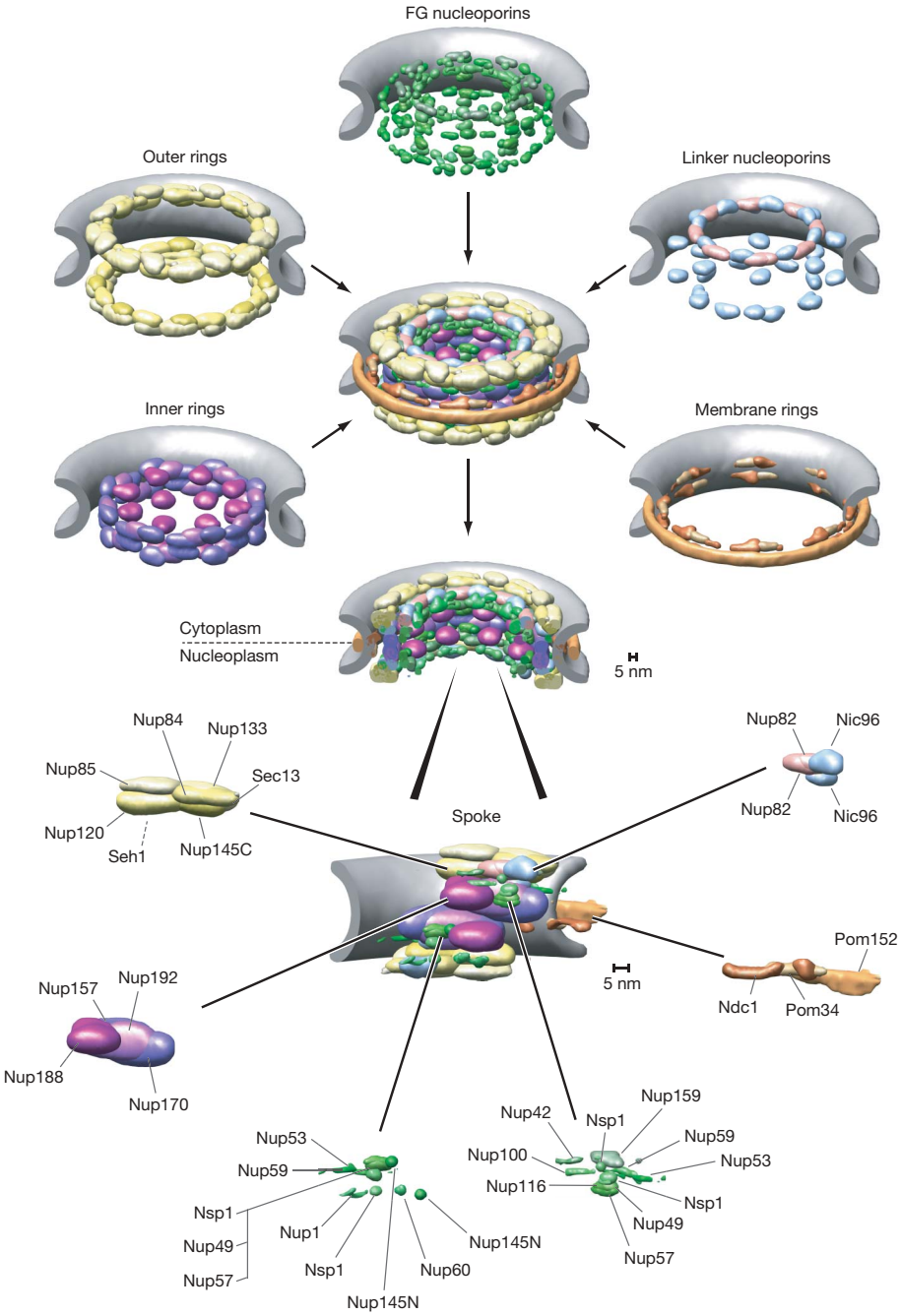


Figure 2 | Localization of major substructures and their component nucleoporins in the NPC. This figure is a single view of data presented in our Supplementary Movie. The nucleoporins are represented by their localization volumes[14] and have been coloured according to their classification into five distinct substructures on the basis of their location and functional properties: the outer rings in yellow, the inner rings in purple, the membrane rings in brown, the linker nucleoporins in blue and pink, and the FG nucleoporins (for which only the structured domains are shown) in green. The pore membrane is shown in grey. A single arbitrary repeat unit, termed the spoke, is shown dissected into its component nucleoporins. Together, the outer and inner rings connect to form the NPC's core scaffold (Fig. 3). Each of the outer rings makes connections with the adjacent linker nucleoporins and inner rings, but connects with few FG nucleoporins and no components of the membrane rings. The two inner rings are closely associated with each other at the NPC's equator and form connections with all three integral membrane proteins in the membrane rings, thereby anchoring the NPC to the nuclear envelope. The bulk of the membrane rings is formed by homo-oligomerization of the C-terminal domain of Pom152. The linker nucleoporins Nic96 and Nup82 are anchored between the inner and outer rings and have a central role in bridging the core scaffold of the NPC with the functionally important FG nucleoporins. On both the cytoplasmic and nucleoplasmic sides of each spoke, one copy of Nic96 is anchored through Nup192 and a second copy through Nup188. Whereas one copy of Nic96 carries the FG nucleoporins Nsp1, Nup57 and Nup49, the second copy forms interactions to another copy of Nsp1 and at the cytoplasmic side also interacts with Nup82. Here, Nup82 associates with the FG nucleoporins Nup159, Nup116, Nsp1 and Nup42. Thus, Nsp1 forms at least two distinct complexes in the NPC: one exclusively cytoplasmic and one disposed symmetrically[52–55]. By contrast, the FG nucleoporins found only on the nucleoplasmic side connect mainly to the inner ring nucleoporins, as do Nup53 and Nup59, both of which also face the pore membrane. The scale bars indicate the average standard deviation of the distance between a pair of neighbouring proteins in the 1,000 best-scoring configurations[14].

©2007 Nature Publishing Group

Attachment at the nuclear pore membrane

The membrane rings form a discrete region of the NPC, containing the three pore membrane proteins Pom152, Pom34 and Ndc1. It is the core scaffold's inner rings that interact with the membrane rings, thus anchoring the NPC to the pore membrane (Fig. 2). A component of the membrane rings (Pom152) homo-oligomerizes at its C terminus to form the ring that equatorially bounds the NPC in the perinuclear lumen[14]. This luminal portion consists of the C-terminal part of Pom152, containing domains predicted to assume the cadherin fold[18]. Members of the cadherin family are transmembrane receptors that form homophilic binding interfaces[29], probably accounting for the oligomeric luminal ring. Perhaps the NPC carries the remnants of an ancient transmembrane receptor, still attached to its vesicle-coating complex.

Transport factor docking sites and nucleocytoplasmic transport

The transport function of the NPC appears to be mediated mainly by the FG nucleoporins. The FG-repeat regions within each FG nucleoporin provide the NPC's docking sites for transport factor–cargo complexes[1,30–33]. The FG nucleoporins and especially their unstructured FG-repeat regions are the least specified part of our structure. Nevertheless, we can still draw conclusions concerning the localization of the FG-repeat regions by using a simplified representation[14]. Because these regions can adopt many different possible configurations in our calculations, on averaging they produce a cloud of low density surrounding their structurally resolved attachment sites, collectively filling and surrounding the central channel and extending into the nucleoplasm and cytoplasm (Figs 1 and 4). This spatial distribution of FG-repeat regions is consistent with 'virtual gating' models explaining the mechanism of nucleocytoplasmic transport[6,31], in which the FG-repeat density represents an effective exclusion filter for macromolecular particles that do not contain FG-repeat binding sites, but is permeable to transport factors that do possess these sites[2,6,31,34–39]. Thus, the cloud of FG-repeat regions forms a zone of selectivity around and across the NPC. The cloud thins radially from the walls of the central channel to the *Z*-axis, limiting the effective diameter of the central channel (Figs 1 and 4). In our structure, this diameter is less than 10 nm, similar to the maximal size of particles that can freely diffuse between the nucleoplasmic and cytoplasmic compartments[2]. Actively transporting cargo–transport factor complexes can displace this diffuse cloud, with the very largest pushing the cloud to the sides of the central channel up to the channel's maximum diameter of ~38 nm.

Nic96 and Nup82 provide anchor points for most of the FG nucleoporins, with connections also being made to the inner ring (Fig. 2). The FG nucleoporins can be divided into three groups according to their localization in the NPC: those that are attached mainly or exclusively to the cytoplasmic or nucleoplasmic side of the NPC, and those attached symmetrically on both sides (Fig. 4)[6]. The distributions of these groups of FG-repeat regions overlap heavily, consistent with the observed long reach of the individual FG-repeat regions[40,41]. The overlap suggests that a transport factor attached to one FG nucleoporin can readily exchange with many other surrounding FG nucleoporins, thus facilitating rapid transit across the NPC.

In contrast to most of the FG nucleoporins, a few transport factor binding sites (in particular Nup53 and Nup59) also face the pore membrane such that they are readily accessible to membrane proteins, as has been previously suggested[42]. These nucleoporins could mediate the transport of transmembrane proteins, in agreement with recent studies showing that active transport is responsible for the translocation of integral membrane proteins from the outer to the inner nuclear membrane[43,44].

Modular duplication in the evolution of the NPC

A striking pattern is revealed when we map the nucleoporins into our NPC structure based on their previously assigned fold types[18]. We find that each spoke can be divided into two parallel columns, in

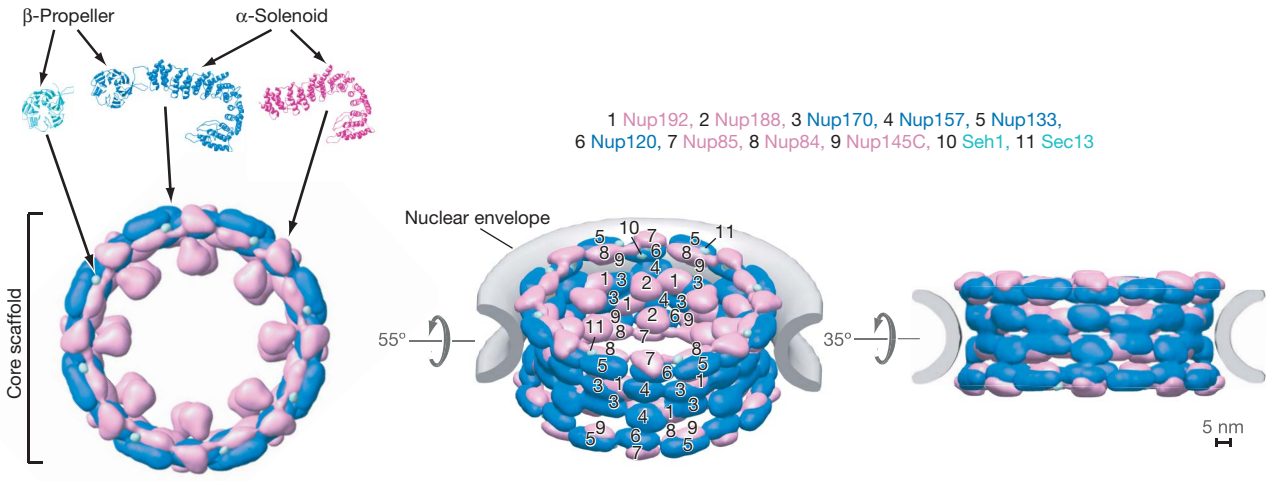


Figure 3 | The core scaffold as a membrane-coating complex. We show here the outer and inner ring nucleoporins comprising the core scaffold. The linker nucleoporins, FG nucleoporins and membrane ring are omitted for clarity. At the top of the left panel are shown the fold types comprising the nucleoporins of the core scaffold: Nup84, Nup85, Nup145C, Nup188 and Nup192 consist mainly of α-solenoid folds (pink); Sec13 and Seh1 are composed of β-propeller folds (cyan); Nup120, Nup133, Nup157 and Nup170 contain both N-terminal β-propeller folds and C-terminal α-solenoid folds (blue), an arrangement shared with clathrin and Sec31. Each of these nucleoporins is present in 16 copies to make the full 176-nucleoporin core scaffold, which is shown in three views related by the indicated rotation around an axis parallel with the NPC's equatorial plane. The localization volumes of all the α-solenoid nucleoporins (pink), all β-propeller nucleoporins (cyan), and all clathrin-like nucleoporins (blue) are indicated. The clathrin-like nucleoporins appear to be located at the outer surface of the core scaffold, adjacent to the surface of the nuclear envelope's pore membrane. Numbers on the middle panel indicate the approximate positions of each nucleoporin. The scale bar indicates the standard deviation of the distance between a pair of neighbouring proteins in the 1,000 best-scoring configurations[14].

©2007 Nature Publishing Group

which almost every nucleoporin in one column contains a counterpart of similar size and fold in a similar position in the adjacent column (Fig. 5). For most of these nucleoporin pairs, the relationship goes beyond structural similarity alone, as they are either clear homologues (for example, Seh1 and Sec13) or duplicate copies (for example, Nic96) (Fig. 5a)[17,18,45–48]. Furthermore, the two columns are analogous in terms of the interactions of their constituents (Fig. 5b). This observation indicates that homologous nucleoporin pairs may share similar functions, which in turn could explain why many single nucleoporin deletions are not lethal in yeast. We can thus consider the NPC as being made of a 16-fold repetition of columns, each with a similar architecture. An underlying 16-fold symmetry of the NPC has been previously suggested by electron microscopy studies[8,49], including yeast NPCs[13] (Supplementary Fig. 22). This pattern may be explained if an ancient duplication event gave rise to the two columns comprising each spoke. In yeast, many of the nucleoporin pairs in the columns (for example, Nup157 and Nup170) resulted from a whole-genome duplication[50]; in other organisms, the duplication is revealed as a single nucleoporin present in two copies (for example, vertebrate Nup155). Moreover, another duplication may have produced the outer and inner rings, as each column can be further divided into units that again are related by pairs of nucleoporins with identical fold types in equivalent positions (one member of each pair in the inner ring, the other in the outer ring). As a result, one can consider the NPC to be made of only a few structural modules, each consisting of only two to four proteins (Fig. 5c). These modules resemble each other, both in terms of being composed of homologous proteins and in their similar arrangement. The architecture of the NPC thus appears to be based on the hierarchical repetition of such structural modules that probably evolved through a series of gene duplications and divergences. Duplications and divergencies of these kinds are seen in other coating complexes, such as the clathrin–adaptin complex and COPI complex[51]. A primordial NPC may have been a multimer containing only a few different modules, rather like the clathrin and COPII coats. It may also have carried only a few kinds of FG nucleoporins, suggesting that the selective barrier was simpler. Just a few different transport factors would have been sufficient to recognize the few kinds of docking sites in this simple barrier; thus, the plethora of transport factors found in modern eukaryotes may have also evolved by duplication events, keeping pace with the evolutionary duplication and diverging specialization of the FG nucleoporins in the NPC's modules.

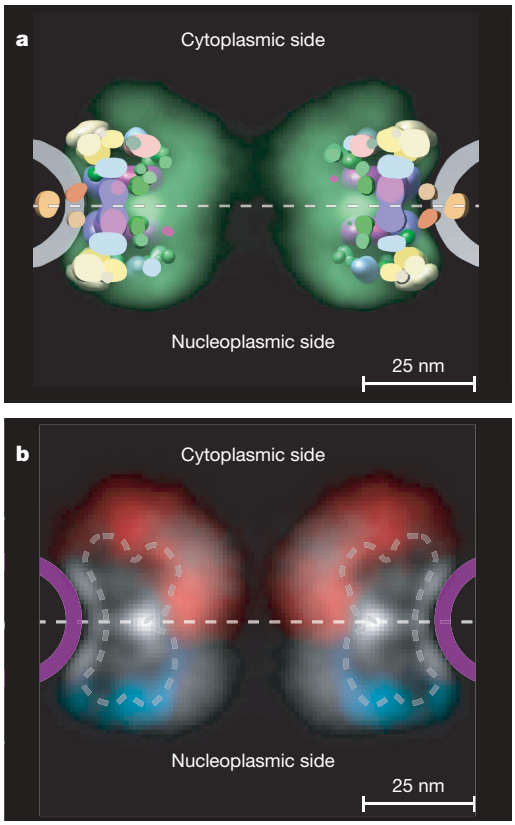


Figure 4 | Distribution of the disordered FG-repeat regions in the NPC. **a**, Slice through the NPC, as in Fig. 1b; here we show the structured domains of all nucleoporins represented by their localization volumes[14] coloured according to their classification into five distinct substructures (Fig. 2). Also shown is the localization probability of the unstructured regions of all FG nucleoporins (green cloud), visualized in Chimera[56]. **b**, Projection of the localization probabilities of the FG-repeat regions from all the FG nucleoporins is shown by a density plot, sampled in a plane perpendicular to the central Z-axis from $X = -5$ nm to $X = +5$ nm. Projections from the FG-repeat regions belonging to FG nucleoporins anchored mainly or exclusively on one side of the NPC are indicated: red for those that are cytoplasmically disposed (Nup42, Nup100, Nup116 and Nup159), blue for those nucleoplasmically disposed (Nup60, Nup145N and Nup1) and white for those FG nucleoporins found equally on both sides (Nup49, Nup53, Nup57, Nup59 and Nsp1). The equatorial plane of the NPC is indicated by a dashed white line, the position of the NPC density in Fig. 1 is indicated by a dashed grey line, and the position of the pore membrane is shown in purple. Although each group concentrates around its distinct anchor sites, considerable overlap can be seen between the three groups of FG-repeat regions. Peaks of density, belonging mainly to the repeat regions of Nup53 and Nup59, also face the pore membrane. A scale bar of 25 nm is shown.

Concluding remarks

We have determined the detailed molecular architecture of the NPC in the yeast *Saccharomyces cerevisiae*. Even though the primary sequence conservation between nucleoporins from different model organisms is low, the high conservation of the overall shape and predicted fold types[18] implies that the overall architecture of the NPC described here is highly conserved among eukaryotes.

Although the NPC is a complex structure, our analysis reveals underlying simplicities in its architecture. At its heart, the NPC contains a highly connected scaffold that attaches to and coats the curved pore membrane. The fold composition of the nucleoporins forming the scaffold is remarkably simple, consisting of only two different domain folds, the configurations of which resemble those found in vesicle-coating complexes—to which the NPC may therefore be evolutionarily related. This scaffold anchors disordered clouds of filaments that fill the central channel and regions proximal to the pore membrane, and project into the nucleoplasm and cytoplasm. These clouds act as a selective barrier to mediate nucleocytoplasmic trafficking. We found that proteins in each half-spoke are either present in duplicate copies or homologous pairings, sharing equivalent interaction patterns; thus, the NPC is another example of how a complicated structure can evolve from the duplication, divergence and elaboration of simple ancestral modules.

The NPC's architecture also suggests a possible assembly process, analogous to the formation of coated vesicles. In this process, the NPC's membrane proteins might serve as the receptors for the attachment of the inner ring to the nuclear envelope and polymerization of the core scaffold, forming a coat on the curved pore membrane. Attachment of the linker and FG nucleoporins to this coat would complete the NPC.

Further elucidation of the evolutionary origin, transport mechanism and assembly pathway of the NPC requires higher resolution information, encompassing the atomic structures of nucleoporins and their intermolecular arrangements. Given that our structure determination method can incorporate such information, we envision continued steady progress in describing the fine architecture of the NPC.

©2007 Nature Publishing Group

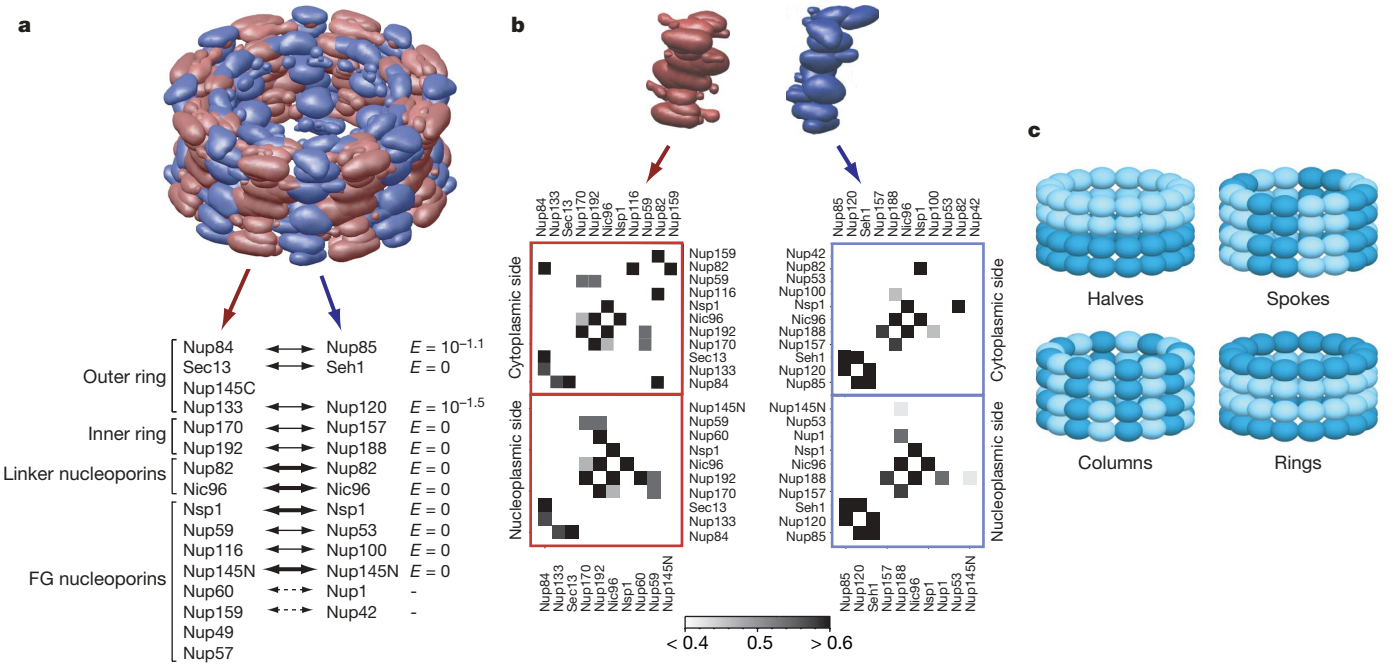


Figure 5 | Modular duplication in the NPC. a, The NPC can be divided into two alternating equivalent groups of nucleoporins arranged in parallel columns, as indicated here in red and blue. The nucleoporins in the red column are listed below left, and those in the blue column are listed below right. Almost every nucleoporin in the red column contains a counterpart in the blue column related by approximate position, as well as size and fold arrangement (pairs linked by dashed arrows), strong sequence similarity (those linked by thin arrows), or as duplicate copies with one copy in each column (thick arrows). The membrane rings and FG repeat regions were removed for clarity. Also shown are the *E*-values generated by HMMsearch[57] for the most significant local matches of the corresponding protein pairs; the *E*-value for a sequence match is the expected number of false positives per database search with a score at least as good as this sequence match. **b**, Network of protein contacts in the cytoplasmic (upper row) and nucleoplasmic (lower row) half-spokes for proteins in the red column (left) and blue column (right), showing that the homologous constituents have equivalent neighbours in each column. The networks are shown as instance contact frequency maps[14] (Supplementary Information). Notably, many contacts among proteins in one column are also present between the equivalent proteins in the second column. **c**, A large portion of the NPC can be divided into pairs of structurally similar modules: nuclear and cytoplasmic halves, eight radially disposed spokes, 16 radially disposed columns, and inner and outer rings.

METHODS SUMMARY

Our approach to structure determination can be seen as an iterative series of four steps: data generation by experiment, translation of the data into spatial restraints, calculation of an ensemble of structures by satisfaction of these restraints, and an analysis of the ensemble to produce the final structure. The experimental and computational methods are described in detail in the accompanying paper[14] and Supplementary Information.

Experimental methods. The localization of the PrA-tagged nucleoporins was determined by immuno-electron microscopy of approximately 10,000 gold particles, obtained from pre-embedding labelled nuclear envelopes isolated from tagged strains[6]. Stokes radii were derived from ultracentrifugation velocity gradient sedimentation experiments on the individual nucleoporins and the Nup84 complex. Affinity purifications of all PrA-tagged nucleoporins were performed from solubilized nuclear envelope, highly enriched NPC fractions[6], or lysates produced by whole cell cryolysis. Approximately 20 variants of extraction buffers were used to generate different complexes on IgG Sepharose resins or IgG-magnetic beads. Co-purified proteins were identified by mass spectrometry[7]. Overlay assays were used to monitor direct binding of a nucleoporin in solution to nucleoporins immobilized on nitrocellulose. The structure and composition of the 'Pom rings' (ref. 14) isolated from NPCs were determined by electron microscopy and mass spectrometry.

Computational methods. The structure calculation is expressed as an optimization problem, a solution of which requires three main components: (1) a representation of the assembly in terms of its constituent parts; (2) a scoring function, consisting of individual spatial restraints that encode all the data; and (3) an optimization of the scoring function, which aims to yield structures that satisfy the restraints.

Each nucleoporin was represented by a flexible chain consisting of a small number of connected spheres. The number and radii of the spheres were chosen to reproduce the nucleoporin mass and Stokes radius. The nuclear envelope in the region of the pore membrane[13] was represented by many partially overlapping spheres, each with a radius of 2.25 nm.

The scoring function captures information about the structure of the NPC and is a sum of restraints of various types. These restraints encode what is known about nucleoporin and nuclear envelope excluded volumes (from the protein sequence and ultracentrifugation), nucleoporin positions (from immuno-electron microscopy), nucleoporin 'connectivity' in composites (from affinity purification), nucleoporin–nucleoporin interactions (from affinity purification and overlay assays), complex diameters (from ultracentrifugation), and the eight-fold and two-fold symmetries of the NPC (from electron microscopy).

The optimization protocol combines the techniques of simulated annealing, molecular dynamics, and conjugate gradient minimization. To sample the space of solutions well, independent optimizations of randomly generated initial configurations were performed until 1,000 structures were generated that satisfied all input restraints (that is, the ensemble); ~200,000 trials were required.

The protein localization probability is the probability that a given volume element of a $100 \times 100 \times 100$ grid with the spacing of 1 nm is occupied by a particular protein. It is calculated for each protein from all superposed configurations in the ensemble. Next, for a given protein, the volume elements are sorted by their localization probability values. The localization volume of the protein then corresponds to the top-ranked elements, the volume of which sums to the protein volume, estimated from its molecular mass. The localization volume of a protein reveals its most probable localization.

Received 20 April; accepted 22 October 2007.

1. Lim, R. Y. & Fahrenkrog, B. The nuclear pore complex up close. *Curr. Opin. Cell Biol.* **18**, 342–347 (2006).
2. Macara, I. G. Transport into and out of the nucleus. *Microbiol. Mol. Biol. Rev.* **65**, 570–594 (2001).
3. Weis, K. Nucleocytoplasmic transport: cargo trafficking across the border. *Curr. Opin. Cell Biol.* **14**, 328–335 (2002).
4. Hetzer, M., Walther, T. C. & Mattaj, I. W. Pushing the envelope: Structure, function, and dynamics of the nuclear periphery. *Annu. Rev. Cell Dev. Biol.* **21**, 347–380 (2005).

©2007 Nature Publishing Group

5. Tran, E. J. & Wente, S. R. Dynamic nuclear pore complexes: life on the edge. *Cell* **125,** 1041–1053 (2006).
6. Rout, M. P. *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148,** 635–651 (2000).
7. Cronshaw, J. M., Krutchinsky, A. N., Zhang, W., Chait, B. T. & Matunis, M. J. Proteomic analysis of the mammalian nuclear pore complex. *J. Cell Biol.* **158,** 915–927 (2002).
8. Akey, C. W. & Radermacher, M. Architecture of the *Xenopus* nuclear pore complex revealed by three-dimensional cryo-electron microscopy. *J. Cell Biol.* **122,** 1–19 (1993).
9. Beck, M. *et al.* Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* **306,** 1387–1390 (2004).
10. Hinshaw, J. E., Carragher, B. O. & Milligan, R. A. Architecture and design of the nuclear pore complex. *Cell* **69,** 1133–1141 (1992).
11. Kiseleva, E. *et al.* Yeast nuclear pore complexes have a cytoplasmic ring and internal filaments. *J. Struct. Biol.* **145,** 272–288 (2004).
12. Stoffler, D. *et al.* Cryo-electron tomography provides novel insights into nuclear pore architecture: implications for nucleocytoplasmic transport. *J. Mol. Biol.* **328,** 119–130 (2003).
13. Yang, Q., Rout, M. P. & Akey, C. W. Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. *Mol. Cell* **1,** 223–234 (1998).
14. Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* doi:10.1038/nature06404 (this issue).
15. Krull, S., Thyberg, J., Bjorkroth, B., Rackwitz, H. R. & Cordes, V. C. Nucleoporins as components of the nuclear pore complex core structure and Tpr as the architectural element of the nuclear basket. *Mol. Biol. Cell* **15,** 4261–4277 (2004).
16. Pante, N. & Kann, M. Nuclear pore complex is able to transport macromolecules with diameters of about 39 nm. *Mol. Biol. Cell* **13,** 425–434 (2002).
17. Devos, D. *et al.* Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol.* **2,** e380 (2004).
18. Devos, D. *et al.* Simple fold composition and modular architecture of the nuclear pore complex. *Proc. Natl Acad. Sci. USA* **103,** 2172–2177 (2006).
19. Fotin, A. *et al.* Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* **432,** 573–579 (2004).
20. Stagg, S. M. *et al.* Structure of the Sec13/31 COPII coat cage. *Nature* **439,** 234–238 (2006).
21. ter Haar, E., Musacchio, A., Harrison, S. C. & Kirchhausen, T. Atomic structure of clathrin: a β propeller terminal domain joins an α zigzag linker. *Cell* **95,** 563–573 (1998).
22. Dokudovskaya, S. *et al.* Protease accessibility laddering: a proteomic tool for probing protein structure. *Structure* **14,** 653–660 (2006).
23. Fath, S., Mancias, J. D., Bi, X. & Goldberg, J. Structure and organization of coat proteins in the COPII cage. *Cell* **129,** 1325–1336 (2007).
24. Antonin, W. & Mattaj, I. W. Nuclear pore complexes: round the bend? *Nature Cell Biol.* **7,** 10–12 (2005).
25. Drin, G. *et al.* A general amphipathic α-helical motif for sensing membrane curvature. *Nature Struct. Mol. Biol.* **14,** 138–146 (2007).
26. Conti, E., Muller, C. W. & Stewart, M. Karyopherin flexibility in nucleocytoplasmic transport. *Curr. Opin. Struct. Biol.* **16,** 237–244 (2006).
27. Akey, C. W. Structural plasticity of the nuclear pore complex. *J. Mol. Biol.* **248,** 273–293 (1995).
28. Hinshaw, J. E. & Milligan, R. A. Nuclear pore complexes exceeding eightfold rotational symmetry. *J. Struct. Biol.* **141,** 259–268 (2003).
29. Bryant, D. M. & Stow, J. L. The ins and outs of E-cadherin trafficking. *Trends Cell Biol.* **14,** 427–434 (2004).
30. Strawn, L. A., Shen, T., Shulga, N., Goldfarb, D. S. & Wente, S. R. Minimal nuclear pore complexes define FG repeat domains essential for transport. *Nature Cell Biol.* **6,** 197–206 (2004).
31. Rout, M. P., Aitchison, J. D., Magnasco, M. O. & Chait, B. T. Virtual gating and nuclear transport: the hole picture. *Trends Cell Biol.* **13,** 622–628 (2003).
32. Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L. & Rexach, M. Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl Acad. Sci. USA* **100,** 2450–2455 (2003).
33. Liu, S. M. & Stewart, M. Structural basis for the high-affinity binding of nucleoporin Nup1p to the *Saccharomyces cerevisiae* importin-β homologue, Kap95p. *J. Mol. Biol.* **349,** 515–525 (2005).
34. Peters, R. Translocation through the nuclear pore complex: selectivity and speed by reduction-of-dimensionality. *Traffic* **6,** 421–427 (2005).
35. Ribbeck, K. & Gorlich, D. Kinetic analysis of translocation through nuclear pore complexes. *EMBO J.* **20,** 1320–1330 (2001).
36. Isgro, T. A. & Schulten, K. Binding dynamics of isolated nucleoporin repeat regions to importin-β. *Structure* **13,** 1869–1879 (2005).
37. Isgro, T. A. & Schulten, K. Association of nuclear pore FG-repeat domains to NTF2 import and export complexes. *J. Mol. Biol.* **366,** 330–345 (2007).
38. Stewart, M. Molecular mechanism of the nuclear protein import cycle. *Nature Rev. Mol. Cell Biol.* **8,** 195–208 (2007).
39. Zilman, A., Di Talia, S., Chait, B. T., Rout, M. P. & Magnasco, M. O. Efficiency, selectivity, and robustness of nucleocytoplasmic transport. *PLoS Comput. Biol.* **3,** e125 (2007).
40. Paulillo, S. M. *et al.* Nucleoporin domain topology is linked to the transport status of the nuclear pore complex. *J. Mol. Biol.* **351,** 784–798 (2005).
41. Lim, R. Y. *et al.* Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc. Natl Acad. Sci. USA* **103,** 9512–9517 (2006).
42. Hawryluk-Gara, L. A., Shibuya, E. K. & Wozniak, R. W. Vertebrate Nup53 interacts with the nuclear lamina and is required for the assembly of a Nup93-containing complex. *Mol. Biol. Cell* **16,** 2382–2394 (2005).
43. King, M. C., Lusk, C. P. & Blobel, G. Karyopherin-mediated import of integral inner nuclear membrane proteins. *Nature* **442,** 1003–1007 (2006).
44. Saksena, S., Summers, M. D., Burks, J. K., Johnson, A. E. & Braunagel, S. C. Importin-α-16 is a translocon-associated protein involved in sorting membrane proteins to the nuclear envelope. *Nature Struct. Mol. Biol.* **13,** 500–508 (2006).
45. Aitchison, J. D., Rout, M. P., Marelli, M., Blobel, G. & Wozniak, R. W. Two novel related yeast nucleoporins Nup170p and Nup157p: complementation with the vertebrate homologue Nup155p and functional interactions with the yeast nuclear pore-membrane protein Pom152p. *J. Cell Biol.* **131,** 1133–1148 (1995).
46. Marelli, M., Aitchison, J. D. & Wozniak, R. W. Specific binding of the karyopherin Kap121p to a subunit of the nuclear pore complex containing Nup53p, Nup59p, and Nup170p. *J. Cell Biol.* **143,** 1813–1830 (1998).
47. Siniossoglou, S. *et al.* A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is essential for normal nuclear pores. *Cell* **84,** 265–275 (1996).
48. Wente, S. R., Rout, M. P. & Blobel, G. A new family of yeast nuclear pore complex proteins. *J. Cell Biol.* **119,** 705–723 (1992).
49. Unwin, P. N. & Milligan, R. A. A large particle associated with the perimeter of the nuclear pore complex. *J. Cell Biol.* **93,** 63–75 (1982).
50. Scannell, D. R., Butler, G. & Wolfe, K. H. Yeast genome evolution-the origin of the species. *Yeast*. (in the press).
51. Schledzewski, K., Brinkmann, H. & Mendel, R. R. Phylogenetic analysis of components of the eukaryotic vesicle transport system reveals a common origin of adaptor protein complexes 1, 2, and 3 and the F subcomplex of the coatomer COPI. *J. Mol. Evol.* **48,** 770–778 (1999).
52. Grandi, P. *et al.* A novel nuclear pore protein Nup82p which specifically binds to a fraction of Nsp1p. *J. Cell Biol.* **130,** 1263–1273 (1995).
53. Belgareh, N. *et al.* Functional characterization of a Nup159p-containing nuclear pore subcomplex. *Mol. Biol. Cell* **9,** 3475–3492 (1998).
54. Bailer, S. M. *et al.* Nup116p associates with the Nup82p-Nsp1p-Nup159p nucleoporin complex. *J. Biol. Chem.* **275,** 23540–23548 (2000).
55. Bailer, S. M., Balduf, C. & Hurt, E. The Nsp1p carboxy-terminal domain is organized into functionally distinct coiled-coil regions required for assembly of nucleoporin subcomplexes and nucleocytoplasmic transport. *Mol. Cell. Biol.* **21,** 7944–7955 (2001).
56. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
57. Soding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33,** W244–W248 (2005).

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

Acknowledgements We thank H. Shio for performing the electron microscopic studies; J. Fanghänel, M. Niepel and C. Strambio-de-Castillia for help in developing the affinity purification techniques; M. Magnasco for discussions and advice; A. Kruchinsky for assistance with mass spectrometry; M. Topf, D. Korkin, F. Davis, M. S. Madhusudan, M.-Y. Shen, F. Foerster, N. Eswar, M. Kim, D. Russell, B. Peterson and B. Webb for many discussions about structure characterization by satisfaction of spatial restraints; C. Johnson, S. G. Parker, and C. Silva, T. Ferrin and T. Goddard for preparation of some figures; and S. Pulapura and X. J. Zhou for their help with the design of the conditional diameter restraint. We are grateful to J. Aitchison for discussion and suggestions. We also thank all other members of the Chait, Rout and Sali laboratories for their assistance. We acknowledge support from an Irma T. Hirschl Career Scientist Award (M.P.R.), a Sinsheimer Scholar Award (M.P.R.), a grant from the Rita Allen Foundation (M.P.R.), a grant from the American Cancer Society (M.P.R.), the Sandler Family Supporting Foundation (A.S.), the Human Frontier Science Program (A.S., L.M.V.), NSF (A.S.), and grants from the National Institutes of Health (B.T.C., M.P.R., A.S.), as well as computer hardware gifts from R. Conway, M. Homer, Intel, Hewlett-Packard, IBM and Netapp (A.S.).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.P.R (rout@rockefeller.edu), A.S. (sali@salilab.org), or B.T.C. (chait@rockefeller.edu).

©2007 Nature Publishing Group

LETTERS

Element-selective imaging of atomic columns in a crystal using STEM and EELS

Koji Kimoto[1], Toru Asaka[1], Takuro Nagai[1], Mitsuhiro Saito[1], Yoshio Matsui[1] & Kazuo Ishizuka[1,2]

Microstructure characterization has become indispensable to the study of complex materials, such as strongly correlated oxides, and can obtain useful information about the origin of their physical properties. Although atomically resolved measurements have long been possible, an important goal in microstructure characterization is to achieve element-selective imaging at atomic resolution. A combination of scanning transmission electron microscopy (STEM) and electron energy-loss spectroscopy (EELS)[1,2] is a promising technique for atomic-column analysis. However, two-dimensional analysis has not yet been performed owing to several difficulties, such as delocalization in inelastic scattering or instrumentation instabilities. Here we demonstrate atomic-column imaging of a crystal specimen using localized inelastic scattering and a stabilized scanning transmission electron microscope[3]. The atomic columns of La, Mn and O in the layered manganite $La_{1.2}Sr_{1.8}Mn_2O_7$ are visualized as two-dimensional images.

The STEM technique is promising for characterizing material microstructures. A small electron probe, whose diameter is comparable to the interatomic distance in crystals, can be scanned on a specimen using a modern STEM. STEM annular dark-field (ADF) imaging, which uses high-angle elastic (or phonon) scattering, enables us to observe high-spatial-resolution images, and it has been applied to various materials[4–6]. The other advantage of ADF imaging is its high compositional sensitivity compared with conventional transmission electron microscopy. ADF imaging is however not straightforward for the evaluation of atomic number, because it only detects the intensity of elastic scattering. Other analytical methods that use element-dependent inelastic scattering (such as EELS) should be implemented to distinguish between elements. Using an electron energy-loss spectrum, elemental and chemical information (for example, valence and atomic coordination) can be obtained with high spatial resolution[7,8]. As STEM is compatible with EELS, there have been several reports on EELS and STEM with high spatial resolution[9,10]. However, two-dimensional atomic-column imaging has not been achieved thus far, because of the difficulties described below.

There are several requirements for realizing the atomic-column imaging of crystalline specimens using STEM and EELS: (1) electron optics suitable for a small intense probe; (2) localization in electron probe propagation; (3) localization in inelastic scattering; and (4) practical technical factors in instrumentation, such as the stability of the electron probe position. Recently, progress towards requirement (1) has been made by using a high-brightness electron source or a spherical aberration corrector. Requirement (2) is related to dynamical electron diffraction in crystals. An electron probe focused on the top of an atomic column propagates essentially along the atomic column (channelling). Under certain experimental conditions (for example, a thick specimen or a large convergence angle), most of the intensity of the electron probe is transferred to the adjacent atomic columns during the propagation, resulting in degradation in spatial resolution (dechannelling). Using multislice simulation software (WinHREM from HREM Research), we confirmed that the intensity maximum of the probe was preserved on the original atomic column under the present experimental conditions: a convergence semi-angle of 15 mrad and a specimen thickness of $<$50 nm. Thus, requirements (1) and (2) basically characterize the spatial resolution of both elastic (ADF) and inelastic (EELS) STEM imaging. As demonstrated in ref. 5, these two requirements are almost fulfilled in modern instrumentation, and atomic-column imaging at a resolution of $<$0.1 nm by ADF imaging is currently achievable.

The third requirement, localization in inelastic scattering (3), is the major factor for EELS and has been discussed by a few authors[1,11,12]. This factor can involve several physical descriptions, such as the impact parameter and non-locality, which respectively correspond to the local and non-local approximation[13,14]. The applicability of these approximations is related to the degree of coherence in inelastic scattering. The local approximation corresponds to the incoherent case, in which the inelastic cross-section can be simply described as the product of the intensity of an incident electron and the object function; the latter is essentially characterized by the impact parameter. By contrast, the non-local description involves a given degree of partial coherence. The inelastic cross-section can no longer be obtained by simple multiplication, but must be described using the wavefunction (not the intensity) of an incident electron and the effective non-local potential, which is deduced from the mixed dynamic form factor. The non-locality is considered as the 'uncertain' region, in which the inelastic scattering is partially coherent. Calculations for several experimental configurations have been reported[15] and have shown the difference between the local and non-local approximations. In the present experiment, we have applied a small convergence angle and a large collection angle, which correspond to the experimental configuration in favour of the local approximation. In this case, the small convergence angle regulates the reciprocal area of the mixed dynamic form factor and suppresses non-dipole transitions in inelastic scattering. The large collection angle is effective in reducing the interference fringes of inelastic electrons.

Under the local approximation, delocalization in inelastic scattering is predominantly characterized by the angular distribution of inelastic scattering. A simple formula has been proposed for the delocalization, $0.5\lambda/\theta_E^{3/4}$, where λ denotes the electron wavelength and θ_E the relativistic characteristic angle[7]. Several experimental results[7,12,16] are in agreement with this formula. Scattering with higher energy-loss is relatively localized on the atomic site. The degree of delocalization decreases from a few nanometres to subnanometre order with increasing energy-loss. As the cross-section of the high energy-loss range is small, we frequently measure EELS spectra in a low energy-loss range of $<$1 keV, which corresponds to a delocalization of $>$0.1 nm. Consequently EELS experiments suffer a substantial delocalization effect in comparison with the incident probe size. We also note that there are

[1]National Institute for Materials Science, Tsukuba, Ibaraki 305-0044, Japan. [2]HREM Research, Inc., Higashimatsuyama, Saitama 355-0055, Japan.

©2007 Nature Publishing Group

often more than one core-loss peaks of the target element within the observable energy-loss range. For example, lanthanum shows the N_{45} edge (99 eV) and the M_5 edge (832 eV)[17]; their delocalizations calculated using the above-mentioned formula are 0.57 nm and 0.11 nm, respectively. In this study, we have clarified the delocalization effect by showing the difference between the N_{45} edge and the M_5 edge, and we have realized atomic-column resolution using the M_5 edge.

The last technical requirement (4), exemplified by the stability of the probe position, is critical in practice[18]. We have developed a stabilized STEM instrument (Hitachi High-Technologies, HD-2300C) with many mechanical and electronic modifications to the original commercial product[3]. We installed the STEM system in an environment-controlled room of a specialized building, which reduces the mechanical vibration of the floor from any adjacent disturbance. The position instability of the STEM electron probe is found to be 0.12 nm min^{-1}, which is about one-tenth that in conventional instrumentation. The spherical aberration coefficient C_s of the objective lens is 0.57 mm and the probe current about 7 pA. The calculated probe size is 0.10 nm (full-width at half-maximum, FWHM) at a defocus of −50 nm (1.3 times the Scherzer focus, $-(C_s\lambda)^{1/2}$). By using the Fourier transform of the ADF images, the spatial resolution of the equipment[3] was estimated to be 0.11 nm. The inner and the outer collection semi-angles of the ADF detector are 36 and 191 mrad, respectively. These optical conditions were applied for both ADF imaging and EELS. Two-dimensional spectrum acquisition[2] was performed using spectrum imaging software and an EELS spectrometer (GIF Tridiem, Gatan). To retain the accuracy of the measurement, the specimen drift during the two-dimensional acquisition was measured and corrected every 30 s using ADF images. The collection semi-angle for EELS was set to a relatively large value of 31 mrad to improve collection efficiency and to reduce the extent of delocalization.

We observed a layered perovskite manganite, $La_{1.2}Sr_{1.8}Mn_2O_7$, which is a typical strongly correlated material[19–21] (Fig. 1a). A single crystal made by a floating-zone method was used. An STEM specimen was prepared by mechanical thinning and argon ion milling at 4 kV. The specimen thickness of the observed area was evaluated to be about 30 nm using the EELS log-ratio method[7].

Figure 1b shows an ADF image along the [010] direction, showing the areas for two-dimensional EELS and spatial-drift correction. The contrast of the ADF image reflects the atomic number of each atomic column. A-site columns (La and Sr, yellow-green spheres in Fig. 1a) are observed as bright spots, and B-site columns (Mn, blue spheres) as relatively dark spots; oxygen columns (red spheres) are not visible. The six arrows in Fig. 1a and in Fig. 1b indicate A-site positions. Note that there are two different A-sites from the crystallographic viewpoint, as indicated by a yellow arrow p (in the perovskite block) and a white arrow r (in the rock-salt layer)[22]. The ADF signal on the yellow arrow p is higher than that on the white arrow r. This suggests that the average atomic number differs between the two crystallographic A-sites (p and r). This can be analysed by atomic-column EELS.

Two-dimensional EELS was performed on a rectangular area, in which 24 pixels in width and 61 pixels in height were analysed—that is, 1,464 electron energy-loss spectra were acquired. An electron probe about 0.1 nm in FWHM was scanned in steps of 0.035 nm. The dwell time on each pixel and the total acquisition time including the drift correction were 2 s and 61 min, respectively. Figure 1c shows the summed spectrum of the analysed area. The core-loss edges of La N_{45}, O K, Mn $L_{3,2}$ and La $M_{5,4}$ are observed. Strontium does not show a clear core-loss peak in this energy range[17]. We evaluated each core-loss intensity after background subtraction, as shown in the filled areas in Fig. 1c.

Figure 2 shows the results of the two-dimensional EELS: the ADF image overlapped with the crystal structure model (Fig. 2a), the calculated incident probe (Fig. 2b), and the core-loss images of La N_{45}, O K, Mn L and La M_5, respectively (Fig. 2c–f). In spite of the high intensity of La N_{45}, its core-loss image (Fig. 2c) does not show significant contrast. This is due to the large delocalization in inelastic scattering, because the delocalization effect of La N_{45} excitation (0.57 nm) is larger than the lanthanum interatomic distances (0.37–0.39 nm) in the crystal. To resolve the shortest inter-column distance (Mn–O column to O column, 0.194 nm), a core-loss signal of more than 420 eV should be used. In contrast with the La N_{45} core-loss image, the O K (Fig. 2d), Mn $L_{3,2}$ (Fig. 2e) and La $M_{5,4}$ (Fig. 2f) core-loss images show column-by-column contrast. The bright dots in each core-loss image correctly correspond to each atomic column. Note that oxygen columns are also visible, although the ADF image does not show bright dots on the oxygen columns. As the contrasts in the core-loss images (see the minimum and maximum values in Fig. 2d–f) are larger than the variation in the total spectrum intensity (for example, about 5%), the contrasts are not due to elastic scattering or to absorption on each atomic site.

In a pioneering study by Suenaga *et al.*[23], a single lanthanoid atom (Gd) embedded in nano-peapods was detected using STEM and EELS. To enhance the detectability, these authors used a 0.5 nm probe and the N_{45} core-loss signal (150 eV), and observed a gadolinium atom approximately 0.5–1 nm in diameter. To resolve atomic columns in the crystal, however, we minimized the probe size (about 0.1 nm) and the degree of delocalization in inelastic scattering using the M_5 edge. Very recently, Bosman *et al.*[24] also reported EELS analysis using a spherical aberration corrector, in which the signal-to-noise ratio of the results is improved. A sub-angstrom probe is effective for STEM-based techniques; however, it should be pointed out that the diameters of the observed atomic columns that we observed (Fig. 2d–f) are larger than that of the incident probe (Fig. 2b). This suggests that the incident probe size is not the predominant limiting factor of spatial resolution

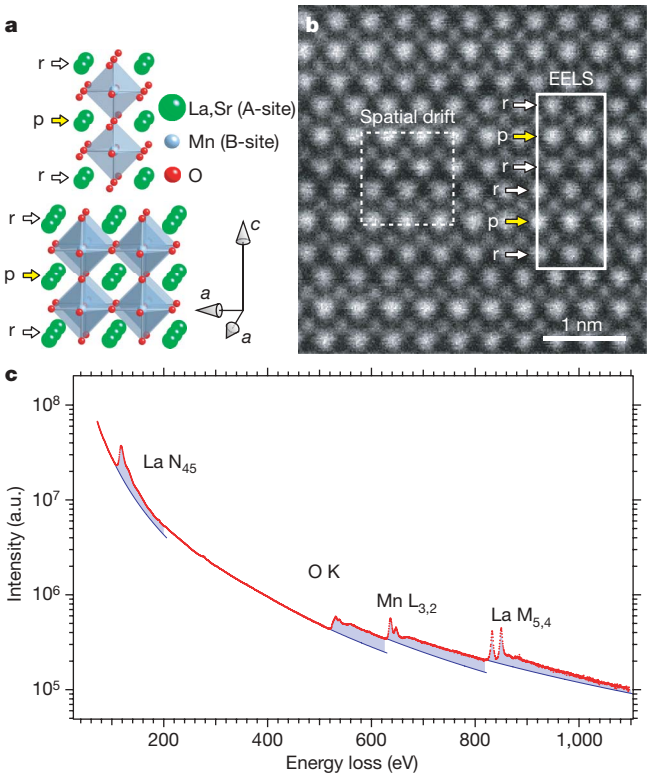


Figure 1 | Conventional STEM-EELS observations. **a**, Crystal structure of layered perovskite manganite, $La_{1.2}Sr_{1.8}Mn_2O_7$. Yellow-green spheres correspond to A-sites (La and Sr), blue spheres to B-sites (Mn) and red spheres to oxygen. There are two different crystallographic A-sites in the perovskite block (yellow arrow p) and the rock-salt layer (white arrow r). **b**, ADF image of the specimen observed along the [010] direction. The areas for two-dimensional EELS and drift measurement are shown by rectangles. **c**, EELS spectrum acquired from the rectangular area for the two-dimensional EELS. The core-loss intensity of each element (blue areas) is visualized as a core-loss image (Fig. 2c–f).

©2007 Nature Publishing Group

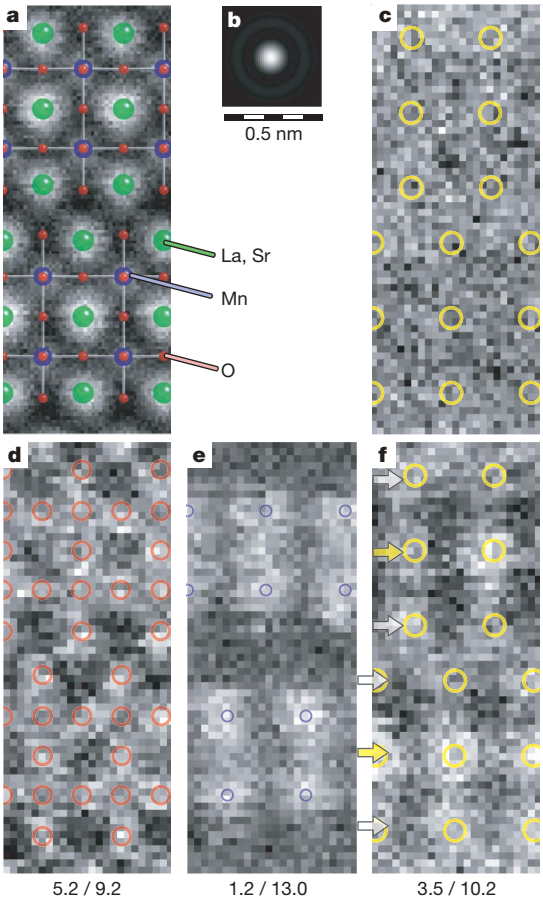


Figure 2 | Atomic-column imaging using STEM and EELS. a, Enlarged ADF image of analysed area overlapped with crystal structure. **b**, Calculated incident probe for EELS and ADF observations. **c–f**, Core-loss images of La N_{45}, O K, Mn $L_{3,2}$ and La $M_{5,4}$. Numbers below panels **d–f** indicate the minimum/maximum of the display range; all values correspond to the number of detected electrons after multiplying by 10^4. Coloured circles in panels **c–f** indicate the corresponding positions of atomic columns.

in the present atomic-column imaging. Although a small incident probe with a large convergence angle is suitable for resolving atomic columns under the classical local approximation, a large convergence angle yields several difficulties in terms of dechannelling and non-locality. It should be emphasized that the essence of the present experimental configuration for atomic column imaging is not to simply make a small incident probe, but to optimize the acquisition condition associated with the delocalization in elastic and inelastic scattering. The present experimental result indicates that incoherent EELS imaging, which allows us to intuitively interpret the core-loss image, is achievable under appropriate experimental conditions.

The core-loss images obtained are informative in materials science. For example, in the lanthanum M_5 core-loss map (Fig. 2f), the intensity on the A-sites (the arrows labelled p) in the perovskite block is higher than that in the rock-salt layer. This is consistent with the above-mentioned ADF image and an X-ray diffraction result[22]. The observation of this kind of local inhomogeneity is important, because it could become a seed for a phase transition or magnetic domain-wall pinning in strongly correlated materials. Thus, this method enables us to discuss the local crystal structure and its material properties beyond the stoichiometric understanding of the average crystal structure. Although the ultimate goal is the quantitative chemical analysis of each atomic column, there are several theoretical and experimental challenges. For example, the dechannelling, delocalization and absorption in electron scattering should be quantitatively evaluated on each atomic column. In addition, we have to eliminate the surface amorphous layer produced by ion milling for STEM specimen preparation, because the surface 'mixing' layer yields a continuous background in atomic-resolution elemental/chemical maps.

In summary, our results clearly demonstrate the atomic-column resolution of EELS spectrum imaging. Using STEM and EELS, elemental and chemical analyses of, for example, site occupancy are possible; such analyses are the next stage of material microstructure characterization.

Received 25 July; accepted 8 October 2007.
Published online 28 October 2007.

1. Pennycook, S. J. & Nellist, P. D. in *Impact of Electron and Scanning Probe Microscopy on Materials Science* (eds Rickerby, D. G., Valdre, G. & Valdre, U.) 161–207 (Kluwer Academic, London, 1999).
2. Jeanguillaume, C. & Colliex, C. Spectrum-image: The next step in EELS digital acquisition and processing. *Ultramicroscopy* **28**, 252–257 (1989).
3. Kimoto, K., Nakamura, K., Aizawa, S., Isakozawa, S. & Matsui, Y. Development of dedicated STEM with high stability. *J. Electron Microsc.* **56**, 17–20 (2007).
4. Pennycook, S. J. & Boatner, L. A. Chemically sensitive structure-imaging with a scanning transmission electron microscope. *Nature* **336**, 565–567 (1988).
5. Batson, P. E., Dellby, N. & Krivanek, O. L. Sub-ångstrom resolution using aberration corrected electron optics. *Nature* **418**, 617–620 (2002).
6. Muller, D. A., Nakagawa, N., Ohtomo, A., Grazul, J. L. & Hwang, H. Y. Atomic-scale imaging of nanoengineered oxygen vacancy profiles in $SrTiO_3$. *Nature* **430**, 657–661 (2004).
7. Egerton, R. F. *Electron Energy-Loss Spectroscopy in the Electron Microscope* 2nd edn (Plenum, New York, 1996).
8. Colliex, C. *International Tables for Crystallography* Vol. C (eds Wilson, A. J. C. & Prince, E.) 387–408 (Kluwer Academic, London, 1999).
9. Batson, P. E. Simultaneous STEM imaging and electron energy-loss spectroscopy with atomic-column sensitivity. *Nature* **366**, 727–728 (1993).
10. Varela, M. *et al.* Spectroscopic imaging of single atoms within a bulk solid. *Phys. Rev. Lett.* **92**, 095502 (2004).
11. Howie, A. Inelastic scattering of electrons by crystals I. The theory of small-angle inelastic scattering. *Proc. R. Soc. Lond. A* **271**, 268–287 (1963).
12. Muller, D. A. & Silcox, J. Delocalization in inelastic scattering. *Ultramicroscopy* **59**, 195–213 (1995).
13. Oxley, M. P., Cosgriff, E. C. & Allen, L. J. Nonlocality in imaging. *Phys. Rev. Lett.* **94**, 203906 (2005).
14. Oxley, M. P. *et al.* Interpreting atomic-resolution spectroscopic images. *Phys. Rev. B* **76**, 064303 (2007).
15. Allen, L. J., Findlay, S. D., Oxley, M. P., Witte, C. & Zaluzec, N. J. Modelling high-resolution electron microscopy based on core-loss spectroscopy. *Ultramicroscopy* **106**, 1001–1011 (2006).
16. Kimoto, K. & Matsui, Y. Experimental investigation of phase contrast formed by inelastically scattered electrons. *Ultramicroscopy* **96**, 335–342 (2003).
17. Ann, C. C. *Transmission Electron Energy Loss Spectrometry in Materials Science and the EELS Atlas* (John Wiley & Sons, New York, 2004).
18. Muller, D. A. & Grazul, J. Optimizing the environment for sub-0.2 nm scanning transmission electron microscopy. *J. Electron Microsc.* **50**, 219–226 (2001).
19. Kimura, T., Tomioka, Y., Asamitsu, A. & Tokura, Y. Anisotropic magnetoelastic phenomena in layered manganite crystals: Implication of change in orbital state. *Phys. Rev. Lett.* **81**, 5920–5923 (1998).
20. Kimura, T. & Tokura, Y. Layered magnetic manganites. *Annu. Rev. Mater. Sci.* **30**, 451–474 (2000).
21. Asaka, T. *et al.* Observation of magnetic ripple and nanowidth domains in a layered ferromagnet. *Phys. Rev. Lett.* **95**, 227204 (2005).
22. Battle, P. D. *et al.* Layered Ruddlesden-Popper manganese oxides: synthesis and cation ordering. *Chem. Mater.* **9**, 552–559 (1997).
23. Suenaga, K. *et al.* Element-selective single atom imaging. *Science* **290**, 2280–2282 (2000).
24. Bosman, M. *et al.* Two-dimensional mapping of chemical information at atomic resolution. *Phys. Rev. Lett.* **99**, 086102 (2007).

Acknowledgements We thank Y. Tokura and T. Kimura for providing the manganite specimen. We thank R. F. Egerton, M. Malac, K. Suenaga and H. Kurata for discussions and encouragement. We also thank M. Barfels, M. Kundmann and J. Hunt (Gatan) for their cooperation in the installation of an EEL spectrometer, and T. Yokosawa (NIMS), K. Nakamura, S. Aizawa and S. Isakozawa (Hitachi High-Technologies) for their cooperation in developing the STEM. This work is partly supported by JST-CREST, a Nanotechnology Support Project by MEXT, Japan, and by a Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.K. (kimoto.koji@nims.go.jp).

©2007 Nature Publishing Group

LETTERS

Visualizing single-molecule diffusion in mesoporous materials

Andreas Zürner[1]*, Johanna Kirstein[1]*, Markus Döblinger[1], Christoph Bräuchle[1] & Thomas Bein[1]

Periodic mesoporous materials formed through the cooperative self-assembly of surfactants and framework building blocks can assume a variety of structures[1–3], and their widely tuneable properties make them attractive hosts for numerous applications[4–7]. Because the molecular movement in the pore system is the most important and defining characteristic of porous materials[8], it is of interest to learn about this behaviour as a function of local structure. Generally, individual fluorescent dye molecules can be used as molecular beacons with which to explore the structure of—and the dynamics within—these porous hosts[9–13], and single-molecule fluorescence techniques provide detailed insights into the dynamics of various processes, ranging from biology[14,15] to heterogeneous catalysis[16]. However, optical microscopy methods cannot directly image the mesoporous structure of the host system accommodating the diffusing molecules, whereas transmission electron microscopy provides detailed images of the porous structure[17], but no dynamic information. It has therefore not been possible to 'see' how molecules diffuse in a real nanoscale pore structure. Here we present a combination of electron microscopic mapping and optical single-molecule tracking experiments to reveal how a single luminescent dye molecule travels through linear or strongly curved sections of a mesoporous channel system. In our approach we directly correlate porous structures detected by transmission electron microscopy with the diffusion dynamics of single molecules detected by optical microscopy. This opens up new ways of understanding the interactions of host and guest.

For the implementation of our approach, key sample requirements include extremely thin mesoporous films on electron-transparent substrates, the inclusion of different types of markers, and optical transparency. We spin-coated a solution with a mixture of tetraethyl orthosilicate, Brij 56, HCl, ethanol and water (in a molar ratio of 1:0.144:0.06:61.69:210), resulting in thin films, measured by ellipsometry to be 100 nm thick. The corresponding two-dimensional grazing-incidence small-angle X-ray scattering (2D-GISAXS) pattern shows a two-dimensional hexagonal arrangement of mesoporous silica (Fig. 1a; for details see Methods). Strongly fluorescent terrylene diimide (TDI) molecules[18,19], acting as molecular beacons for single-molecule tracking, polystyrene beads (280 nm in diameter) as markers that are visible in both transmission electron microscopy (TEM) and optical microscopy, and gold colloids (5 nm in diameter) for merging several electron micrographs were added to the synthesis solutions of the mesoporous films. The dye molecules were incorporated into the pores during the evaporation-induced self-assembly of the material[20].

For the overlay of optical and electron microscopic images, we used Si_3N_4 membranes 30 nm in thickness supported by a small silicon wafer with a 500 μm × 500 μm window (Fig. 1b, c). The mesoporous film was spin-coated into the interior of the window. Scanning electron micrographs (Supplementary Fig. 1) show that the polystyrene beads are embedded in the mesoporous film. We emphasize that no further sample preparation, such as ion milling, was needed for electron microscopy at low and high resolution or for the optical wide-field measurements. The whole membrane was mapped with successive white-light transmission images (Supplementary Fig. 2a; an enlargement is shown in Fig. 1d) before single-molecule fluorescence tracking. For the subsequent analysis by TEM, the whole membrane was first mapped at low resolution (Supplementary Fig. 2b), and then at higher magnification to distinguish between single polystyrene beads (Fig. 1e). In this map we were able to retrieve the same geometric arrangement of polystyrene beads as in the white-light images (Fig. 1f, g).

Figure 2 demonstrates the spatial correlation between the single-molecule trajectories (movie of 1,000 images, 200 ms per frame; Fig. 2b) and TEM (Fig. 2c). To overlay the trajectories on the TEM images, the positions of the polystyrene beads were determined from the same movie (red crosses in Fig. 2b). In the area of the tracked trajectories, 400 (20 × 20) TEM images were taken at ×40,000 magnification and were merged by using image cross-correlation and gold particle markers to obtain a map of an area of about 3.5 μm × 3.5 μm (Fig. 2c). From the geometric relationship of the polystyrene beads we fitted the scaling factor and rotating angle for overlaying the two images by the use of a least-square algorithm (Fig. 2d). In this procedure the bead positions in the TEM images were used as fixed references, and the wide-field images were scaled, rotated and translated during the fitting procedure. A more detailed description of the overlay procedure is given in Methods. As an example the bead positions and the deviations of the positions in the TEM and wide-field images of Fig. 2 are tabulated in Supplementary Table 1. We take the average deviation for the position of the polystyrene beads in both TEM and wide-field imaging as a measure for the accuracy of the overlay. It amounts to 26 nm for the overlaid images in Fig. 2d, calculated from the values in the last column of Supplementary Table 1. From the correlation of five different 3.5 μm × 3.5 μm TEM maps with the respective tracking data, we calculated the deviation of the bead positions in TEM and wide-field images to be 29 ± 13 nm (mean ± s.d.).

To show the orientation of the pores in the TEM images better, fast Fourier transformation (FFT) was applied to adjacent square regions of 133 nm × 133 nm (Fig. 2a). The results of the FFT are plotted as black lines. The orientation of each 'director' depicts the average orientation of the pores in the square region around it, and the line thickness is a measure for the intensity of the maxima in the FFT fit and thus for the degree of structural order in this region. In addition to the orientation of the channels, these directors give a good overview of the domain size and domain borders (Fig. 2a, c, d). The enlargement in Fig. 2e directly shows the channels and the

[1]Department of Chemistry and Biochemistry and Center for NanoScience (CeNS), University of Munich, Butenandtstrasse 5-13 (E), D-81377 Munich, Germany.
*These authors contributed equally to this work.

©2007 Nature Publishing Group

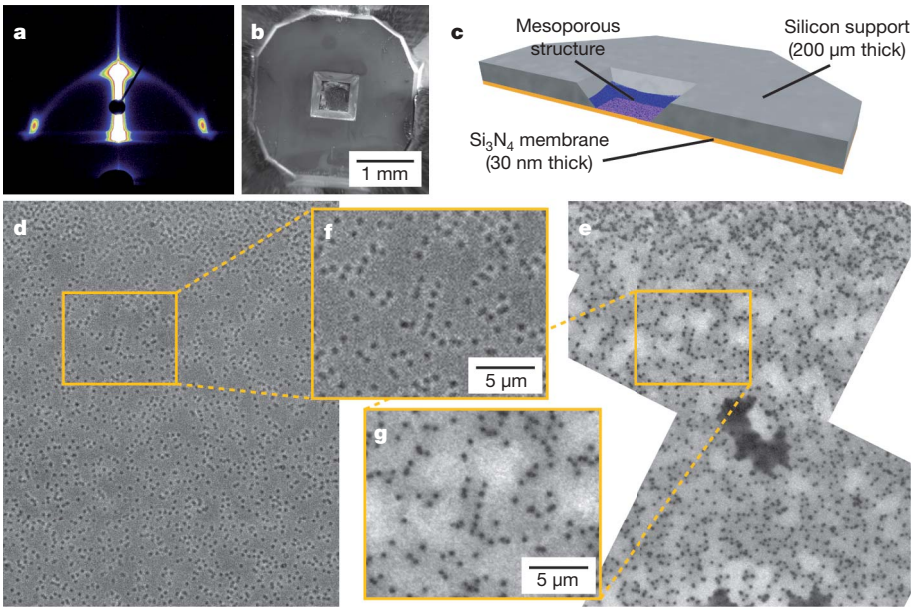


Figure 1 | Sample systems. **a**, A 2D-GISAXS pattern of the mesoporous film on a silicon wafer. **b**, Scanning electron micrograph of a coated Si_3N_4 membrane window on the silicon support. **c**, Diagram of a coated membrane. **d**, White-light transmission image of the region of interest. **e**, TEM image of the region of interest. **f**, **g**, The same characteristic formation of polystyrene beads shown by white-light transmission (**f**) and by TEM (**g**).

corresponding FFT directors, as well as the trajectories of the dye molecules along the channels.

The combination of the two techniques provides the first direct proof that the molecular diffusion pathway through the pore system correlates with the pore orientation of the two-dimensional hexagonal structure. In addition, the influence of specific structural features of the host on the diffusion behaviour of the guest molecules can be clearly seen. With this approach we can uncover, in unprecedented detail, how a single fluorescent dye molecule travels through linear or strongly curved sections of the hexagonal channel system in a thin film of mesoporous silica (as sketched in Fig. 3a, b), how it changes speed in the channel structure, and how it bounces off a domain boundary with a different channel orientation (Fig. 3c). Furthermore, we can show how molecular travel is stopped at a less ordered region (Fig. 3d), or how lateral motions between 'leaky' channels allow a molecule to explore different parallel channels within an otherwise well-ordered periodic structure (Fig. 3e). Figure 4 depicts several examples of specific pore ordering, indicated by the FFT directors, in combination with the overlaid trajectories, showing the various structural features that are shown in Fig. 3. The positioning error for the single molecule trajectories is shown by the light blue boxes, which indicate the standard deviation of the fitted (x,y) positions. It is in the range of only 10–20 nm; the molecular positions can therefore be assigned to an ensemble of about five to ten parallel channels.

Figure 4a depicts a trajectory of a molecule that is following the porous system along different structural domains. The corresponding wide-field movie and the tracked trajectory of the single molecule overlaid with the TEM images are provided in Supplementary Movies 1 and 2. In this case the average deviation of the overlaid images is 34 nm (see Methods and Supplementary Information).

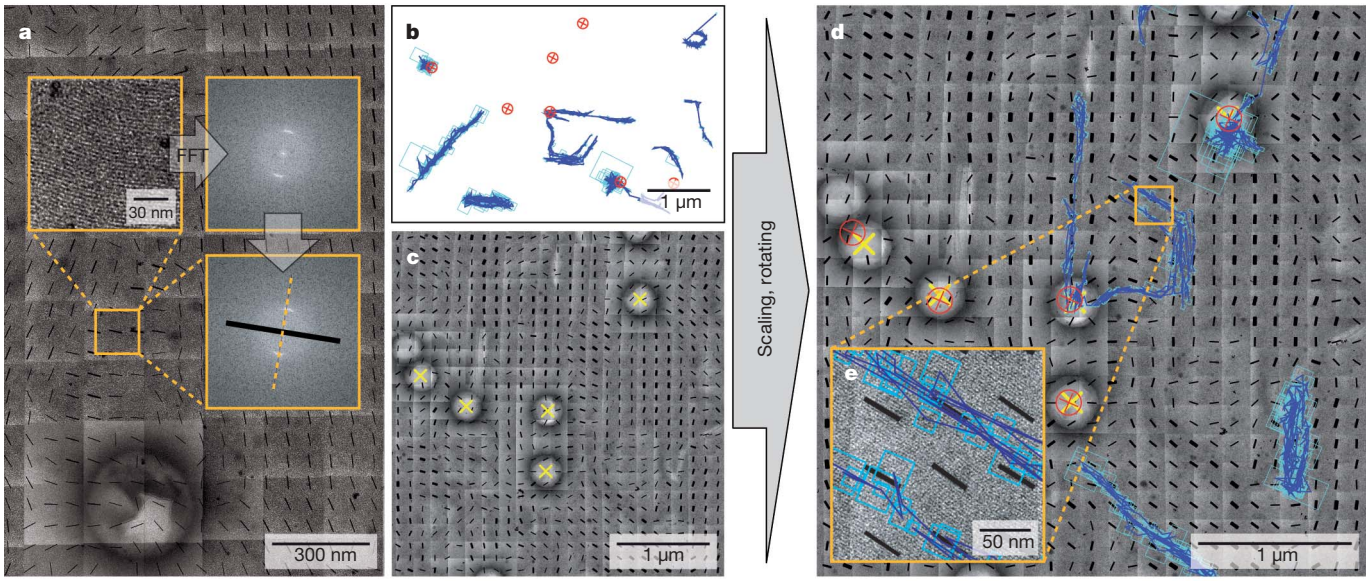


Figure 2 | Merging of the single-molecule trajectories and the TEM micrographs. **a**, The merged TEM map is redivided into equal-sized squares that are processed by FFT as shown in the insets. The direction of the channels is perpendicular to a straight line through the maxima. The thickness of the black directors, pointing along the pore direction, correlates with the structural quality. **b**, Diffusion pathways of single molecules through the pore system (blue trajectories) and positions of the polystyrene beads (red crosses). **c**, The same area as in **b** is imaged by TEM, with lines indicating the direction of the channels (FFT directors) and yellow crosses marking the centres of the polystyrene beads. **d**, Final overlay of trajectories on TEM images; this was obtained by fitting to the best overlay of the polystyrene bead positions. **e**, Enlargement of a region in **d**, showing the trajectories running along the channels.

©2007 Nature Publishing Group

The molecule in Fig. 4a and Supplementary Movie 2 diffuses along linear pores in the middle part of the trajectory and follows the curvature on the right and on the left side. In addition, it turns at a domain boundary, as shown in more detail in Fig. 4b. The insets in the upper right corners of Fig. 4a, b show the different structural elements of this trajectory that were sketched in Fig. 3. To analyse the diffusion along this trajectory we projected the data points onto a manually defined backbone of the trajectory. Then we plotted the averaged mean-square displacement $\langle r^2 \rangle$ along the backbone as a function of time. The linear relation of $\langle r^2 \rangle$ with time was fitted in accordance with the Einstein–Smoluchowski equation for one-dimensional (1D) diffusion for the first few time lags[21], resulting in a 1D diffusion coefficient D_{1D} of $(3.2 \pm 0.1) \times 10^{-2}\,\mu m^2\,s^{-1}$. It is important to analyse the 1D diffusion along the backbone, because the r^2 in the curved region would otherwise tend towards smaller values for longer time lags.

Figure 4c shows another example of a molecule faithfully following the pores and mapping out specific elements of the host structure (see also Supplementary Movie 3). Here the average deviation of the overlay is 48 nm. In this case only one single bead and two pairs of closely neighbouring beads could be used for the overlay, which explains this relatively high value (see Methods). The molecule was blinking several times for intervals of up to 6.6 s, and the trajectory is therefore divided into 11 parts, with durations ranging from 0.6 s to 11.8 s, shown in different colours. The perfect overlay of the S-shaped trajectory on the pore system is shown well by the FFT directors. In Fig. 4d–f the specific regions marked with rectangular boxes in Fig. 4c are shown, and the insets sketch the different behaviours at specific regions of the trajectory (as in Fig. 3). Although the resolution of the optical microscope was not high enough to assign the position of a molecule to one specific channel, the width of the middle of this trajectory (magnified in Fig. 4d) shows that the molecule was moving in different straight parallel channels of the same domain. Especially interesting is the short element at the end of the yellow trajectory where the molecule makes a U-turn and diffuses back in a parallel track that is different from the original track, as shown at the left of Fig. 4d. On the basis of the spatial resolution of the optical pathways, we conclude that the molecule is diffusing in different parallel pores. At the left end of the trajectory in Fig. 4c the molecule diffuses in a well-ordered straight structure and bounces back repeatedly from an amorphous region having no apparent pore ordering, as shown in Fig. 4e. At the right of Fig. 4c a domain boundary region is visible; this structure is shown in more detail in Fig. 4f. Here the molecule bounces back from the domain boundary with channels having different orientations, as sketched in Fig. 3c and in the inset. The orange lines in Fig. 4f are situated slightly above the region of linear channels. However, if the average deviation of the overlay of 48 nm is taken into account it is highly probable that the molecule was actually diffusing along the linear structure in the bottom region of the figure. Furthermore, one should keep in mind that we are sampling diffusion at discrete points in time and space. The connecting lines are just a method of visualizing the trajectories; they do not represent the molecules' exact path.

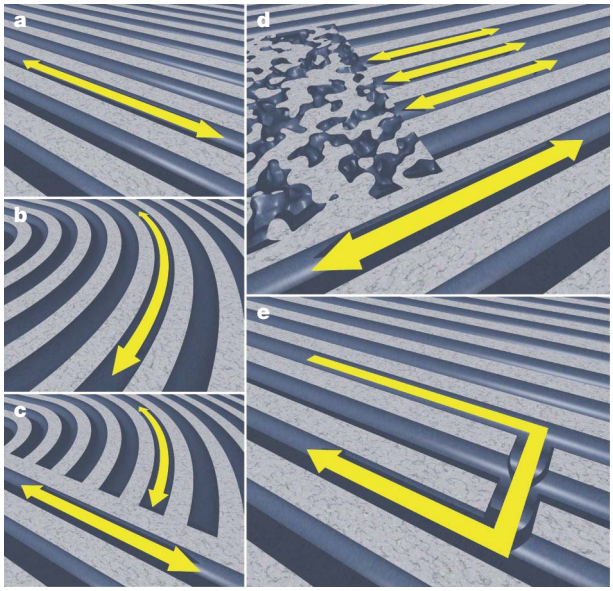


Figure 3 | Structural elements found in real two-dimensional hexagonal mesoporous silica film. **a**, **b**, Straight (**a**) and curved (**b**) segments. **c**, Domain boundaries forcing molecules to turn back. **d**, Molecular travel stopped at less ordered regions. **e**, Lateral motion between neighbouring channels.

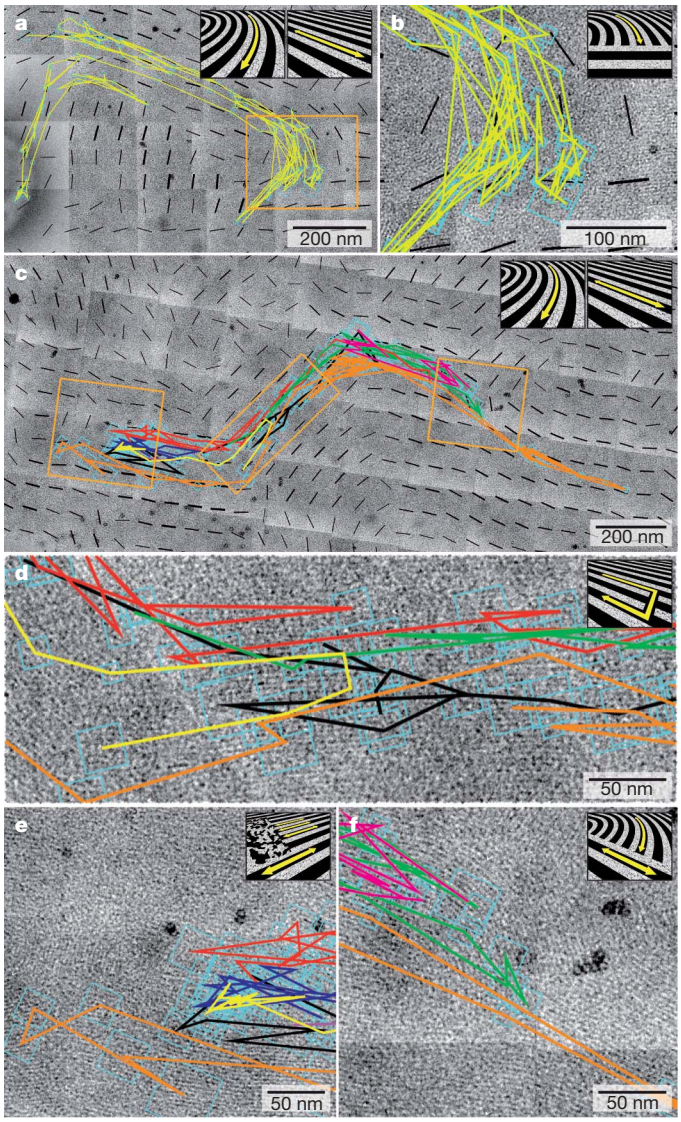


Figure 4 | Structural elements and molecular trajectories found in real two-dimensional hexagonal mesoporous silica film. **a**, Molecule exploring regions of parallel channels, strongly curved regions and domain boundaries. **b**, Magnified area from **a** showing the domain boundary at which the molecule turns back. **c**, Trajectory of another molecule, showing an S shape corresponding to the underlying channel structure. The trajectory is divided into several sections as a result of blinking of the fluorescent molecule; the sections are plotted in different colours. **d**, Lateral motions between 'leaky' channels (yellow trajectory). **e**, Area in which the molecular movement is stopped at a less ordered region. **f**, Forcing of the fluorescent molecule to turn back (green pathway) at a domain boundary. In all panels the light blue boxes depict the standard deviation of the fit to the single-molecule signals; that is, the positioning accuracy. Movies showing the diffusion of the single molecules in **a** and **c** are available as Supplementary Movies 1–3.

©2007 Nature Publishing Group

Projection of the trajectory shown in Fig. 4c onto its backbone and calculation of the average diffusion coefficient along the backbone gives $D_{1D} = (3.0 \pm 0.1) \times 10^{-2}\,\mu m^2\,s^{-1}$. In addition, we calculated the following individual average diffusion coefficients for the parts of the trajectory showing no blinking for at least 20 frames: red (frames 1–33), $D_{1D} = (3.6 \pm 0.2) \times 10^{-2}\,\mu m^2\,s^{-1}$; green (frames 37–82), $D_{1D} = (1.1 \pm 0.1) \times 10^{-2}\,\mu m^2\,s^{-1}$; blue (frames 100–119), $D_{1D} = (1.0 \pm 0.1) \times 10^{-2}\,\mu m^2\,s^{-1}$; yellow (frames 135–157), $D_{1D} = (2.0 \pm 0.1) \times 10^{-2}\,\mu m^2\,s^{-1}$; orange (frames 176–244), $D_{1D} = (5.5 \pm 0.1) \times 10^{-2}\,\mu m^2\,s^{-1}$; and violet (frames 308–328), $D_{1D} = (1.3 \pm 0.1) \times 10^{-2}\,\mu m^2\,s^{-1}$. Parts of the trajectory with fewer than 20 points are drawn in black. The noticeable changes of the 'local' diffusion coefficient show that the diffusion behaviour of the molecules is strongly dependent on the surrounding pore structure. When the molecule is 'bouncing back' at an unstructured region, as in the dark blue part of the trajectory, D_{1D} is smaller than in regions where the molecule is diffusing along the trajectory (orange and red). On average the values are in the same range as the diffusion coefficient for the U-shaped trajectory shown in Fig. 4a and also many other trajectories observed in the same sample (not shown).

In this study we correlate directly the dynamic information from the diffusion trajectories of single guest molecules with the detailed structure of the porous host in regions up to 3.5 μm × 3.5 μm in size. This approach reveals the nature of the real-space porous defect structure, as detected by the movement of fluorescent single molecules, containing linear or strongly curved sections of the hexagonal channel system, domain boundaries, boundaries between ordered and disordered sections, and leaky channels that permit lateral travel. We emphasize that our approach provides detailed information on the real defect structure of porous materials with high spatial resolution that cannot be obtained by scattering methods because of the intrinsic averaging effects. Similarly, we also obtain highly resolved dynamic information in real time that cannot be extracted from conventional diffusion techniques because of ensemble averaging. This new methodology is expected to provide detailed insights into the real structure and dynamics of other classes of porous materials and important host–guest systems, such as bioactive molecules in porous materials for drug delivery or reactants in porous catalysts.

METHODS SUMMARY

Synthesis. First, 2.08 g (0.01 mol) of tetraethyl orthosilicate (Sigma-Aldrich) were mixed with 3 g of 0.2 M HCl, 1.8 g of water and 7.9 g of ethanol and heated at 60 °C for 1 h to accomplish acid-catalysed hydrolysis–condensation of the silica precursor. This solution was mixed with a second solution containing 600 mg of Brij 56 and 12.5 g of ethanol. Finally, 75 μl of this mixture were combined with 4 μl of a highly diluted TDI solution ($\sim 10^{-8}\,mol\,l^{-1}$ in ethanol), 20 μl of a gold colloid solution (diameter 5 nm, concentration corresponding to ~0.01% $HAuCl_4$; Sigma-Aldrich), 15 μl of a polystyrene bead solution (Polybead microspheres, diameter 0.281 ± 0.014 μm, 2.6% solids (latex); Polysciences) and 40 μl of deionized water. The samples were prepared in dry air by spin-coating at 3,000 r.p.m. and then analysed as synthesized.

Ellipsometry. Ellipsometry measurements were performed on a Woollam ESM-300 ellipsometer.

TEM. TEM images were obtained with a JEOL 2011 transmission electron microscope operated at 200 kV by using the software-controlled mapping function of the charge-coupled device camera (model TVIPS F114).

2D-GISAXS. 2D-GISAXS was performed at beamline BW4 (HASYLAB, Hamburg).

Wide-field microscopy and single-particle tracking. Fluorescence images were recorded with a wide-field setup as described in Methods and in more detail in refs 9, 11. The fluorescent dye molecules were excited with a HeNe laser at 633 nm, and movies of 1,000 images were recorded to follow their diffusion. The examples presented here were measured with an exposure time of 200 ms per image. The individual molecule patterns are fitted by a gaussian function

$$f(x,y,A,w) = Ae^{-\left(\frac{x-x0}{w}\right)^2} e^{-\left(\frac{y-y0}{w}\right)^2}$$

with a positioning accuracy of down to 10 nm, where A and w are the amplitude and the width of the gaussian curve, respectively.

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

Received 13 August; accepted 18 October 2007.

1. Beck, J. S. *et al.* A new family of mesoporous molecular sieves prepared with liquid-crystal templates. *J. Am. Chem. Soc.* **114,** 10834–10843 (1992).
2. Sen, T., Tiddy, G. J. T., Casci, J. L. & Anderson, M. W. Synthesis and characterization of hierarchically ordered porous silica materials. *Chem. Mater.* **16,** 2044–2054 (2004).
3. Zhao, D. *et al.* Triblock copolymer syntheses of mesoporous silica with periodic 50 to 300 angstrom pores. *Science* **279,** 548–552 (1998).
4. Dag, Ö., Ozin, G. A., Yang, H., Reber, C. & Bussière, G. Photoluminescent silicon clusters in oriented hexagonal mesoporous silica film. *Adv. Mater.* **11,** 474–480 (1999).
5. Davis, M. E. Ordered porous materials for emerging applications. *Nature* **417,** 813–821 (2002).
6. Shen, J. L. *et al.* Photoluminescence sites on MCM-48. *Micropor. Mesopor. Mater.* **64,** 135–143 (2003).
7. Yang, C.-M., Cho, A.-T., Pan, F.-M., Tsai, T.-G. & Chao, K.-J. Spin-on mesoporous silica films with ultralow dielectric constants, ordered pore structures, and hydrophobic surfaces. *Adv. Mater.* **13,** 1099–1102 (2001).
8. Kukla, V. *et al.* NMR studies of single-file diffusion in unidimensional channel zeolites. *Science* **272,** 702–704 (1996).
9. Hellriegel, C., Kirstein, J. & Bräuchle, C. Tracking of single molecules as a powerful method to characterize diffusivity of organic species in mesoporous materials. *New J. Phys.* **7,** 23 (2005).
10. Jung, C., Hellriegel, C., Michaelis, J. & Bräuchle, C. Single-molecule traffic in mesoporous materials: translational, orientational, and spectral dynamics. *Adv. Mater.* **19,** 956–960 (2007).
11. Kirstein, J. *et al.* Exploration of nanostructured channel systems with single-molecule probes. *Nature Mater.* **6,** 303–310 (2007).
12. McCain, K. S., Hanley, D. C. & Harris, J. M. Single-molecule fluorescence trajectories for investigating molecular transport in thin silica sol-gel films. *Anal. Chem.* **75,** 4351–4359 (2003).
13. Werley, C. A. & Moerner, W. E. Single-molecule nanoprobes explore defects in spin-grown crystals. *J. Phys. Chem. B* **110,** 18939–18944 (2006).
14. Schmidt, T., Schütz, G. J., Baumgartner, W., Gruber, H. J. & Schindler, H. Imaging of single molecule diffusion. *Proc. Natl Acad. Sci. USA* **93,** 2926–2929 (1996).
15. Seisenberger, G. *et al.* Real-time single-molecule imaging of the infection pathway of an adeno-associated virus. *Science* **294,** 1929–1932 (2001).
16. Roeffaers, M. B. J. *et al.* Spatially resolved observation of crystal-face-dependent catalysis by single turnover counting. *Nature* **439,** 572–575 (2006).
17. Sakamoto, Y. *et al.* Direct imaging of the pores and cages of three-dimensional mesoporous materials. *Nature* **408,** 449–453 (2000).
18. Holtrup, F. O. *et al.* Terrylenimides: new NIR fluorescent dyes. *Chem. Eur. J.* **3,** 219–225 (1997).
19. Jung, C. *et al.* A new photostable terrylene diimide dye for applications in single molecule studies and membrane labeling. *J. Am. Chem. Soc.* **128,** 5283–5291 (2006).
20. Brinker, C. J., Lu, Y., Sellinger, A. & Fan, H. Evaporation-induced self-assembly: nanostructures made easy. *Adv. Mater.* **11,** 579–585 (1999).
21. Saxton, M. J. & Jacobson, K. Single-particle tracking: applications to membrane dynamics. *Annu. Rev. Biophys. Biomol. Struct.* **26,** 373–399 (1997).

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

Acknowledgements We thank K. Müllen for providing the TDI dye; Siltronic AG for the silicon wafers; the group of P. Müller-Buschbaum for the 2D-GISAXS measurements; and S. Schmidt and B. Platschek for assistance with electron microscopy. This work was funded by two Collaborative Research Centres ('Manipulation of matter at the nanometer length scale' and 'Dynamics and intermediates of molecular transformations') of the German Research Foundation (DFG) and by the Nanosystems Initiative Munich (NIM), as well as the Center for Integrated Protein Science Munich (CiPSM).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to C.B. (christoph.braeuchle@cup.uni-muenchen.de) or T.B. (bein@lmu.de).

©2007 Nature Publishing Group

METHODS

Films for characterization by X-ray diffraction and ellipsometry. For X-ray diffraction and ellipsometry measurements, 200 μl of the final precursor solution were spin-coated on silicon wafers (28 mm × 15 mm). The corresponding 2D-GISAXS pattern (Fig. 1a) shows a two-dimensional hexagonal order for the mesoporous silica, with the peaks corresponding to the *d*-spacings of $d(10) = 5.5$ nm and $d(01) = 5.9$ nm, respectively. The pore-to-pore distances are therefore $a(01) = 7.0$ nm (parallel to the substrate plane) and $a(10) = 6.5$ nm. The shrinkage of the pores normal to the substrate is 7%. The thickness determined by ellipsometry is about 100 nm.

Overlay of optical and electron microscopy images. For the preparation of the films, 50 μl of the final precursor solution were spin-coated on Si_3N_4 membranes 30 nm in thickness supported by a small silicon wafer with a 500 μm × 500 μm window (Fig. 1b, c; PLANO; Wetzlar). The film on the central region of the supported membrane was thin enough for electron microscopy. The polystyrene beads and the gold nanoparticles were enclosed in the mesoporous film (Supplementary Fig. 1).

The wide-field microscopy setup was based on an Eclipse TE200 (Nikon) epifluorescence microscope with a high-numerical-aperture (NA) oil-immersion objective (Nikon Plan Apo 100×/1.40 NA oil). The molecules were excited at 633 nm with a HeNe gas laser (75 mW max. at 632.8 nm; Coherent) with an intensity of about 0.3 kW cm^{-2}, and their fluorescence was detected with a back-illuminated EM-CCD camera in frame transfer mode (iXon DV897, 512 pixels × 512 pixels; Andor). Incident laser light was blocked by a dichroic mirror (640 nm cutoff; AHF) and a bandpass filter (730 nm/140 nm; AHF). Additional lenses ($f = 150$ mm, $f = 200$ mm, achromat; Thorlabs) in the detection pathway give an overall magnification of ×133. This means that the side of one square pixel on the camera chip (16 μm) should represent 120 nm on the sample. By using the USAF test target (1951 Chromium positive; Melles Griot) this value was experimentally determined to 122 nm, which agrees well with the theoretical value. Because the films were much thinner than the focal depth of the microscope objective used (>1 μm), images contain data from molecules at all heights inside and on the surface of the sample. Series of 1,000 images were acquired with a temporal resolution of 200 ms per frame. In each movie frame, single molecules show up as bright spots. Because the Si_3N_4 membranes give a relatively high and inhomogeneous background signal, the background was subtracted frame-to-frame from the films before tracking of the molecules. Single-molecule trajectories were then built up by fitting the positions of the spots from frame to frame.

To overlay these trajectories on the TEM images, the positions of the polystyrene beads had to be determined from the same movie. For this purpose the laser shutter was closed for the last 240 frames of the movie and the polystyrene beads were imaged in transmission in white light (Supplementary Movie 1). Their positions were fitted in inverted images with the same gaussian fit routine as the single molecules (red crosses in Fig. 2b). Mean and standard deviations of the fitted positions were calculated and tabulated.

After optical microscopy the sample was analysed with TEM. Again, during the first step the whole membrane was mapped at low resolution (Supplementary Fig. 2b). The central region with the appropriate thickness was then mapped at a magnification high enough to distinguish single polystyrene beads (Fig. 1e). In this map we could retrieve the same geometric arrangement of polystyrene beads as in the white-light transmission images (Fig. 1f, g). In all, 400 (20 × 20) images in the neighbourhood of the tracked trajectories were taken at ×40,000 magnification and were merged by using image cross-correlation to obtain a map of an area of about 3.5 μm × 3.5 μm. The centre positions of the polystyrene spheres in the TEM images were determined by fitting a circle on the outside margin of their patterns.

The overlay of the TEM and wide-field images was accomplished by fitting the best overlay of the polystyrene bead positions in the respective images by the use of a least-square algorithm (Fig. 2b–d). In this iterative process the sum of the squared deviations of the bead positions in TEM images and wide-field tracking data was minimized by scaling, rotating and translating the tracking data. As an example, Supplementary Table 1 shows the bead positions in the TEM images and in wide-field tracking data of Fig. 2 before and after the overlay procedure. In Fig. 2, the positions of four well-separated beads were used, yielding an average deviation of 26 nm after the overlay. The two beads at the left edge of the TEM image in Fig. 2c are too close to be resolved by optical microscopy. They are therefore not included in the overlay process. However, in some cases such pairs of closely neighbouring beads had to be used for the overlay to have a minimum of three anchor points for the overlay. In such cases, the individual positions of the closely neighbouring beads were determined in the TEM images. The centre position between the two beads was then correlated with the centre position of the single spot resulting from the two beads in the respective white-light transmission images, fitted by the gaussian function. For example, in Fig. 4d–f two such doubled beads had to be used for correlating tracking data and TEM images, which resulted in an average deviation of 48 nm.

©2007 Nature Publishing Group

LETTERS

A diffusion mechanism for core–mantle interaction

Leslie A. Hayden[1] & E. Bruce Watson[1]

Understanding the geochemical behaviour of the siderophile elements — those tending to form alloys with iron in natural environments — is important in the search for a deep-mantle chemical 'fingerprint' in upper mantle rocks, and also in the evaluation of models of large-scale differentiation of the Earth and terrestrial planets. These elements are highly concentrated in the core relative to the silicate mantle, but their concentrations in upper mantle rocks are higher than predicted by most core-formation models[1,2]. It has been suggested that mixing of outer-core material back into the mantle following core formation may be responsible for the siderophile element ratios observed in upper mantle rocks[3]. Such re-mixing has been attributed to an unspecified metal–silicate interaction in the reactive D″ layer just above the core–mantle boundary[4]. The siderophile elements are excellent candidates as indicators of an outer-core contribution to the mantle, but the nature and existence of possible core–mantle interactions is controversial[5]. In light of the recent findings that grain-boundary diffusion of oxygen through a dry intergranular medium may be effective over geologically significant length scales[6] and that grain boundaries can be primary storage sites for incompatible lithophile elements[7], the question arises as to whether siderophile elements might exhibit similar (or greater) grain-boundary mobility. Here we report experimental results from a study of grain-boundary diffusion of siderophile elements through polycrystalline MgO that were obtained by quantifying the extent of alloy formation between initially pure metals separated by ~1 mm of polycrystalline MgO. Grain-boundary diffusion resulted in significant alloying of sink and source particles, enabling calculation of grain-boundary fluxes. Our computed diffusivities were high enough to allow transport of a number of siderophile elements over geologically significant length scales (tens of kilometres) over the age of the Earth. This finding establishes grain-boundary diffusion as a potential fast pathway for chemical communication between the core and mantle.

The distribution of the chemical elements in the Earth has traditionally been characterized according to their partitioning behaviour towards minerals, melts and fluid phases. In general, grain boundaries have not been considered significant storage sites for most elements, though this view is changing following the recent demonstration that localization at grain boundaries is potentially important for at least some elements[8]. The capacity of grain boundaries to transport a given element (as opposed to simply store it) is a function of both the partitioning of that element into the grain boundary and its grain-boundary diffusivity. These combine to determine the potential grain-boundary flux. Grain-boundary diffusivities of cations and oxygen in oxides and silicates are typically several orders of magnitude higher than lattice values[9,10]. To our knowledge, no data are available for grain-boundary diffusion of electrically neutral, highly incompatible metal atoms in oxides, as conventional studies of grain-boundary diffusion rely upon the compatibility of the diffusant in the crystal lattice to detect grain-boundary transport[11]. The highly and moderately siderophile elements (HSE and MSE) are not only very incompatible in the lattices of mantle phases, but also likely to have very low lattice diffusivities—an extreme case of 'type C' grain-boundary diffusion in the terminology of Harrison[12]. These elements may nevertheless be present and mobile in the grain boundaries of mantle phases, which would lead to a previously unrecognized reservoir and pathway for storage and transport. Grain-boundary concentrations are expected to be very low, so direct detection is beyond any means currently available. In this study we chose to explore the grain-boundary mobility of siderophile elements in polycrystalline MgO (periclase). A simple, dense oxide that is stable over a wide range of pressure–temperature conditions, this mineral is a suitable analogue for deep-mantle materials that can be studied in adequately sized samples at conditions accessible to solid–media devices.

All elements selected for this study are mobile in grain boundaries of polycrystalline MgO. Grain-boundary transport caused 0.07–95% alloying with Pt, depending on the source element and the run duration. Overall, Os and Re consistently displayed the smallest grain-boundary fluxes, while Au, W and Rh displayed the largest. The amount of alloying with Pt differed not only between different elements but also from particle to particle within the same experiment by up to a factor of 5. Not all of the Pt sink particles contained the same concentration of the source element. The distribution of alloying source atoms was also not homogenous in the Pt foils used in place of particles in some experiments. This implies that not all grain boundaries are equally 'conductive' pathways.

Results are shown in Fig. 1. The grain-boundary fluxes vary over approximately three orders of magnitude among the elements studied. There is good reproducibility among experiments, as results are generally within the same order of magnitude. The elements in

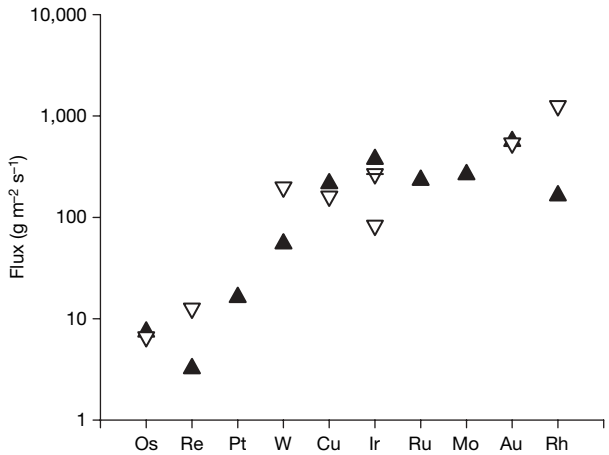


Figure 1 | Results of single-sink experiments. All the experiments shown were run at 1,600 °C and 2.5 GPa for 10 h, with a sink layer located ~1 mm from the source layer. Filled triangles represent initial series of experiments; open triangles represent experiments repeated in duplicate or triplicate.

[1]Department of Earth and Environmental Sciences, Rensselaer Polytechnic Institute, Troy, New York 12180, USA.

©2007 Nature Publishing Group

Fig. 1 are listed in order of increasing flux; at this time there does not appear to be a clear correlation between an element's grain-boundary diffusivity and any single atomic property, such as mass, volatility or siderophility.

For selected siderophile elements, additional experiments were performed with modified experimental conditions such that the sink layer was moved farther away from the source, the experiments were run for a longer time or both conditions were modified. The results of this series of experiments are shown in Fig. 2. The flux values for these modified runs have been normalized (to $x = 1$ mm and $t = 10$ h, where x is the distance between the source and sink, and t is the run duration) in order to better compare the results. There is again very good agreement among samples, and this series produced results that make sense in the context of well-understood diffusion principles[13]. A run duration of 50 h resulted in a lower time-averaged flux than one of 10 h. Increasing the distance between the source and sink horizons from 1 mm to 5 mm resulted in a markedly decreased average flux. The time-integrated flux should be proportional to time for steady-state transport. Steady-state conditions are assumed to hold here for the purpose of obtaining diffusivity estimates, but this is not strictly realistic because of the dynamic nature of the system, which exhibits coarsening of MgO grains and compositional changes in both the source and sink particles. These results are also consistent with the data in Fig. 1.

Back-scattered electron imaging of the polycrystalline MgO revealed a distinct increase in average grain size with increasing run duration. This is due to normal grain coarsening, but its occurrence in these experiments is potentially significant because it reduces the total grain-boundary area (or volume) of the MgO. The bulk flux of siderophile diffusant atoms through the MgO must undergo a gradual diminution with time as some grain-boundary pathways are modified and others eliminated completely. It is possible that coarsening effects are responsible for some of the scatter in Fig. 2. With the exception of Pt, the (normalized) time integrated fluxes of source atoms were smaller in the longer runs, due in part to grain growth causing decreased total grain-boundary volume.

Because of the low concentrations of the siderophile elements in the grain boundaries (below detection with an electron microprobe) and the very narrow (~1 nm) width of individual grain boundaries[13,14], direct measurement of the atoms of interest in the grain boundaries was not possible. Ideally, knowledge of the concentration gradient along the grain boundaries from source to sink is necessary for calculation of diffusivity. In order to translate the calculated relative fluxes of the elements into an estimate of diffusivity, a series of experiments was run using multiple evenly spaced Pt foil sink layers in the MgO. Results of these multi-sink experiments are shown as the solid line in Fig. 3. These experiments also differ in that they each had two elements in the source layer, effectively setting up a diffusion 'race'. Figure 3 shows the fluxes from the source layer to the first sink layer only, and these results are generally consistent with the results of the single-sink experiments. The presence of a competing element does not seem to have a significant effect on the mobility compared to that element diffusing alone. The concentrations at each sink horizon were determined and plotted as a function of distance from the source layer, then fitted with a simple error function solution to estimate diffusivity: $C_{(x,t)} = C_0[1 - \mathrm{erf}(x/\sqrt{4Dt})]$, where C_0 is the initial concentration of the source element (~1 for the pure starting metal), x is the distance from the source layer, D is the diffusivity and t is the experiment duration. Given local equilibrium between the metal particles and the grain boundaries, the varying concentrations in the metal particles as a function of distance from the source provide a relative measure of concentration variation in a grain boundary that connects a source particle to a sink. The resulting diffusivities range from 10^{-12} m^2 s^{-1} to 10^{-7} m^2 s^{-1} (Supplementary Fig. 3). These diffusivities represent minimum values due to the fact that some of the diffusing source particles are 'trapped' upon reaching the initial Pt sink layer, and source concentrations in the more distant sinks are lower than they would be in the absence of this chemical (and physical) boundary.

To put these diffusivity values in perspective, the characteristic diffusion distance ($x^2 = Dt$) over the Earth's history for these elements is on the order of 1–100 km (Fig. 4). Grain-boundary diffusion of siderophile elements is thus fast enough to allow significant core–mantle communication, especially if material diffused into the lowermost mantle from the outer core is swept into the mantle convective cycle.

When considering the implications of this experimental work to the environment at the core–mantle boundary (CMB), there are a few parameters to take into consideration. First, these experiments were done at temperatures and pressures significantly lower than the conditions at the CMB. Increasing temperature and pressure generally have opposing effects on diffusion; that is, diffusivity increases with temperature and decreases with pressure. Preliminary work[15] has shown that tungsten mobility increases dramatically between 1,450 and 1,600 °C. Second, although MgO (as magnesiowüstite) is widely believed to be present in the lower mantle[16], the predominant phase is Mg-perovskite. It seems unlikely that the grain-boundary mobility of HSE and MSE observed in MgO would be hampered by the presence of Mg-perovskite, but it is premature to assume that perovskite grain boundaries would also serve as high-diffusivity pathways for these

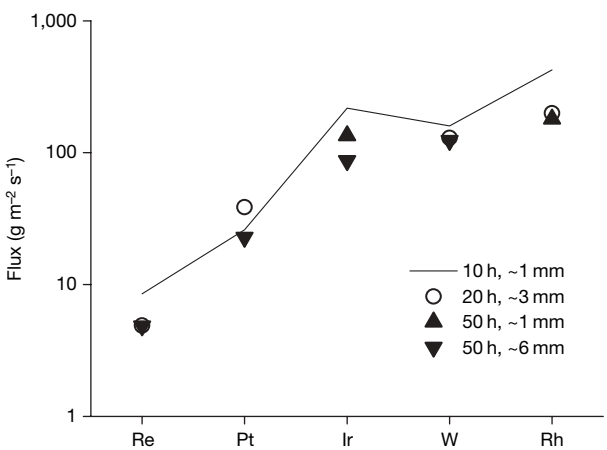


Figure 2 | Time-averaged flux values. Flux values for selected elements were derived from the time series. The solid line represents the 'standard' set-up with the shortest run duration. The open circles, triangles and inverted triangles represent experiments with modified run conditions, as shown in the key.

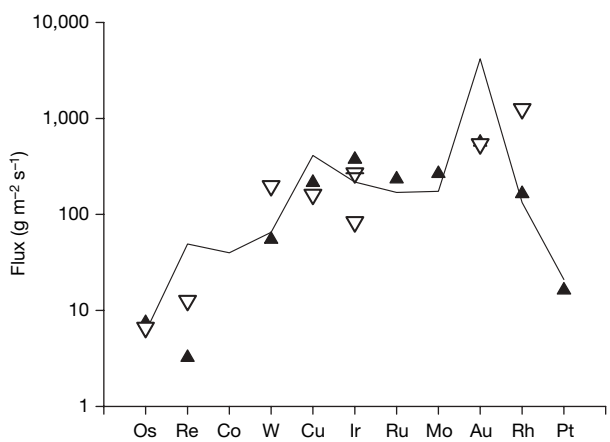


Figure 3 | Comparison of single and multi-sink experiments. The solid line represents multi-sink runs with three sink horizons of Pt foil located ~1.5, 3 and 4.5 mm from the source layer. Elements were run in pairs: Os-Rh, Re-Ru, Co-Au, W-Ir and Cu-Mo. Solid and open triangles represent initial and replicate experiments, respectively.

©2007 Nature Publishing Group

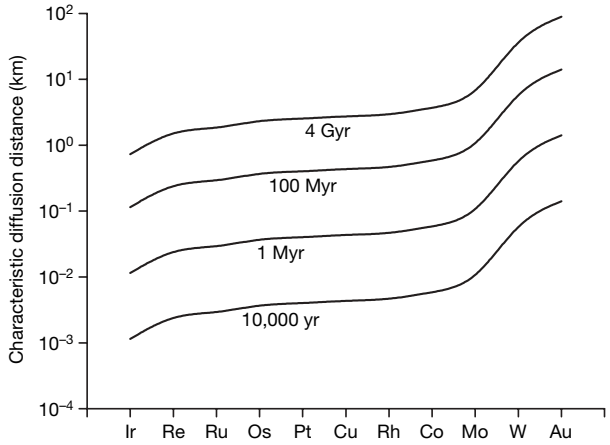


Figure 4 | Characteristic diffusion distance. This value was determined using the relation $x = \sqrt{Dt}$ based on calculated grain-boundary diffusivity for MSE and HSE over various time periods: 10,000 yr, 1 Myr, 100 Myr and 4 Gyr.

elements. The contiguity (extent of grain-to-grain contact) of magnesiowüstite, in addition to the possible existence of discontinuous zones or layers in the lower mantle phases, may also influence the potential for long-distance transport. Finally, diffusion between two pure metals is not representative of CMB conditions, but represents a simple experimental design from which data can be obtained over laboratory timescales and within analytical constraints.

Here we have shown that a large number of siderophile elements are mobile in the grain boundaries of polycrystalline MgO. The diffusivities of these elements are high enough to suggest that grain-boundary diffusion could occur on a geologically significant length scale. Grain-boundary diffusion (combined with convective overturn of the mantle) is thus a plausible mechanism for chemical interaction at and above the core-mantle boundary. Understanding the grain-boundary segregation and diffusion behaviour of these elements would provide a new tool for interpreting the abundance patterns and isotope systematics of siderophile elements in mantle rocks and thus understanding the geochemical processes in the Earth's mantle.

METHODS SUMMARY

The general experimental strategy was to establish a one-dimensional gradient in the chemical potential of each metal of interest (Os, Ir, Ru, Rh, Pt, Re, Au, W, Mo, Co or Cu) across a volume of polycrystalline MgO having an equilibrium microstructure. Our technique involved pre-fabrication of MgO wafers (~3 mm diameter × 1 mm thick), the opposing faces of which were smeared with metal powders—a 'source' of the diffusant of interest on one side and a suitable 'sink' metal (Pt powder or foil) on the other. During the actual diffusion anneals, the faces created source and sink horizons of metal particles separated by ~1 mm of polycrystalline MgO, typically ~50 μm in grain size. The wafers were placed in a standard 12.7 mm diameter piston-cylinder assembly and held at 1,600 °C and 2.5 GPa for 10 to 50 h. Viewed in terms of a single planar grain boundary, the 'source-sink' experimental strategy works as follows: diffusant atoms (for example, Au) are partitioned into the grain boundary at the source and diffuse along the boundary until they encounter the sink phase (Pt) into which they dissolve. The time-integrated grain-boundary flux (J_{gb}) can be evaluated by measuring the total number of diffusant atoms (n) accumulated in the sink in a given time:

$$n = A_{int} \int_{t=0}^{t} J_{gb} \mathrm{d}t \qquad (1)$$

where A_{int} is the effective cross-sectional area of the grain boundary that is intersected by the sink phase. The implementation of this idea for a volume of MgO having numerous grain boundaries of diverse orientation is discussed in the Methods section. The run products were analysed with an electron microprobe by measuring the concentration of the source element in the Pt (essentially determining n in equation (1)). The alloying (or 'communication') of the two initially pure metals separated by polycrystalline MgO is taken as proof of the existence of grain-boundary diffusion pathways in the MgO.

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

Received 17 May; accepted 3 October 2007.

1. O'Neill, H. St C. The origin of the moon and the early history of the Earth—a chemical model. Part 2: the Earth. *Geochim. Cosmochim. Acta* **55,** 1159–1172 (1991).
2. Jones, J. H. & Drake, M. J. Geochemical constraints on core formation in the Earth. *Nature* **322,** 221–228 (1986).
3. Walker, R. J. & Walker, D. Does the core leak? *Eos* **86,** 237–244 (2005).
4. Brandon, A. D., Walker, R. J., Morgan, J. W., Norman, M. D. & Pritchard, H. M. Coupled ^{186}Os and ^{187}Os evidence for core-mantle interaction. *Science* **280,** 1570–1573 (1998).
5. Brandon, A. D. & Walker, R. J. The debate over core-mantle interaction. *Earth Planet. Sci. Lett.* **232,** 211–225 (2005).
6. Van Orman, J. A., Fei, Y., Hauri, E. H. & Wang, J. Diffusion in MgO at high pressures: Constraints on deformation mechanisms and chemical transport at the core-mantle boundary. *Geophys. Res. Lett.* **30,** 1056–1059 (2003).
7. Hiraga, T., Anderson, I. M. & Kohlstedt, D. L. Chemistry of grain boundaries in mantle rocks. *Am. Mineral.* **88,** 1015–1019 (2003).
8. Hiraga, T., Anderson, I. M. & Kohlstedt, D. L. Grain boundaries as reservoirs of incompatible elements in the Earth's mantle. *Nature* **427,** 699–703 (2004).
9. Farver, J. R. & Yund, R. A. Measurement of oxygen grain boundary diffusion in natural, fine-grained, quartz aggregates. *Geochim. Cosmochim. Acta* **97,** 1597–1607 (1991).
10. Kingery, W. D. *Introduction to Ceramics* 2nd edn 250–257 (Wiley, New York, 1976).
11. Fisher, J. C. Calculation of diffusion penetration curves for surface and grain-boundary diffusion. *J. Appl. Phys.* **22,** 74–77 (1951).
12. Harrison, L. G. Influence of dislocations on diffusion kinetics in solids with particular reference to the alkali halides. *Trans. Faraday Soc.* **57,** 1191–1199 (1961).
13. Ricoult, D. L. & Kohlstedt, D. L. Structural width of low-angle grain boundaries in olivine. *Phys. Chem. Mineral.* **9,** 133–138 (1983).
14. Sutton, A. P. & Balluffi, R. W. *Interfaces in Crystalline Materials* 305–310 (Oxford Univ. Press, New York, 1995).
15. Watson, E. B. Mobility of siderophile elements in grain boundaries of periclase and periclase/olivine aggregates. *Eos* **83**(Spring Meet. Suppl.), abstr. V52B-05 (2002).
16. Ringwood, A. E. Phase transformations and their bearing on the constitution and dynamics of the mantle. *Geochim. Cosmochim. Acta* **55,** 2083–2110 (1991).

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

Acknowledgements This work was supported by a National Science Foundation grant to E.B.W.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.A.H. (haydel@rpi.edu).

©2007 Nature Publishing Group

METHODS

Traditional methods of characterizing grain-boundary diffusion depend on the compatibility of the element of interest in the crystal lattice. However, in cases where the diffusant can be presumed to exist primarily in the grain boundaries, as with the HSEs, its mobility may be undetected by most bulk analysis methods. Therefore the characterization of grain-boundary transport of siderophile elements in oxides requires a different experimental strategy, using 'source' and 'sink' layers of metal to detect movement through grain boundaries.

In the most basic type of experiment, cylindrical MgO wafers, approximately 1 mm in length, were presynthesized to obtain an equilibrium microstructure. Powdered metal particles were smeared on each side of the wafer; the upper horizon with the diffusant or 'source' (Os, Ir, Ru, Rh, Pt, Re, Au, W, Mo, Co, Cu) and the lower horizon with the 'sink' element, Pt (in some experiments thin Pt foil was used instead of metal powder) (Supplementary Fig. 1). As the source elements are highly incompatible in the lattice, alloying of the source element with the initially pure Pt 'sink' particles implies a grain-boundary diffusion pathway. The wafers were placed in a graphite furnace in a standard 12.7 mm piston-cylinder assembly in which the furnace was separated from the NaCl pressure medium by a thermally insulating silica glass sleeve (Supplementary Fig. 1). Experiments were run at 1,600 °C and 2.5 GPa for 10–50 h. There was also a series of experiments containing multiple sink layers involving three Pt foil horizons located ~1.5, 3 and 4.5 mm from the source layer; as well as a 'time series' for selected elements in which either the run duration (up to 50 h) or diffusion distance (up to 6 mm) was modified.

Although the oxygen fugacity during the run was not measured or quantified, it can be constrained to a reducing environment. The graphite furnace shows the presence of reduced carbon. The inherent stability (absence of oxidation) of the W-Re thermocouple and the stability of the metal phase in experiments run with oxide-forming siderophiles (Mo-MoO_2) are also indications of reducing conditions.

The run products were analysed with a Cameca SX-100 electron microprobe by measuring the concentration of the source element in the sink particles/foils. The source particles were also analysed for the presence of sink materials. Analyses were performed with a 1 or 10 μm beam (for analyses of particles and foils, respectively) at 15 kV accelerating potential and sample currents of 100 nA. K_α X-rays were collected through LiF crystals for Co and Cu; L_α X-rays were collected through PET (pentaerythritol) crystals for Rh, Ru and Mo and through a LiF crystal for Pt; and M_α X-rays were collected through PET crystals for Au, Ir, Os, Re and W. Acquisition times were 60 s for all elements except Pt, for which it was 10 s. These analytical procedures resulted in detection limits on the order of ~10 p.p.m. for Mo and Os; ~300 p.p.m. for Co, Cu, Ir, Re, Rh and Ru; and ~1,000 p.p.m. for Au, Pt and W. In the experiments with Pt powder in the sink horizon, owing to beam limitations only particles larger than a few micrometres in diameter were analysed. In experiments with foils, analyses were done along the length of the foil in approximately ~40 μm steps. To ensure that neither the furnace nor the silica glass sleeves were breached by NaCl, energy dispersive X-ray spectrometer analyses for NaCl were performed on both the polycrystalline MgO medium and the graphite and silica sleeves. Back-scattered electron images were also taken of the sectioned samples.

The raw microprobe data were used in conjunction with SigmaScan image software to determine the grain-boundary volume in the sample and ultimately the flux of source atoms through the MgO (Supplementary Fig. 2). This was done by first calibrating the scale of the back-scattered electron image, then using the software to determine the area of the Pt particles or foils. The two-dimensional image was assigned a depth of 10 μm, which seems a reasonable estimate given the general shape of the Pt particles. This volume was then converted to a mass of Pt via its density. The number of atoms and then the mass of the source element in each particle (or foil) can then be determined. The average concentration (for runs with Pt particles) is essentially the summation of the total mass of source element divided by the summation of the total mass of sink element.

The time-integrated grain-boundary flux (Jgb) can be evaluated by measuring the total number of diffusant atoms (n) accumulated in the sink in a given time via equation (1) in the main text.

In order to determine the grain-boundary volume, the total length of the grain boundaries was carefully measured, again with the software, and assigned a width of 1 nm. The distance between the source and sink horizons was measured and then multiplied by a factor of 1.7 to represent the tortuosity of the pathway. These three parameters—mass transport, grain-boundary volume and diffusion distance—plus the duration of the experiment were then used to calculate the flux of the source element through the grain boundaries. This amounts to a practical implementation of equation (1) for a complex, polycrystalline sample.

The flux alone has value as a qualitative indicator of the effectiveness of grain-boundary diffusant transport, but additional information or assumptions are needed to calculate diffusivity. Assuming the system quickly establishes a (near) steady-state condition, the relationship between diffusivity and flux in the grain boundary is given approximately by

$$D_{\mathrm{gb}} = -J_{\mathrm{gb}}\left[\frac{\mathrm{d}c}{\mathrm{d}x}\right] \tag{2}$$

where $\mathrm{d}c/\mathrm{d}x$ is the concentration gradient of the diffusant in the grain boundary. Complications arise in this simplistic approach, of course, because neither the absolute concentrations in the grain boundary nor $\mathrm{d}c/\mathrm{d}x$ is expected to be directly measurable (or strictly constant), but in the total absence of data even approximate diffusivities have value. A semi-infinite model for diffusion

$$C_{(x,t)} = C_0[1-\mathrm{erf}(x/\sqrt{4Dt})] \tag{3}$$

was used to calculate D from the 'concentration profile' from multi-sink experiments. This approach is appropriate—even for fast diffusing elements like Au and W, the mass transfer via diffusion is very small compared to the mass of the source layer. However, this is by no means a rigorous mathematical treatment, because no such treatment is available for these and similar experiments. It was implemented to approximate diffusivity within an order of magnitude.

©2007 Nature Publishing Group

LETTERS

Small molecule activators of SIRT1 as therapeutics for the treatment of type 2 diabetes

Jill C. Milne[1]*, Philip D. Lambert[1]*, Simon Schenk[2]*, David P. Carney[1], Jesse J. Smith[1], David J. Gagne[1], Lei Jin[1], Olivier Boss[1], Robert B. Perni[1], Chi B. Vu[1], Jean E. Bemis[1], Roger Xie[1], Jeremy S. Disch[1], Pui Yee Ng[1], Joseph J. Nunes[1], Amy V. Lynch[1], Hongying Yang[1], Heidi Galonek[1], Kristine Israelian[1], Wendy Choy[1], Andre Iffland[1], Siva Lavu[1], Oliver Medvedik[1], David A. Sinclair[3], Jerrold M. Olefsky[2], Michael R. Jirousek[1], Peter J. Elliott[1] & Christoph H. Westphal[1]

Calorie restriction extends lifespan and produces a metabolic profile desirable for treating diseases of ageing such as type 2 diabetes[1,2]. SIRT1, an NAD$^+$-dependent deacetylase, is a principal modulator of pathways downstream of calorie restriction that produce beneficial effects on glucose homeostasis and insulin sensitivity[3–9]. Resveratrol, a polyphenolic SIRT1 activator, mimics the anti-ageing effects of calorie restriction in lower organisms and in mice fed a high-fat diet ameliorates insulin resistance, increases mitochondrial content, and prolongs survival[10–14]. Here we describe the identification and characterization of small molecule activators of SIRT1 that are structurally unrelated to, and 1,000-fold more potent than, resveratrol. These compounds bind to the SIRT1 enzyme–peptide substrate complex at an allosteric site amino-terminal to the catalytic domain and lower the Michaelis constant for acetylated substrates. In diet-induced obese and genetically obese mice, these compounds improve insulin sensitivity, lower plasma glucose, and increase mitochondrial capacity. In Zucker *fa/fa* rats, hyperinsulinaemic-euglycaemic clamp studies demonstrate that SIRT1 activators improve whole-body glucose homeostasis and insulin sensitivity in adipose tissue, skeletal muscle and liver. Thus, SIRT1 activation is a promising new therapeutic approach for treating diseases of ageing such as type 2 diabetes.

To identify activators of human SIRT1 a high-throughput *in vitro* fluorescence polarization assay was developed and used to screen a large collection of small molecules. The hits identified were structurally distinct from resveratrol[11] and subsequently optimized for *in vitro* enzyme activity using a high-throughput mass spectrometry assay. Potency was tracked by determining the concentration of compound required to increase enzyme activity by 50% ($EC_{1.5}$) and the percentage maximum activation achieved at the highest doses of compound tested (resveratrol $EC_{1.5} = 46.2$ μM and maximum activation = 201%; SRT1460 $EC_{1.5} = 2.9$ μM and maximum activation = 447%; SRT2183 $EC_{1.5} = 0.36$ μM and maximum activation = 296%; and SRT1720 $EC_{1.5} = 0.16$ μM and maximum activation = 781%) (Fig. 1a, b). The variation in maximum activation is reminiscent of agonist activity observed with receptors, where compounds can vary in their ability to mediate full receptor activation as compared to an endogenous ligand[15]. These compounds are selective for activation of SIRT1 versus the closest sirtuin homologues, SIRT2 and SIRT3 (SIRT2: resveratrol, SRT1460 and SRT2183 $EC_{1.5} > 300$ μM, SRT1720 $EC_{1.5} = 37$ μM; SIRT3: resveratrol, SRT1460, SRT1720 and SRT2183 $EC_{1.5} > 300$ μM).

The compounds were tested for functional activity in a SIRT1 cell-based (U2OS cells) deacetylation assay (Fig. 1c)[16,17]. SRT2183 was used as a positive control because it is well tolerated in cellular assays. The SIRT1 activators decreased the acetylation state of p53, a known SIRT1 substrate. A selective SIRT1 inhibitor, 6-chloro-2,3,4,9-tetrahydro-1-*H*-carbazole-1-carboxamide[18], abrogated the deacetylation activity, indicating that the effect was dependent on SIRT1.

The *Sir2* gene in *Saccharomyces cerevisiae* and *Drosophila melanogaster* extends lifespan when overexpressed[19,20], and the ability of

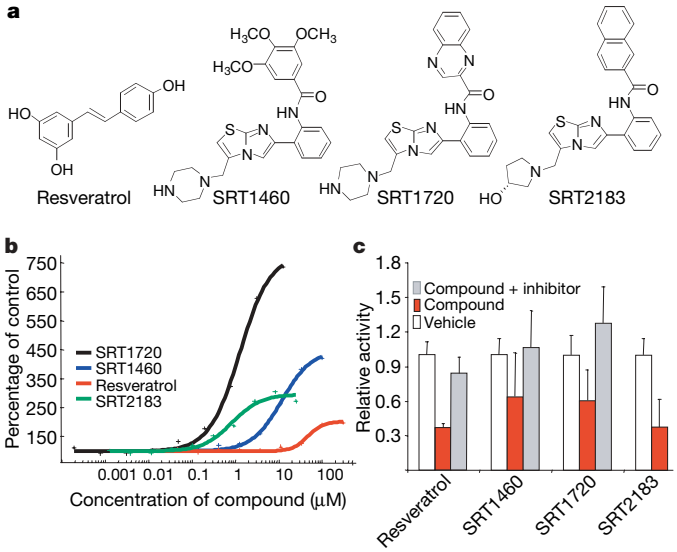


Figure 1 | Identification of potent SIRT1 activators unrelated to resveratrol. **a**, Chemical structures of SIRT1 activators, resveratrol, SRT1460, SRT2183, and SRT1720. **b**, The effect of activators on human SIRT1 enzyme activity measured by mass spectrometry. **c**, Cellular activity was measured using an ICW that monitors the degree of p53 deacetylation in U2OS cells using β-tubulin as a normalization control. Compounds were tested at concentrations of: resveratrol (100 μM), SRT2183 (10 μM), SRT1460 (10 μM), SRT1720 (0.10 μM). Each concentration represents the approximate EC_{50} for each compound. $n = 6$ for all compounds tested except for resveratrol, where $n = 3$. Data are expressed as mean ± s.d. The activation of SIRT1 resulting in p53 deacetylation could be blocked by a SIRT1 inhibitor, 6-chloro-2,3,4,9-tetrahydro-1-*H*-carbazole-1-carboxamide (10 μM). $n = 3$ for all compounds tested.

[1]Sirtris Pharmaceuticals Inc., 790 Memorial Drive, Cambridge, Massachusetts 02139, USA. [2]Division of Endocrinology and Metabolism, Department of Medicine, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. [3]Department of Pathology, Paul F. Glenn Laboratories for the Biological Mechanisms of Aging, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.
*These authors contributed equally to this work.

©2007 Nature Publishing Group

calorie restriction to extend lifespan is abrogated when the gene is deleted[20–22]. We have observed lifespan extension in yeast with analogues structurally related to SRT1460 that activate the yeast SIR2 enzyme (data not shown).

To address the mechanism of activation of the human SIRT1 enzyme, we determined the effect of compound on the V_{max} (velocity of enzyme-catalysed reaction at infinite concentration of substrate) and the K_m (Michaelis constant) of SIRT1 for its two substrates, NAD^+ and acetylated peptide. None of the compounds affected the K_m for NAD^+ or the V_{max} (data not shown); however, all compounds decreased the K_m of SIRT1 for acetylated peptide substrate (Fig. 2a). The magnitude of the K_m effect positively correlated with the $EC_{1.5}$ values for the compounds. A similar K_m-type activation mechanism has been proposed for small molecule activators that bind to an allosteric site of glucokinase[23].

The energetics of binding of SRT1460 to purified SIRT1 enzyme were studied by isothermal titration calorimetry. The titration of SRT1460 against purified human SIRT1-C did not result in detectable binding (data not shown). In the presence of acetylated peptide substrate SRT1460 exhibited signs of protein-binding-site saturation best fitting a one site binding model (Fig. 2b). The dissociation constant (K_d) for SRT1460 was 16.2 μM and the enthalpy (ΔH) was −6.1 kcal mol^{-1}. The data suggest that SRT1460, and other related SIRT1 activators, bind to a SIRT1–peptide substrate complex and promote a more productive conformation that enhances catalytic activity. Binding of substrate may induce a conformational change that leads

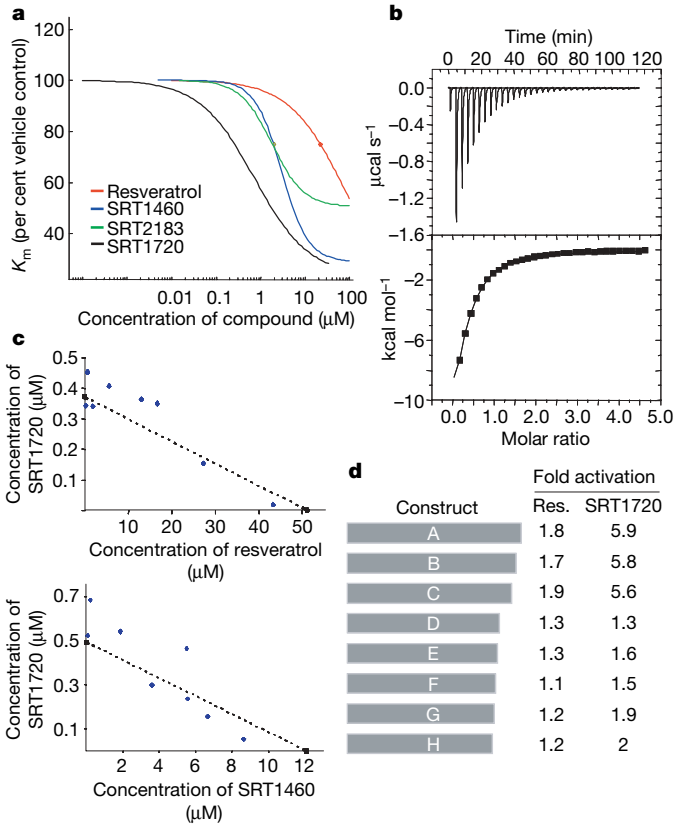


Figure 2 | *In vitro* characterization of activators of human SIRT1. **a**, The effect of SIRT1 activators on peptide substrate K_m. **b**, Calorimetric titrations of SIRT1-C–peptide substrate complex with the activator SRT1460. Top panel: heat of binding SRT1460 to enzyme–peptide complex. Bottom panel: integrated fit with a one-site binding model. **c**, Isobologram analysis of resveratrol versus SRT1720 and SRT1720 versus SRT1460. The experimental data are best fit to the theoretical line of additivity (dashed line). **d**, SIRT1 N-terminal truncations define the allosteric compound binding site. The ability of resveratrol and SRT1720 to activate SIRT1 was examined against a series of N-terminal deletions in the mass spectrometry assay.

to the exposure of an allosteric binding site. In support of this hypothesis, it has been demonstrated that NAD^+ does not bind in the absence of acetylated peptide substrate to a closely related sirtuin, SIRT2 (ref. 24). A K_m-type mechanism is supported by isothermal titration calorimetry experiments that show a roughly twofold increase in the affinity for peptide substrate binding to SIRT1-C in the presence of SRT1460 ($K_{d(-SRT1460)}$ = 25.7 μM and $K_{d(+SRT1460)}$ = 14.1 μM; data not shown).

To determine whether SRT1460 and SRT1720 bind and activate the enzyme at the same molecular site as resveratrol, an isobologram analysis was performed. A concentration matrix of two compounds, resveratrol versus SRT1720 and SRT1720 versus SRT1460, was examined to determine whether the combination was antagonistic, additive or synergistic. In both cases the compound combination resulted in additivity consistent with the hypothesis that a single allosteric site exists on the SIRT1–substrate complex to which structurally diverse compounds can bind (Fig. 2c).

To define the compound binding site a series of SIRT1 truncation mutants was generated (Fig. 2d). SIRT1-A (156–664) and SIRT1-B (172–664) retain similar activity and activation to the parent construct (SIRT1-C), whereas SIRT1-D (219–664) exhibits a marked decrease in enzyme activity. SIRT1-D expresses poorly in *Escherichia coli* and may lose activity simply due to misfolding. The SIRT1-E (225–664), SIRT1-F (230–664), SIRT1-G (235–664), and SIRT1-H (240–664) constructs exhibit enzymatic activity comparable to SIRT1-C, but the ability to be activated by SRT1720 and resveratrol is significantly decreased. The SIRT1 truncation data suggest that amino acids 240–664 comprise the core catalytic domain, and that the amino acids 183–225 N-terminal to the core domain are important in defining the compound binding site.

The therapeutic potential of SIRT1 activators to treat insulin resistance and diabetes was tested in three *in vivo* models of type 2 diabetes. SRT1720 exhibited a pharmacokinetic profile (Fig. 3a) suitable for *in vivo* evaluation in both mouse (bioavailability = 50%, terminal $t_{1/2}$ = ~5 h, Area Under the Curve (AUC) = 7,892 ng h^{-1} ml^{-1}) and rat (bioavailability = 25%, terminal $t_{1/2}$ = ~8.4 h, AUC = 3,714 ng h^{-1} ml^{-1}). SRT501, a reformulated version of resveratrol with improved bioavailability (11% bioavailability, terminal $t_{1/2}$ of ~ 1 h and an AUC of 10,524 ng h^{-1} ml^{-1}), was also examined in genetically obese mice ($Lep^{ob/ob}$) and diet-induced obesity (DIO) mice. SRT501 (500 or 1,000 mg per kg (body weight)) and SRT1720 (100 mg per kg (body weight)) were dosed in all efficacy studies once daily by oral gavage. During these studies we did not measure food intake or metabolic rates so there may be uncertainty about how the compound is working. We did measure body weight during these studies and this did not change.

In DIO mice, fasting blood glucose levels are elevated (120–150 mg dl^{-1} range) after being placed on a high-fat diet. Administration of SRT1720 reduced fed glucose levels after 1 week of treatment with further reduction after 3 weeks of treatment (Fig. 3b) that continued through 10 weeks of dosing. Glucose excursion during an intraperitoneal glucose tolerance test was also significantly reduced in the SRT1720 group, and comparable to rosiglitazone, a PPARγ activator that has been used to treat type 2 diabetes (Fig. 3c; glucose AUC: control, 603 ± 32 mg h^{-1} dl^{-1}; SRT1720, 462 ± 25 mg h^{-1} dl^{-1}; rosiglitazone, 496 ± 20 mg h^{-1} dl^{-1}). SRT1720 did not have an effect on fasting glucose in chow-fed mice (data not shown), indicating that pharmacological SIRT1 activation is unlikely to induce hypoglycaemia.

DIO mice are insulin resistant and hyperinsulinaemic (3.9 ± 0.7 ng dl^{-1}) compared to normal chow fed controls (0.4 ± 0.1 ng dl^{-1}). SRT1720 significantly reduced the hyperinsulinaemia after 4 weeks, partially normalizing elevated insulin levels similar to rosiglitazone treatment (Fig. 3d). Consistent with the effects on improved glucose tolerance and fasting insulin levels, a reduction in glucose levels during an insulin tolerance test was greater in SRT1720 or rosiglitazone treated animals (Fig. 3e). SRT1720 did not reduce insulin levels in

©2007 Nature Publishing Group

mice fed a normal chow diet that are not insulin resistant (Fig. 3d). The effect of SRT1720 on blood glucose and insulin levels is not secondary to weight loss as there was no difference in the weight of mice treated with SRT1720 compared to vehicle-treated DIO mice (data not shown).

As observed in calorie restriction[6,25], SRT1720 treatment (10 weeks) increases mitochondrial capacity by 15% in gastrocnemius muscle as measured by citrate synthase activity (Fig. 3f). In DIO mice SRT1720 mimics several of the effects observed after calorie restriction including improved insulin sensitivity, normalized glucose and insulin levels, and elevated mitochondrial capacity.

To study SRT1720 in a genetic model of type 2 diabetes, $Lep^{ob/ob}$ mice were placed on a high-fat diet to induce a diabetic state and then dosed with SRT1720. SRT1720 treatment significantly reduced fasting blood glucose to near normal levels (Fig. 3g) after 1 week. SRT501, a much less potent SIRT1 activator, also lowered fasting blood glucose in $Lep^{ob/ob}$ mice after 3 weeks (1,000 mg per kg (body weight)) and in DIO mice after 4 weeks (500 mg per kg (body-weight)) of oral dosing (Fig. 3h, i). SRT501 also reduces hyperinsulinaemia in DIO mice measured as fed insulin levels (data not shown). As observed with SRT1720, no difference in body weight was evident after treatment with SRT501 in $Lep^{ob/ob}$ mice.

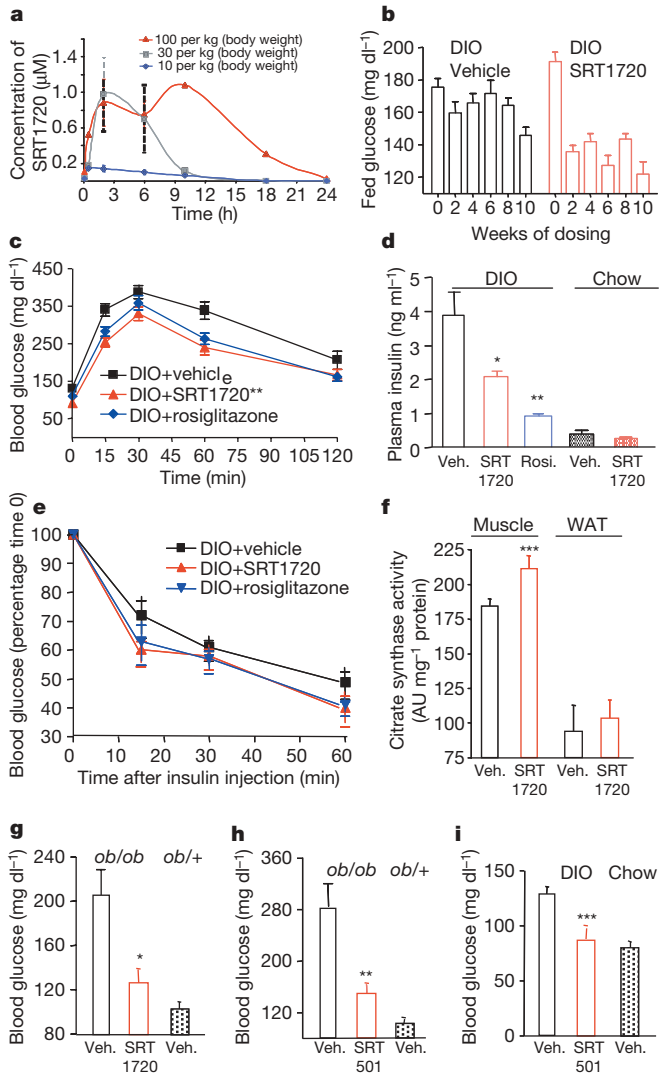


Figure 3 | SIRT1 activators in mouse models of type 2 diabetes. **a**, Plasma levels of SRT1720 after administration by oral gavage. **b**, Effect of SRT1720 treatment over 10 weeks on fed plasma glucose levels in DIO mice. **c**, Glucose excursion during an intraperitoneal glucose tolerance test (5 weeks) in DIO mice treated with the indicated compounds. **d**, Plasma insulin levels after 10 weeks treatment with indicated compounds. **e**, Glucose response in DIO mice during an insulin tolerance test after 10 weeks treatment with the indicated compounds. **f**, Skeletal muscle citrate synthase activity (V_{max} mg^{-1} protein, 11 weeks, $n = 5$). **g**, Effect of SRT1720 treatment in diabetic $Lep^{ob/ob}$ mice (1 week). **h**, Effect of SRT501 treatment (1,000 mg per kg (body weight)) in diabetic $Lep^{ob/ob}$ mice after 2 weeks. **i**, Effect of SRT501 treatment (500 mg per kg (body weight), 4 weeks) in DIO mice. All studies consisted of ten mice per group unless noted. Statistics were conducted as an ANOVA; asterisk $P < 0.05$, double asterisk $P < 0.01$ and triple asterisk $P < 0.001$. Data are expressed as mean ± s.e.m.

We also performed *in vivo* studies in the Zucker *fa/fa* rat, which is a genetically obese rodent model for studying insulin resistance. Rats were treated with vehicle or SRT1720 for 4 weeks. Body weight was not different between the two groups throughout the study. After 3 weeks treatment, fed blood glucose was significantly lower in *fa/fa* rats (Fig. 4a) treated with SRT1720 compared to vehicle-treated (140 ± 5 versus 110 ± 3 mg dl^{-1}, vehicle versus SRT1720, $P < 0.001$). In support of improved glucose metabolism after SIRT1 activation, both the glucose (AUC: 4,550 ± 463 versus 2,490 ± 236 mg min^{-1}dl^{-1}, vehicle versus SRT1720, $P < 0.01$) and insulin (AUC: 277 ± 32 vs. 127 ± 25 ng min^{-1}dl^{-1}, vehicle versus SRT1720, $P < 0.01$) responses during an oral glucose tolerance test were significantly improved in the SRT1720-treated group (Fig. 4b, c). The decrease in glucose in the face of reduced hyperinsulinaemia indicates improved insulin sensitivity. To assess more definitively insulin action after SRT1720 treatment, hyperinsulinaemic-euglycaemic clamps were performed on *fa/fa* rats after 4 weeks of treatment. Consistent with improved glucose tolerance, the glucose infusion rate required to maintain euglycaemia was ~35% higher in SRT1720-treated *fa/fa* rats (Fig. 4d, $P < 0.01$), and the total glucose disposal rate was increased by ~20% (Fig. 4e, $P < 0.05$). The insulin-stimulated glucose disposal rate, which primarily represents glucose disposal into skeletal muscle, was 55% higher in the SRT1720-treated group (Fig. 4f, $P < 0.05$). Along with the increased glucose disposal rate and glucose infusion rate, liver insulin sensitivity was significantly enhanced in the SRT1720 group, as evidenced by a greater suppression of hepatic glucose production during the clamp (Fig. 4g, $P < 0.05$). Adipose tissue insulin sensitivity was also improved after SIRT1 activation (Fig. 4h), as indicated by lower (~40%) plasma free fatty acid concentration during the clamp in the SRT1720-treated group (0.48 ± 0.03 versus 0.82 ± 0.10 mmol l^{-1}, SRT1720 versus vehicle, $P < 0.05$).

Recent studies have suggested that SIRT1 might be important in regulating hepatic gluconeogenesis, through deacetylation and activation of PGC-1α[26,27]. However, transgenic mice that overexpress SIRT1 exhibit improved insulin sensitivity and are not hyperglycaemic[28]. To investigate if SRT1720 treatment alters hepatic gluconeogenesis, *fa/fa* rats were subjected to a pyruvate tolerance test (PTT) after 3 weeks of treatment. The glucose response during the PTT was markedly blunted in SRT1720-treated *fa/fa* rats relative to vehicle (Fig. 4i), demonstrating decreased hepatic gluconeogenesis. These results show that the overall *in vivo* effect of SIRT1 activation is to ameliorate the heightened gluconeogenic capacity of Zucker *fa/fa* rats and not to stimulate hepatic glucose production. In addition, because Zucker *fa/fa* rats exhibit hepatic insulin resistance, the decreased gluconeogenesis after SRT1720 treatment is further evidence of improved hepatic insulin sensitivity.

To address the therapeutic potential of SIRT1 activation for the improvement of age-related diseases, we have identified and characterized novel small molecule activators of SIRT1 both *in vitro* and *in vivo*. These compounds are 1,000-fold more potent activators than, and structurally unrelated to, resveratrol. Similar to resveratrol, these compounds bind directly to the SIRT1-acetylated peptide complex at the same site and lower the K_m for peptide substrate resulting in a more productive catalytic complex. SIRT1 deletion experiments indicate that amino acids 183–225 are critical for maintaining activation of SIRT1 by these compounds and define the allosteric binding site. It is possible that acetylated peptide binding to SIRT1 induces a

©2007 Nature Publishing Group

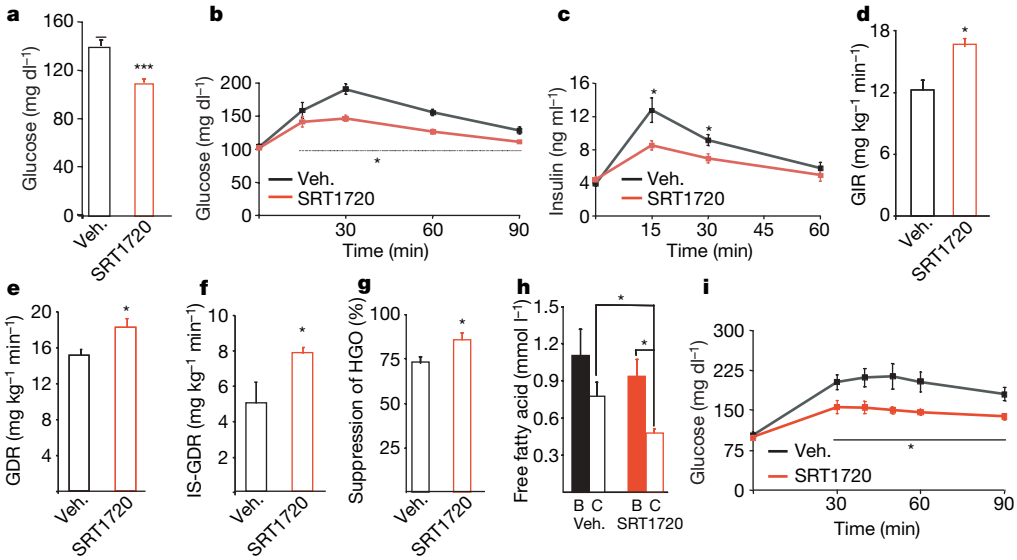


Figure 4 | SIRT1 activator SRT1720 in the Zucker *fa/fa* rat model. a, Post-absorptive blood glucose (3 weeks). **b**, **c**, Glucose and insulin responses during an oral glucose tolerance test (3.5 weeks, ANOVA). After 4 weeks a hyperinsulinaemic-euglycaemic clamp study was conducted. **(d–f)** Glucose infusion rate (GIR), glucose disposal rate (GDR) and insulin-stimulated glucose disposal rate (IS-GDR) were significantly enhanced. **g**, Hepatic glucose output (HGO) suppression (%). **h**, Plasma fatty acid concentration during the clamp was significantly lower in SRT1720-treated animals (ANOVA). B, basal conditions. C, clamp conditions. **i**, Glucose excursion during a PTT was reduced, indicating reduced gluconeogenic capacity. All studies consisted of $n \geq 5$ rats per group. Statistics were conducted as student *t*-test unless noted otherwise; asterisk $P < 0.05$, double asterisk $P < 0.01$, and triple asterisk $P < 0.001$. Data are expressed as mean ± s.e.m.

conformational change that exposes an allosteric site in this region of the enzyme. An endogenous regulator of SIRT1 has yet to be identified, and it is tempting to speculate that an endogenous activator of SIRT1 exists and may be increased after calorie restriction and other mild physiological stresses. Selective SIRT1 activators ablate insulin resistance and diabetes in DIO mice fed a high-fat diet and in diabetic $Lep^{ob/ob}$ mice. In addition, these new SIRT1 activators ameliorate the metabolic disturbances in Zucker *fa/fa* rats. SIRT1 activators improve glucose homeostasis and insulin sensitivity in key metabolic tissues including liver, muscle and fat. These compounds seem to mimic the beneficial effects of calorie restriction on mitochondrial and metabolic function in mammals *in vivo* and hold promise for treating diseases of ageing such as type 2 diabetes.

METHODS SUMMARY

The full details of the materials and methods are presented in the Supplementary Information. The compounds, SRT1460, SRT1720, and SRT2183, were synthesized at Sirtris Pharmaceuticals. SIRT1 constructs were expressed in *E. coli* and purified by affinity, size exclusion, and/or ion exchange chromatography. The SIRT1 fluorescence polarization assay used a peptide based on p53 linked to biotin at the N-terminal end and to a fluorescent tag at the C-terminal end. The reaction was a coupled enzyme assay where the first reaction was the deacetylation reaction and the second was cleavage by trypsin at the exposed lysine. The mass spectrometry assay was used for all SAR studies and employs the same peptide as the HTS assay. For the in cell western assay (ICW) the acetylation state of p53 was determined using a polyclonal antibody against the acetylated lysine residue at position 382 in p53. The effect of test compounds on the K_m of SIRT1 for peptide substrate was examined using the mass spectrometry assay. Isothermal titration calorimetry was performed using a VP-ITC (MicroCal, Inc.). For the isobologram analysis a concentration matrix of two compounds was created and tested against SIRT1 using the mass spectrometry assay. For pharmacokinetic studies compounds were administered by oral gavage or intravenously to C57BL/6 male mice at the doses indicated. For all *in vivo* efficacy studies vehicle or test compound was administered once daily by oral gavage. For the DIO studies, C57BL/6 male mice were fed a high-fat diet until their mean body weight reached approximately 40 g. $Lep^{ob/ob}$ mice and heterozygous Lep^{ob+} mice were placed on a high-fat diet for a minimum of 1 week before the start of a study and remained on the high-fat diet for the duration of the study. Male fatty (*fa/fa*) Zucker (ZF) rats (Harlan Sprague–Dawley, Inc.) were used for the studies.

Received 3 August; accepted 17 September 2007.

1. Facchini, F. S., Hua, N., Abbasi, F. & Reaven, G. M. Insulin resistance as a predictor of age-related diseases. *J. Clin. Endocrinol. Metab.* **86**, 3574–3578 (2001).
2. Barzilai, N., Banerjee, S., Hawkins, M., Chen, W. & Rossetti, L. Caloric restriction reverses hepatic insulin resistance in aging rats by decreasing visceral fat. *J. Clin. Invest.* **101**, 1353–1361 (1998).
3. Bordone, L. & Guarente, L. Calorie restriction, SIRT1 and metabolism: understanding longevity. *Nature Rev. Mol. Cell Biol.* **6**, 298–305 (2005).
4. Cohen, H. Y. *et al.* Calorie restriction promotes mammalian cell survival by inducing the SIRT1 deacetylase. *Science* **305**, 390–392 (2004).
5. Heilbronn, L. K. *et al.* Glucose tolerance and skeletal muscle gene expression in response to alternate day fasting. *Obes. Res.* **13**, 574–581 (2005).
6. Nisoli, E. *et al.* Calorie restriction promotes mitochondrial biogenesis by inducing the expression of eNOS. *Science* **310**, 314–317 (2005).
7. Frye, R. A. Characterization of five human cDNAs with homology to the yeast SIR2 gene: Sir2-like proteins (sirtuins) metabolize NAD and may have protein ADP-ribosyltransferase activity. *Biochem. Biophys. Res. Commun.* **260**, 273–279 (1999).
8. Frye, R. A. Phylogenetic classification of prokaryotic and eukaryotic Sir2-like proteins. *Biochem. Biophys. Res. Commun.* **273**, 793–798 (2000).
9. Imai, S., Armstrong, C. M., Kaeberlein, M. & Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800 (2000).
10. Baur, J. A. *et al.* Resveratrol improves health and survival of mice on a high-calorie diet. *Nature* **444**, 337–342 (2006).
11. Howitz, K. T. *et al.* Small molecule activators of sirtuins extend *Saccharomyces cerevisiae* lifespan. *Nature* **425**, 191–196 (2003).
12. Jarolim, S. *et al.* A novel assay for replicative lifespan in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **5**, 169–177 (2004).
13. Lagouge, M. *et al.* Resveratrol improves mitochondrial function and protects against metabolic disease by activating SIRT1 and PGC-1α. *Cell* **127**, 1109–1122 (2006).
14. Wood, J. G. *et al.* Sirtuin activators mimic caloric restriction and delay ageing in metazoans. *Nature* **430**, 686–689 (2004).
15. Brink, C. B., Harvey, B. H., Bodenstein, J., Venter, D. P. & Oliver, D. W. Recent advances in drug action and therapeutics: relevance of novel concepts in G-protein-coupled receptor and signal transduction pharmacology. *Br. J. Clin. Pharmacol.* **57**, 373–387 (2004).
16. Luo, J. *et al.* Negative control of p53 by Sir2α promotes cell survival under stress. *Cell* **107**, 137–148 (2001).
17. Vaziri, H. *et al.* hSIR2(SIRT1) functions as an NAD-dependent p53 deacetylase. *Cell* **107**, 149–159 (2001).
18. Napper, A. D. *et al.* Discovery of indoles as potent and selective inhibitors of the deacetylase SIRT1. *J. Med. Chem.* **48**, 8045–8054 (2005).
19. Kaeberlein, M., McVey, M. & Guarente, L. The SIR2/3/4 complex and SIR2 alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Genes Dev.* **13**, 2570–2580 (1999).

©2007 Nature Publishing Group

20. Rogina, B. & Helfand, S. L. Sir2 mediates longevity in the fly through a pathway related to calorie restriction. *Proc. Natl Acad. Sci. USA* **101,** 15998–16003 (2004).
21. Anderson, R. M., Bitterman, K. J., Wood, J. G., Medvedik, O. & Sinclair, D. A. Nicotinamide and PNC1 govern lifespan extension by calorie restriction in *Saccharomyces cerevisiae. Nature* **423,** 181–185 (2003).
22. Lin, S. J., Defossez, P. A. & Guarente, L. Requirement of NAD and SIR2 for life-span extension by calorie restriction in *Saccharomyces cerevisiae. Science* **289,** 2126–2128 (2000).
23. Sarabu, R. & Grimsby, J. Targeting glucokinase activation for the treatment of type 2 diabetes–a status review. *Curr. Opin. Drug Discov. Dev.* **8,** 631–637 (2005).
24. Borra, M. T., Langer, M. R., Slama, J. T. & Denu, J. M. Substrate specificity and kinetic mechanism of the Sir2 family of NAD+-dependent histone/protein deacetylases. *Biochemistry* **43,** 9877–9887 (2004).
25. Heilbronn, L. K. *et al.* Effect of 6-month calorie restriction on biomarkers of longevity, metabolic adaptation, and oxidative stress in overweight individuals: a randomized controlled trial. *J. Am. Med. Assoc.* **295,** 1539–1548 (2006).
26. Rodgers, J. T. *et al.* Nutrient control of glucose homeostasis through a complex of PGC-1α and SIRT1. *Nature* **434,** 113–118 (2005).
27. Rodgers, J. T. & Puigserver, P. Fasting-dependent glucose and lipid metabolic response through hepatic sirtuin 1. *Proc. Natl Acad. Sci. USA* **104,** 12861–12866 (2007).
28. Banks, A. *et al.* Overexpression of the Sirtuin SIRT1 increases insulin sensitivity in aging mice. *Diabetes* 56 (S1), 0234-OR (2007).

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

Acknowledgements We thank C. Ozbal and W. LaMarr from BioTrove, Inc. for running the mass spectrometry samples; S. Schaertl, D. Winkler, N. Fay and T. Hesterkamp for work on the SIRT1 fluorescence polarization assay development; M. Saberi, P. P. Li , M. Lu, and A. Hevener for assistance and advice with the Zucker *fa/fa* studies; P. Romero, K. Normington, and M. Dipp for experimental advice and comments on the manuscript; M. Inghilterra for help in data analysis and data mining. D.A.S. is supported by an Ellison Medical Foundation Senior Scholarship, and grants from NIH/NIA and the Paul F. Glenn Medical Foundation. J.M.O. is supported by a University of California Discovery Biostar grant and NIH. S.S. is supported by a Mentor-Based Postdoctoral Fellowship from the American Diabetes Association awarded to J.M.O.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full text HTML version of the paper at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.H.W. (cwestphal@sirtrispharma.com).

©2007 Nature Publishing Group

LETTERS

Interpretation of the sonic hedgehog morphogen gradient by a temporal adaptation mechanism

Eric Dessaud[1], Lin Lin Yang[2]†, Katy Hill[1], Barny Cox[1], Fausto Ulloa[1], Ana Ribeiro[1], Anita Mynett[1], Bennett G. Novitch[2]† & James Briscoe[1]

Morphogens act in developing tissues to control the spatial arrangement of cellular differentiation[1,2]. The activity of a morphogen has generally been viewed as a concentration-dependent response to a diffusible signal, but the duration of morphogen signalling can also affect cellular responses[3]. One such example is the morphogen sonic hedgehog (SHH). In the vertebrate central nervous system and limbs, the pattern of cellular differentiation is controlled by both the amount and the time of SHH exposure[4–7]. How these two parameters are interpreted at a cellular level has been unclear. Here we provide evidence that changing the concentration or duration of SHH has an equivalent effect on intracellular signalling. Chick neural cells convert different concentrations of SHH into time-limited periods of signal transduction, such that signal duration is proportional to SHH concentration. This depends on the gradual desensitization of cells to ongoing SHH exposure, mediated by the SHH-dependent upregulation of patched 1 (PTC1), a ligand-binding inhibitor of SHH signalling[8]. Thus, in addition to its role in shaping the SHH gradient[8–10], PTC1 participates cell autonomously in gradient sensing. Together, the data reveal a novel strategy for morphogen interpretation, in which the temporal adaptation of cells to a morphogen integrates the concentration and duration of a signal to control differential gene expression.

How both concentration and duration of morphogen signalling determine cell pattern is poorly understood[1] (see Supplementary Fig. 1a). We focused on three transcription factors that respond to differential SHH signalling in progenitors of the neural tube. OLIG2 and NKX2.2, expressed in the ventral neural tube of chick (Fig. 1a), depend on SHH signalling for their expression[4,11,12]. In contrast, PAX7 expression is repressed by SHH signalling[13] and is restricted to dorsal neural tube progenitors (Fig. 1a). We confirmed that these proteins respond to different levels of SHH signalling using an *ex vivo* assay of intermediate region naive neural plate explants[14] (Fig. 1b–d). In agreement with previous studies[13,15], changes in SHH concentration controlled the expression of these genes in a manner corresponding to their *in vivo* expression patterns (Fig. 1e). Similar gene expression responses were obtained by generating, *in vivo*, a gradient of GLI transcriptional activity[16,17], the transcriptional effectors of SHH signalling. Furthermore, manipulation of the activity of the transmembrane protein smoothened (SMO), which transduces SHH signalling intracellularly[18], was also sufficient to confer graded responses to neural cells (Supplementary Fig. 2).

We asked how the response of cells to SHH develops over time. We assayed intermediate region neural plate explants exposed to SHH for 6 h to 24 h (Fig. 1f, g and Supplementary Table 1). NKX2.2 induction was delayed compared to OLIG2, taking >12 h compared to ~6 h for OLIG2. Moreover, NKX2.2-inducing concentrations of SHH (≥2 nM) produced a transient expression of OLIG2 (Fig. 1f, g). Thus, during the first 12 h, ≥1 nM SHH generated similar amounts of OLIG2 induction (Supplementary Table 1). Only after 12 h were distinct responses apparent. By 18 h, OLIG2 and NKX2.2 co-expression was evident in some progenitors treated with ≥2 nM SHH (Fig. 1f). These results are consistent with the sequential onset of OLIG2 and NKX2.2 expression *in vivo*[10,17]. Furthermore, the explant data predict that, *in vivo*, NKX2.2 should be induced in cells that previously expressed OLIG2. Genetic lineage tracing in mice harbouring an *Olig2* allele engineered to encode Cre recombinase confirmed this (Fig. 1h). Thus, compared to OLIG2, induction of NKX2.2 requires a higher concentration and longer duration of SHH exposure.

To investigate the reason for the temporal and concentration dependence of the response, we analysed the output of the SHH signal transduction pathway. We assayed GLI activity using a reporter plasmid[19] (GBS-Luc; see Fig. 2a and Methods). Taking advantage of the short half-life of luciferase[20] (~3 h), we measured GBS-Luc activity every 6 h. For concentrations of SHH ≥1 nM, GLI activity was similar during the first 12 h (Fig. 2a), the period when these concentrations induce OLIG2. Then, with the exception of the highest SHH concentration, GLI activity decreased over time, with a rate inversely proportional to SHH concentration (Fig. 2a). Notably, the time at which 1 nM and 4 nM SHH produced differences in the level of GLI activity corresponded to the detection of differences in gene expression (Fig. 1f, g and Supplementary Table 1). The data indicate, therefore, that cells become progressively desensitized to ongoing SHH signalling (Supplementary Fig. 1). Initially 1 nM and 4 nM SHH produce similar levels of GLI activity, then the level of GLI activity begins to fall, with a higher rate of decrease in cells exposed to lower concentrations. This suggests a mechanism for gradient sensing in which 'temporal adaptation' to the ligand transforms the extracellular concentration of morphogen into a time-limited period of signal transduction, such that the duration of signalling is proportional to ligand concentration.

This model predicts that the response of cells relies not only on the level but also on the duration of intracellular signal transduction. To test this, we compared GLI activity and gene expression in neural cells containing an endogenous source of SHH. In explants consisting of SHH-producing notochord and floor plate together with ventral regions of the neural tube (hereafter called NVF explants), the expression of OLIG2 and NKX2.2 was induced sequentially, 12 h and 18 h after the start of culture, respectively (Fig. 2f–h). During this period the level of GLI activity remained approximately constant, confirming that the switch to NKX2.2 expression was not associated with an increase in GLI activity (Fig. 2b, f–h). To examine whether the maintenance of an OLIG2-expressing state depends on the downregulation

[1]Developmental Neurobiology, National Institute for Medical Research Mill Hill, London NW7 1AA, UK. [2]Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. †Present address: Department of Neurobiology, David Geffen School of Medicine, University of California, Los Angeles, California 90095, USA.

©2007 Nature Publishing Group

of GLI activity, NVF explants were cultured for 12 h and then transferred to media containing cyclopamine, a small-molecule antagonist of SHH signalling (Supplementary Fig. 2). GLI activity and gene expression were monitored at 18 h (Fig. 2c–e, i). Addition of 200 nM cyclopamine at 12 h resulted in a twofold decrease in GLI activity and a failure to induce NKX2.2 expression (Fig. 2c, e, i) without inhibiting OLIG2 expression. Furthermore, addition of 400 nM cyclopamine inhibited GLI activity to background levels, leading to a complete loss of NKX2.2 expression and a decrease in the number of OLIG2-expressing cells (Fig. 2c, e, and data not shown). These data indicate that the duration of GLI activity is crucial for determining the cellular response to SHH. To test the converse prediction of the model, we assayed the effect of extending the period of SHH signalling (Fig. 2j). Consistent with the model, prolonging SHH signalling resulted in peak OLIG2 and NKX2.2 induction at lower SHH concentrations after 48 h of exposure compared to 24 h and 36 h (Fig. 2j).

Two mechanisms that could account for the adaptation of cells to SHH signalling are the loss of a factor necessary for signal transduction or the induction of inhibitors of signal transduction, such as the SHH-binding protein PTC1 (refs 21–23). We assessed GLI activity in a time series of intermediate region neural plate explants in which SHH signalling was induced using purmorphamine, a small-molecule agonist of SMO[24] (Fig. 3a and Supplementary Fig. 2). In contrast to the response of neural cells to SHH (Fig. 2a), even though a diminution in GLI activity was observed in treated cells, there was no correlation between the rate of decrease and the concentration of purmorphamine (Fig. 3a). This suggests that the profile of GLI activity evoked by SHH is shaped mainly by an adaptation mechanism acting upstream of SMO.

As PTC1 is a well-established inhibitor of SHH signalling[8], we tested whether it is required cell autonomously for the temporal adaptation of cells to SHH. Small interfering RNAs (siRNAs) were used to block chick PTC1 (cPTC1) induction (Fig. 3b and Supplementary Figs 3–6). *In ovo* transfection of *cPTC1* siRNAs inhibited the upregulation of *cPTC1* but did not completely abolish its expression (Supplementary Fig. 4, and data not shown). As a result, and consistent with observations in mouse embryos containing an un-inducible allele of *Ptc1*[10], ventralization of the neural tube, but no ligand-independent SHH signalling, was observed in siRNA-transfected embryos or in intermediate region neural plate explants (Supplementary Figs 3 and 5). We therefore assayed whether blocking cPTC1 upregulation altered

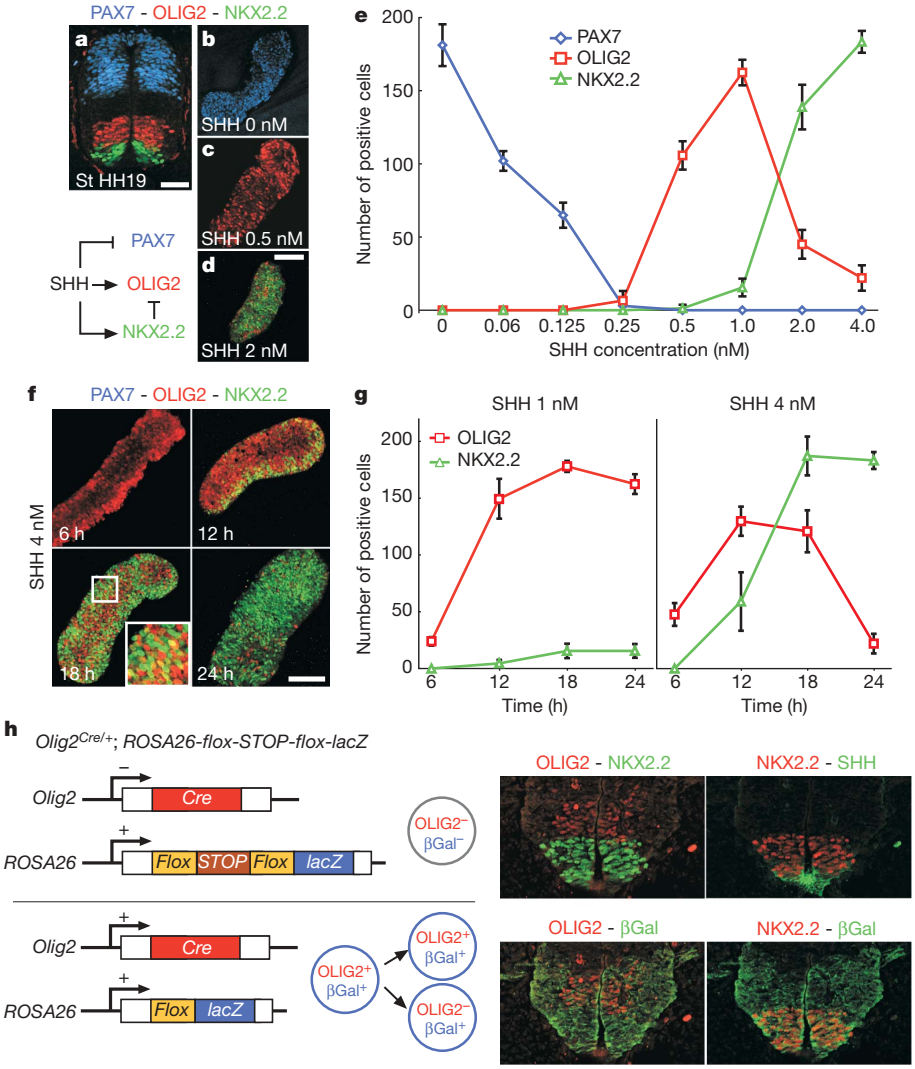


Figure 1 | The temporal and concentration dependence of the response of neural cells to SHH. **a**, PAX7, OLIG2 and NKX2.2 expression in HH stage 19 chick neural tube (scale bar, 50 μm). **b–d**, Intermediate region neural plate explants cultured for 24 h in the indicated concentrations of SHH and assayed simultaneously for the expression of PAX7, OLIG2 and NKX2.2. OLIG2 expression decreases as NKX2.2 is induced, consistent with the ability of NKX2.2 to repress OLIG2[12]. **e**, Quantification of cells expressing PAX7, OLIG2 and NKX2.2 in intermediate region neural plate explants ($n \geq 5$; number of cells per unit ± s.d.). **f**, PAX7, OLIG2 and NKX2.2 expression in intermediate region neural plate explants grown with 4 nM SHH for 6 h, 12 h, 18 h or 24 h (scale bar, 100 μm). **g**, Quantification of cells expressing OLIG2 and NKX2.2 in intermediate region neural plate explants ($n \geq 5$; number of cells per unit ± s.d.). **h**, As well as OLIG2$^+$ progenitors, NKX2.2$^+$ progenitors express LacZ in *Olig2*$^{Cre/+}$;*ROSA26-flox-STOP-flox-lacZ* E10.5 mouse embryos.

©2007 Nature Publishing Group

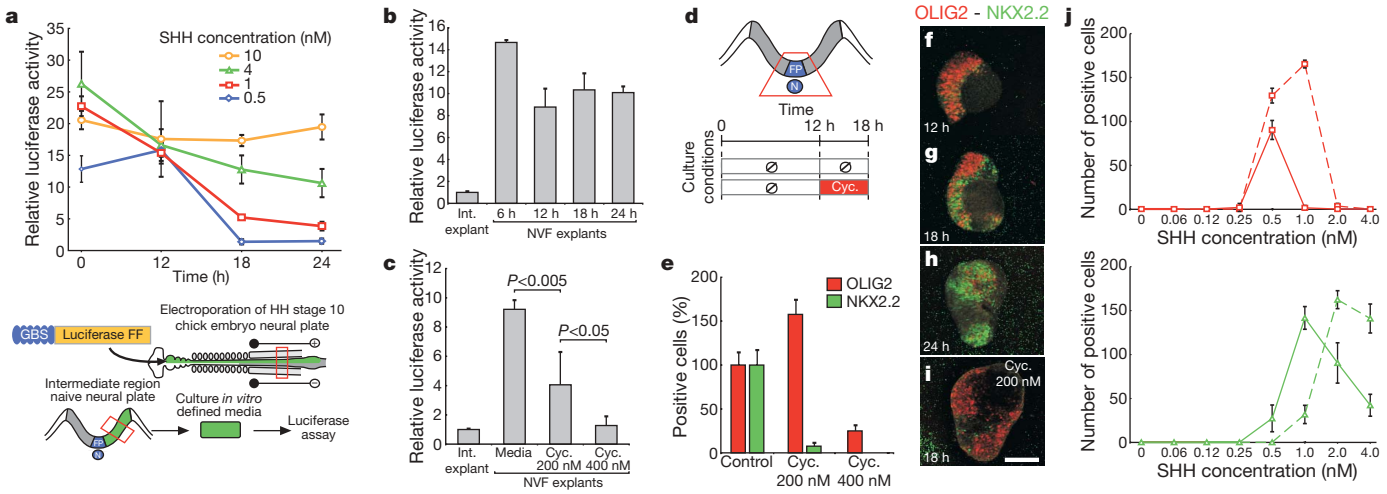


Figure 2 | Sustained GLI activity is required for discrimination between *OLIG2* and *NKX2.2* gene induction. **a**, GLI activity (relative GLI activity ± s.e.m.) measured with GBS-Luc in intermediate region neural plate explants treated with the indicated concentrations of SHH for the designated times. GLI activity is measured relative to untreated explants. **b**, GLI activity (relative GLI activity ± s.e.m.) assayed in cells exposed to an endogenous source of SHH. GBS-Luc assays were performed using NVF explants (**d**) compared to intermediate region neural plate explants, which lack SHH expression. **c**, Graph of relative GLI activity ± s.e.m. according to conditions in **d** (*P*-values from student's *t*-test). Cyc., cyclopamine. Decreasing the duration of GLI activity in NVF explants alters gene expression (**e**, **i**). **e**, Quantification of OLIG2$^+$ and NKX2.2$^+$ cells relative to control condition (18 h culture; ± s.d.) in NVF explants. **f**–**h**, Temporal profile of NKX2.2 and OLIG2 expression in NVF explants. **j**, Numbers of cells expressing OLIG2 (red) and NKX2.2 (green) in intermediate region neural plate explants ($n \geq 5$; number of cells per unit ± s.d.) exposed to the indicated concentrations of SHH for 36 h (dashed lines) and 48 h (solid lines).

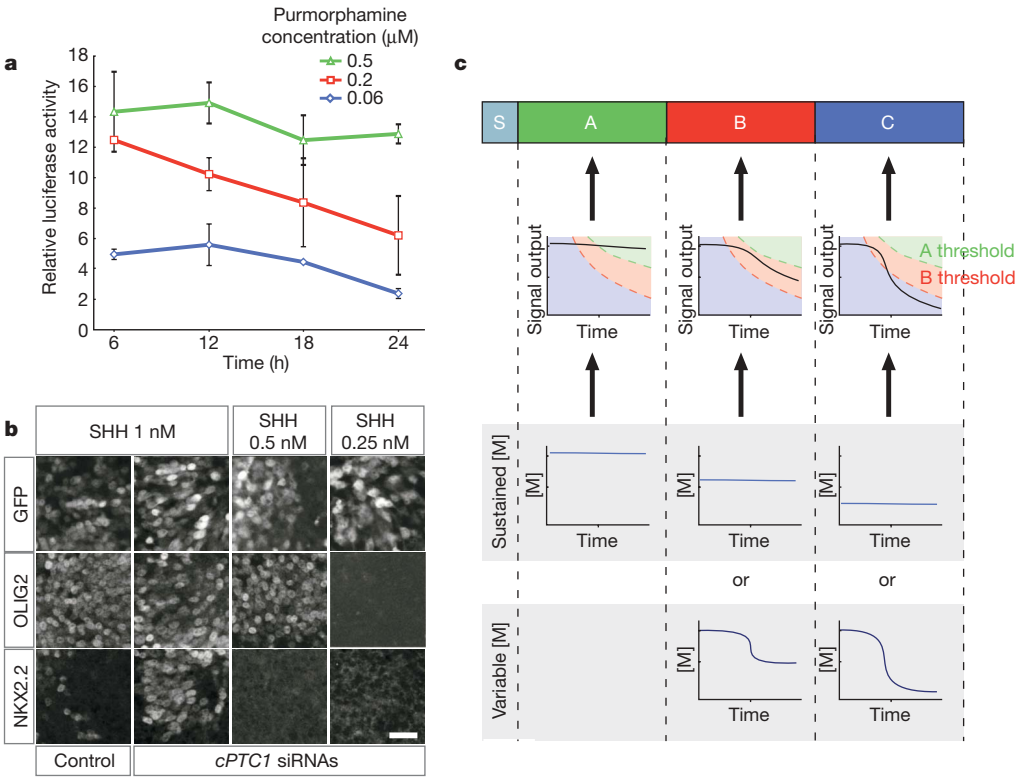


Figure 3 | A cell autonomous function for PTC1-mediated desensitization is required for gradient sensing. **a**, Temporal profile of GLI activity, measured with GBS-Luc, in intermediate region neural plate explants treated with the indicated concentrations of purmorphamine for the designated times (relative GLI activity ±s.e.m.). **b**, Blocking cPTC1 upregulation disrupts the interpretation of high concentrations of SHH. Control or *cPTC1* siRNA electroporated intermediate region neural plate explants where cultured with the indicated concentrations of SHH for 24 h and assayed for NKX2.2 and OLIG2 (scale bar, 25 μm; $n = 10$ for control siRNA; $n = 12$ for *cPTC1* siRNAs). **c**, A model for the morphogen patterning of a tissue divided into three domains characterized by differential gene expression (A, B and C). Graded morphogen signalling (M) from the source (S) is transduced into an episode of signal transduction (signal output). Initially, each M concentration ([M]) generates similar levels of signal output. However, signal output declines more rapidly in cells exposed to lower [M]. Similar profiles of signal output can be generated by a sustained [M] or by transient exposure to M (variable [M]). This model illustrates how two different parameters of a graded morphogen—amount of ligand and duration of exposure—are transformed to an equivalent output by a responding cell.

©2007 Nature Publishing Group

the response of cells to SHH exposure. Explants were transfected with control or *cPTC1* siRNAs and exposed to 0.25–1 nM SHH. Transfection of control siRNAs did not change the gene expression profile induced by SHH (Fig. 3b). However, in contrast to control explants exposed to 1 nM SHH that predominately expressed OLIG2, increased numbers of NKX2.2-expressing cells were observed in 1 nM SHH-treated intermediate region neural plate explants transfected with *cPTC1* siRNAs. A similar result was obtained when siRNAs were transfected individually (Supplementary Fig. 6). The siRNA effect was dependent on SHH concentration: *cPTC1* siRNAs did not alter the response of intermediate region neural plate explants exposed to $<$0.5 nM SHH, indicating that the blockade of cPTC1 upregulation does not result in NKX2.2 induction in response to all concentrations of SHH. These data indicate that gradient sensing requires upregulation of PTC1. This complements the previously identified ability of PTC1 to influence the spread of ligand by sequestering SHH protein[8–10]. Thus, upregulation of PTC1 by SHH has both cell autonomous and non-autonomous functions in morphogen interpretation.

Adaptation to extracellular signals has previously been proposed to permit the sensing and transduction of a signal over large concentration ranges[25,26]. This study indicates that the same strategy is used to measure the concentration of a ligand during developmental patterning. The progressive adaptation of cells to SHH transforms ligand exposure into periods of increased GLI activity, the duration of which is proportional to SHH concentration (Fig. 3c). This allows cells to integrate both the concentration and duration of a graded morphogen (Supplementary Fig. 1). The transduction of a morphogen concentration into a duration of transcriptional activity offers an alternative to the established mechanisms of graded signalling[27,28], which explains how the amount and duration of SHH exposure determine similar cellular responses[3,6,7].

METHODS SUMMARY

Neural plate explants, immunohistochemistry and *in situ* hybridization. Explants isolated from Hamburger and Hamilton (HH) stage 10 chick embryos were cultured as described[13,14]. Cyclopamine (Toronto Research Chemicals) and purmorphamine (Calbiochem) were dissolved in 100% ethanol and DMSO, respectively. Reagents and protocols for immunohistochemistry and *in situ* hybridization have been described[12–14,29].

Chick *in ovo* electroporation, luciferase assays and siRNAs. HH stage 10 embryos were electroporated *in ovo* and incubated for 2 h before dissecting explants. Luciferase assays with GBS-Luc[19] were carried out as previously described[17]. RNA duplexes targeting chick *PTC1* and, as a control, human lamin were from Dharmacon. These were co-electroporated *in ovo* with a plasmid expressing green fluorescent protein.

***Olig2-Cre* lineage tracing.** *Olig2-Cre* mice were generated in the Jessell laboratory, Columbia University. Standard homologous recombination procedures were used to replace exon 2 of the *Olig2* locus with Cre recombinase. *Olig2-Cre*/+ mice were mated with *ROSA26* reporter mice[30], and embryos analysed at E10.5.

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

Received 6 September; accepted 3 October 2007.

1. Gurdon, J. B. & Bourillot, P. Y. Morphogen gradient interpretation. *Nature* **413**, 797–803 (2001).
2. Wolpert, L. One hundred years of positional information. *Trends Genet.* **12**, 359–364 (1996).
3. Pages, F. & Kerridge, S. Morphogen gradients. A question of time or concentration? *Trends Genet.* **16**, 40–44 (2000).
4. Briscoe, J. & Ericson, J. Specification of neuronal fates in the ventral neural tube. *Curr. Opin. Neurobiol.* **11**, 43–49 (2001).
5. Ahn, S. & Joyner, A. L. Dynamic changes in the response of cells to positive hedgehog signaling during mouse limb patterning. *Cell* **118**, 505–516 (2004).
6. Harfe, B. D. *et al.* Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* **118**, 517–528 (2004).
7. Yang, Y. *et al.* Relationship between dose, distance and time in Sonic Hedgehog-mediated regulation of anteroposterior polarity in the chick limb. *Development* **124**, 4393–4404 (1997).
8. Chen, Y. & Struhl, G. Dual roles for patched in sequestering and transducing Hedgehog. *Cell* **87**, 553–563 (1996).
9. Briscoe, J., Chen, Y., Jessell, T. M. & Struhl, G. A hedgehog-insensitive form of patched provides evidence for direct long-range morphogen activity of sonic hedgehog in the neural tube. *Mol. Cell* **7**, 1279–1291 (2001).
10. Jeong, J. & McMahon, A. P. Growth and pattern of the mammalian neural tube are governed by partially overlapping feedback activities of the hedgehog antagonists patched 1 and Hhip1. *Development* **132**, 143–154 (2005).
11. Lei, Q. *et al.* Wnt signaling inhibitors regulate the transcriptional response to morphogenetic Shh-Gli signaling in the neural tube. *Dev. Cell* **11**, 325–337 (2006).
12. Novitch, B. G., Chen, A. I. & Jessell, T. M. Coordinate regulation of motor neuron subtype identity and pan-neuronal properties by the bHLH repressor Olig2. *Neuron* **31**, 773–789 (2001).
13. Ericson, J. *et al.* Pax6 controls progenitor cell identity and neuronal fate in response to graded Shh signaling. *Cell* **90**, 169–180 (1997).
14. Yamada, T., Pfaff, S. L., Edlund, T. & Jessell, T. M. Control of cell pattern in the neural tube: motor neuron induction by diffusible factors from notochord and floor plate. *Cell* **73**, 673–686 (1993).
15. Ericson, J., Morton, S., Kawakami, A., Roelink, H. & Jessell, T. M. Two critical periods of Sonic Hedgehog signaling required for the specification of motor neuron identity. *Cell* **87**, 661–673 (1996).
16. Lei, Q., Zelman, A. K., Kuang, E., Li, S. & Matise, M. P. Transduction of graded Hedgehog signaling by a combination of Gli2 and Gli3 activator functions in the developing spinal cord. *Development* **131**, 3593–3604 (2004).
17. Stamataki, D., Ulloa, F., Tsoni, S. V., Mynett, A. & Briscoe, J. A gradient of Gli activity mediates graded Sonic Hedgehog signaling in the neural tube. *Genes Dev.* **19**, 626–641 (2005).
18. Lum, L. & Beachy, P. A. The Hedgehog response network: sensors, switches, and routers. *Science* **304**, 1755–1759 (2004).
19. Sasaki, H., Hui, C., Nakafuku, M. & Kondoh, H. A binding site for Gli proteins is essential for HNF-3β floor plate enhancer activity in transgenics and can respond to Shh *in vitro*. *Development* **124**, 1313–1322 (1997).
20. Thompson, J. F., Hayes, L. S. & Lloyd, D. B. Modulation of firefly luciferase stability and impact on studies of gene regulation. *Gene* **103**, 171–177 (1991).
21. Chuang, P. T. & McMahon, A. P. Vertebrate Hedgehog signalling modulated by induction of a Hedgehog-binding protein. *Nature* **397**, 617–621 (1999).
22. Goodrich, L. V., Johnson, R. L., Milenkovic, L., McMahon, J. A. & Scott, M. P. Conservation of the hedgehog/patched signaling pathway from flies to mice: induction of a mouse patched gene by Hedgehog. *Genes Dev.* **10**, 301–312 (1996).
23. Marigo, V. & Tabin, C. J. Regulation of patched by sonic hedgehog in the developing neural tube. *Proc. Natl Acad. Sci. USA* **93**, 9346–9351 (1996).
24. Sinha, S. & Chen, J. K. Purmorphamine activates the Hedgehog pathway by targeting Smoothened. *Nature Chem. Biol.* **2**, 29–30 (2006).
25. Baker, M. D., Wolanin, P. M. & Stock, J. B. Signal transduction in bacterial chemotaxis. *Bioessays* **28**, 9–22 (2006).
26. Matthews, H. R. & Reisert, J. Calcium, the two-faced messenger of olfactory transduction and adaptation. *Curr. Opin. Neurobiol.* **13**, 469–475 (2003).
27. Bourillot, P. Y., Garrett, N. & Gurdon, J. B. A changing morphogen gradient is interpreted by continuous transduction flow. *Development* **129**, 2167–2180 (2002).
28. Saha, K. & Schaffer, D. V. Signal dynamics in Sonic hedgehog tissue patterning. *Development* **133**, 889–900 (2006).
29. Briscoe, J., Pierani, A., Jessell, T. M. & Ericson, J. A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. *Cell* **101**, 435–445 (2000).
30. Soriano, P. Generalized *lacZ* expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21**, 70–71 (1999).

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

Acknowledgements We thank H. Sasaki for the GBS-Luc reporter construct. We are grateful to S. Kales, B. Han and M. Mendelsohn for help in generating *Olig2-Cre* mice and to A. Gould, J. P. Vincent and V. Ribes for discussions. This work was supported by the MRC (UK). B.G.N. is supported by grants from the Whitehall Foundation, the March of Dimes Foundation, the NINDS and the University of Michigan Biomedical Scholars Program. *Olig2-Cre* mice were generated with support from HHMI and NINDS to T. M. Jessell. E.D. is supported by EMBO and Marie Curie Fellowships.

Author Contributions E.D. and J.B. conceived the work, performed the experiments, analysed data and wrote the manuscript. L.L.Y and B.G.N. analysed *Olig2-Cre* mice. K.H., B.C., F.U., A.R. and A.M. provided materials and technical assistance.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.B. (james.briscoe@nimr.mrc.ac.uk).

©2007 Nature Publishing Group

METHODS

Neural plate explant culture. Neural plate tissue was isolated from Hamburger and Hamilton (HH) stage 10 chick embryos and cultured as described[14]. Explants consisted of the intermediate region or the notochord, floor plate and adjacent ventral region (NVF). SHH protein was generated as described[15]. Concentration of SHH was determined by comparison with previous batches. Cyclopamine (Toronto Research Chemicals) and purmorphamine (Calbiochem) were dissolved in 100% ethanol and DMSO, respectively. Quantifications of PAX7, OLIG2 and NKX2.2 expressing cells were performed on representative unit squares (containing 200 ± 20 cells) in intermediate region neural plate explants ($n \geq 5$ explants for each data point).

Immunohistochemistry and *in situ* hybridization. Antibody reagents and protocols have been described[12–14,29]. The monoclonal antibodies against PAX6 and PAX7 were obtained from the Developmental Studies Hybridoma Bank (University of Iowa). Rabbit Nkx2.2 antibody[31] (T. Jessell) was used at 1:5,000 and guinea-pig Olig2 antibody[12] was used at 1:10,000. Mouse antibody for Tuj1 (Covance Research) and goat antibody for Sox2 (ISL) were used at 1:1,000 and 1:500, respectively. Sheep antibody for GFP (Biogenesis) and rabbit antibody for GFP (Invitrogen) were used at 1:1,000. Analysis was carried out using a Leica TCS SP2 confocal microscope or a Zeiss Axioplan 2 and processed with Adobe Photoshop 7.0 software (Adobe Systems). *In situ* hybridization was performed as described[32] using a chick PTC1 probe[33] labelled with DIG-labelling reagents (Roche Diagnostics) and filtered NBT/BCIP substrate (Sigma) in 0.5% Tween (Sigma).

Generation of *cPTC1* siRNAs. RNA duplexes targeting *cPTC1* mRNA were ordered in 2′ deprotected, annealed and desalted form (Option A4, http://www.dharmacon.com). The RNA duplexes used in this study are: *cPTC1* siRNA 1 sense 5′-UGAGAGCUAUCUGCAAUAAUU-3′; *cPTC1* siRNA 2 sense 5′-AUAUGGACCAGAUUAUAGAUU-3′; *cPTC1* siRNA 3 sense 5′-CUAAUGUACUCACAACAGAUU-3′. siRNA (target sequence 5′-AACUGGACUUCCAGAAGAACA-3′) against human lamin was used as control (Dharmacon).

Chick *in ovo* electroporation and luciferase assay in intermediate region explants. HH stage 10 embryos were electroporated and incubated for 2 h before dissecting intermediate region neural plate explants. For luciferase assays, the firefly-luciferase reporter gene carrying eight copies of the GLI-binding site from the FoxA2 floor plate enhancer (GBS-Luc) was used[19]. Chick embryos were co-electroporated with GBS-Luc and CMV-Renilla luciferase normalization plasmid (Promega). Assays were carried out on explants as previously described[17]. Explants were homogenized in passive lysis buffer on ice and measurement of firefly and Renilla luciferase activities was performed using the Dual Luciferase Reporter Assay System (Promega). Each point corresponds to triplicates of 7–10 explants. Short interfering RNAs were electroporated at $2.5\,\mu g\,\mu l^{-1}$ final concentration. The neural tube of embryos was co-injected with a plasmid expressing GFP to control for electroporation efficiency.

***Olig2-Cre* lineage tracing.** Standard homologous recombination procedures were used to replace exon 2 of the *Olig2* locus with a gene encoding a nuclear-localized form of Cre recombinase. Details of the gene-targeting strategy are available on request. *Olig2-Cre/+* mice were mated with *ROSA26* reporter mice[30], and embryos were collected at E10.5 and subjected to immunohistochemical analysis as described above.

31. Briscoe, J. *et al.* Homeobox gene *Nkx2.2* and specification of neuronal identity by graded Sonic hedgehog signalling. *Nature* **398**, 622–627 (1999).
32. Dessaud, E., Salaun, D., Gayet, O., Chabbert, M. & deLapeyriere, O. Identification of lynx2, a novel member of the ly-6/neurotoxin superfamily, expressed in neuronal subpopulations during mouse development. *Mol. Cell. Neurosci.* **31**, 232–242 (2006).
33. Pearse, R. V. II, Vogan, K. J. & Tabin, C. J. Ptc1 and Ptc2 transcripts provide distinct readouts of Hedgehog signaling activity during chick embryogenesis. *Dev. Biol.* **239**, 15–29 (2001).

©2007 Nature Publishing Group

LETTERS

p53 regulates maternal reproduction through LIF

Wenwei Hu[1]*, Zhaohui Feng[1]*, Angelika K. Teresky[1] & Arnold J. Levine[1,2]

Extensive studies have shown that p53 is important in tumour prevention[1]. However, little is known about its normal physiological function. Here we show that p53 is important in reproduction, in a gender-specific manner. Significant decreases in embryonic implantation, pregnancy rate and litter size were observed in matings with *p53*$^{-/-}$ female mice but not with *p53*$^{-/-}$ male mice. The gene encoding leukaemia inhibitory factor (LIF), a cytokine critical for implantation[2], was identified as a p53-regulated gene that functions as the downstream mediator of this effect. p53 can regulate both basal and inducible transcription of LIF. Loss of p53 decreased both the level and function of LIF in uteri. Lower LIF levels were observed in the uteri of *p53*$^{-/-}$ mice than in those of *p53*$^{+/+}$ mice, particularly at day 4 of pregnancy, when transiently induced high levels of LIF were crucial for embryonic implantation. This observation probably accounts for the impaired implantation of embryos in *p53*$^{-/-}$ female mice. Administration of LIF to pregnant *p53*$^{-/-}$ mice restored maternal reproduction by improving implantation. These results demonstrate a function for p53 in maternal reproduction through the regulation of LIF. Evidence is accumulating that p53 may have a similar function in humans.

p53, the guardian of the genome, regulates the cellular response to various stresses[1]. Most studies have concentrated on *p53* as a tumour suppressor gene, and very little is known about its function in normal cellular processes. Two observations suggest that p53 could have a function in reproduction. First, some alleles of *p53* and its negative regulator, MDM2, are under positive evolutionary selection[3,4]. Second, we observed that crosses of *p53*$^{+/+}$, *p53*$^{+/-}$ and *p53*$^{-/-}$ mice indicated an unusual pattern of maternal effects on pregnancy rate and litter size. In this study, mice of two inbred strains (C57BL/6J and 129SV^{sl}) with different p53 statuses were bred, and p53 was found to be involved in reproduction in a gender-specific manner. In C57BL/6J mice, pregnancy rate and litter size for *p53*$^{-/-}$ female mice in breeding pairs with *p53*$^{+/+}$, *p53*$^{+/-}$ or *p53*$^{-/-}$ males were markedly lower, whereas they were unaffected when *p53*$^{-/-}$ male mice were mated with *p53*$^{+/+}$ female mice. The decrease was very significant compared with wild-type breeding pairs and was most severe when *p53*$^{-/-}$ females were mated with *p53*$^{-/-}$ males, with the embryo genotype of *p53*$^{-/-}$ (Table 1). A similar, although less severe, phenotype was observed with the 129SV^{sl} strain (Table 1), indicating the impact of genetic background. To minimize the effect of genetic background, crosses were performed by mating *p53*$^{-/-}$ mice from the two different genetic backgrounds (C57BL/6J male and 129SV^{sl} female mice). A clear reduction in pregnancy rate and litter size was observed (Table 1), indicating that the defect in reproduction was due to the loss of p53, which could not be compensated for by cross-breeding two different *p53*$^{-/-}$ inbred strains. Furthermore, the extent of decrease was very close to that in the 129SV^{sl} strain (the female breeder) indicating that the severity of the phenotype depended on the genetic background of female mice. These data clearly demonstrate that p53 deficiency in female mice and fetuses results in reduced maternal reproduction, and strongly suggest a crucial function for p53 in maternal reproduction.

As a transcription factor, p53 mainly functions through the transcriptional regulation of its target genes by binding to degenerate DNA-responsive elements (Fig. 1a)[5]. To explore whether p53 regulates maternal reproduction through transcriptional regulation of its target genes, p53 algorithms[6] were employed to search for potential target genes for p53 involved in maternal reproduction. The gene encoding leukaemia inhibitory factor (LIF), a secreted cytokine that is critical for blastocyst implantation[2,7], was identified as a potential p53 target gene with a putative p53-binding element in intron 1 of both mouse and human *LIF* genes (Fig. 1a). Chromatin immunoprecipitation (ChIP) assays were performed in tet-off *p53*-inducible human H1299-WTp53 cells, and in mouse Val5 fibroblasts containing a temperature-sensitive mutant *p53* plasmid. Immunoprecipitation of the chromatin fragments corresponding to the binding elements in both human and mouse *LIF* genes with anti-p53 antibody was observed only when the functional p53 protein was expressed (Fig. 1b). Similar results were obtained in murine embryonic fibroblast (MEF) cells with or without wild-type p53 (Supplementary Fig. 1). These data demonstrate an interaction of p53 with these putative p53-binding elements *in vivo*. Further, pGL2 luciferase reporter plasmids containing these binding elements exhibited a p53-dependent transcriptional activity in *p53*-null human H1299 cells (Fig. 1c), Saos2 and HCT116 *p53*$^{-/-}$ cells (data not shown).

Table 1 | Reproductive data from C57BL/6J and 129SV^{sl} mice with different *p53* genotypes

Strain		Genotype of *p53*		Litter size	Pregnancy rate (%)	Number of breeding pairs
Male	Female	Male	Female			
C57BL/6J	C57BL/6J	+/+	+/+	6.71 ± 0.43	100	14
		−/−	+/+	6.41 ± 0.61	100	11
		+/−	+/−	4.79 ± 0.61	79	19
		−/−	+/−	4.93 ± 0.21	89	89
		+/+	−/−	2.18 ± 0.83*	36†	11
		+/−	−/−	1.42 ± 0.55*	33†	12
		−/−	−/−	0.69 ± 0.16*	27†	69
129SV^{sl}	129SV^{sl}	+/+	+/+	5.41 ± 0.35	95	22
		−/−	+/+	5.58 ± 0.61	100	6
		+/−	+/−	5.21 ± 0.64	100	8
		−/−	+/−	4.8 ± 0.18	84	32
		−/−	−/−	2.35 ± 0.24‡	58§	90
C57BL/6J	129SV^{sl}	+/+	+/+	5.64 ± 0.58	100	11
		−/−	−/−	2.62 ± 0.84	63	8

Eight-week-old mice with different *p53* genotypes were set up for mating as indicated and were housed together for at least 6 weeks. Pregnancy rate was calculated as the ratio of the number of females with confirmed pregnancy to the number of females who served as breeders. Where shown, errors are s.e.m.
* $P < 0.001$ versus wild type, *t*-test.
† $P < 0.001$ versus wild type, χ^2 test.
‡ $P < 0.01$ versus wild type, *t*-test.
§ $P < 0.01$ versus wild type, χ^2 test.

[1]Cancer Institute of New Jersey, University of Medicine and Dentistry of New Jersey, New Brunswick, New Jersey 08903, USA. [2]The Simons Center for Systems Biology, The Institute for Advanced Study, Princeton, New Jersey 08540, USA.
*These authors contributed equally to this work.

©2007 Nature Publishing Group

Regulation of *LIF* expression by p53 was explored in Val5 and Vm10 cells, both containing a temperature-sensitive mutant *p53* plasmid. Activation of p53 at 32 °C increased LIF levels significantly in both cells (up to sevenfold at the messenger RNA level and 14-fold at the protein level) and culture supernatant (up to 16-fold at the protein level) (Fig. 1d). This induction was p53 dependent; there was no increase in 10(1) cells, their *p53*-null parental cells. This p53-dependent induction was also observed in human H1299-WTp53 cells (data not shown). Irradiation with γ-rays resulted in a large induction of *Lif* transcription in spleen (more than 70-fold) and thymus (more than 12-fold), two radiosensitive tissues, in wild-type mice but not in $p53^{-/-}$ mice (Fig. 1e). A clear p53-dependent induction of *Lif* was also observed in uterine tissues under conditions of both whole-body γ-irradiation and *in vitro* irradiation of cultured uterine tissues (Fig. 1e). This induction showed tissue specificity; it was also observed in liver, skin and small intestine but not in kidney, heart or muscle tissues (data not shown). For the sake of comparison, the induction of a panel of known p53-target genes by γ-irradiation in different tissues is listed in Supplementary Fig. 2. Furthermore, endogenous p53, under no apparent stress conditions, had a significant impact on the basal transcription levels of *Lif*. In two pairs of isogenic cell lines (MEF $p53^{+/+}$ and $p53^{-/-}$, and HCT116 $p53^{+/+}$ and $p53^{-/-}$), the basal *Lif* mRNA levels were significantly (twofold to threefold) higher in $p53^{+/+}$ cells than in $p53^{-/-}$ cells (Fig. 1f). These data show that *LIF* is a previously unidentified target of p53, which regulates both basal and inducible transcription of *LIF*.

LIF has a crucial function in blastocyst implantation[2,7,8]. Transiently increased uterine *LIF* expression coincides with onset of implantation[9]. Similarly to our observation in $p53^{-/-}$ mice, $Lif^{-/-}$ mice have a defect in maternal reproduction as a result of the failure of blastocyst implantation[2]. To investigate the impact of loss of p53 on LIF levels and function during pregnancy, uterine *Lif* expression levels were measured in pregnant mice with different p53 statuses. Simultaneously with increased oestrogen levels, the *Lif* expression levels increased significantly (fivefold to eightfold) in the $p53^{+/+}$ uterus at day 4 (Fig. 2a), the exact onset time for implantation, although no significant increase in the p53 levels (activation) was observed (Supplementary Fig. 3). This transient induction of *Lif* showed tissue specificity; it was not observed in spleen or thymus, where the highest induction by p53 activation after γ-irradiation was observed. The uterine *Lif* expression levels were significantly lower in $p53^{-/-}$ mice than in $p53^{+/+}$ mice, both non-pregnant (day 0) and at day 4 of pregnancy (Fig. 2b). *Lif* is an oestrogen-responsive gene[7] and it has been suggested that oestrogen is involved in the induction of *Lif* expression at day 4 of pregnancy[7]. The impact of the p53 status on *LIF* regulation by oestrogen was examined in a pair of isogenic oestrogen-responsive cell lines: MCF7 ($p53^{+/+}$) and MCF7-p53siRNA (stably expressing a *p53* short interfering RNA (siRNA)). The basal transcription levels of *LIF* in MCF7-p53siRNA were one-third of those in MCF7 cells (Fig. 2c). The *LIF* expression levels increased significantly in both cell lines after treatment with oestrogen. However, significantly lower *LIF* levels were still observed in MCF7-p53siRNA cells (more than threefold to fourfold lower). As a control, there were no significant differences in the levels of *pS2*, an oestrogen-responsive gene not regulated by p53, between these two cell lines, either before or after treatment with oestrogen (data not shown). These results are similar to the observation (Fig. 2b) of different uterine *Lif* levels in day 4 pregnant mice of various p53 statuses, indicating that p53 regulates *LIF* expression in coordination with oestrogen during implantation, which may involve an interaction between p53, oestrogen and oestrogen receptor. Along with the decreased uterine expression of *Lif*, we observed impaired implantation in the $p53^{-/-}$ mice. Examining early embryo implantation sites at day 5 of pregnancy by Chicago blue dye staining demonstrated a significantly smaller number of implantation sites in $p53^{-/-}$ mice than in $p53^{+/+}$ mice (Fig. 2d, e). A significant number of unimplanted blastocysts were recovered from $p53^{-/-}$ mice at day 5 of pregnancy by uterine flushing, whereas

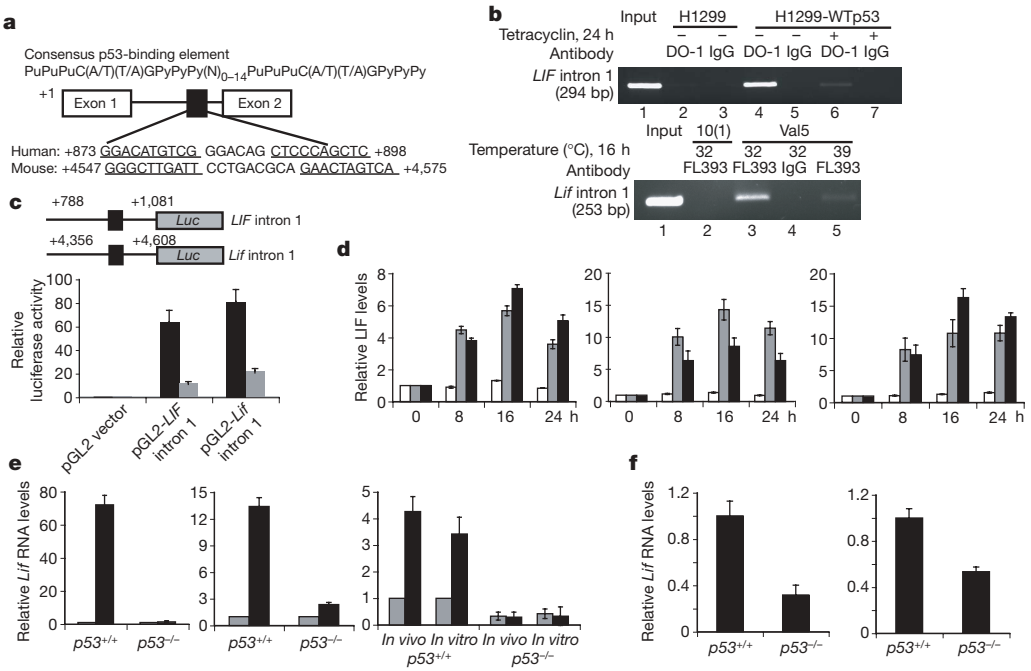


Figure 1 | p53 regulates the expression of *LIF* genes. **a**, Putative p53-consensus binding elements in *LIF* genes. Pu, purine; Py, pyrimidine. **b**, p53 binds to the p53-consensus binding elements in the *LIF* genes (human, *LIF*; mouse, *Lif*) as determined by ChIP assay. bp, base pairs. **c**, p53 transactivates the p53-consensus binding elements in the *LIF* genes as determined by luciferase (*Luc*) reporter assay. Black bars, WTp53; grey bars, mutant p53. **d**, p53 activation increases LIF expression and secretion. Left, RNA; centre, protein (cell lysate); right, protein (culture supernatant). White bars, 10(1) cells; grey bars, Val5 cells; black bars, Vm10 cells. **e**, p53 activation increases *Lif* expression in mice. C57BL/6J mice or uterine tissues from C57BL/6J mice were γ-irradiated (5 Gy) and tissues were harvested 6 h after irradiation ($n = 3$). Left, spleen; centre, thymus; right, uterus. Grey bars, control; black bars, irradiated. **f**, p53 regulates the basal transcription levels of *Lif*. Left, MEF cells; right, HCT116 cells. Results are means and s.e.m.

©2007 Nature Publishing Group

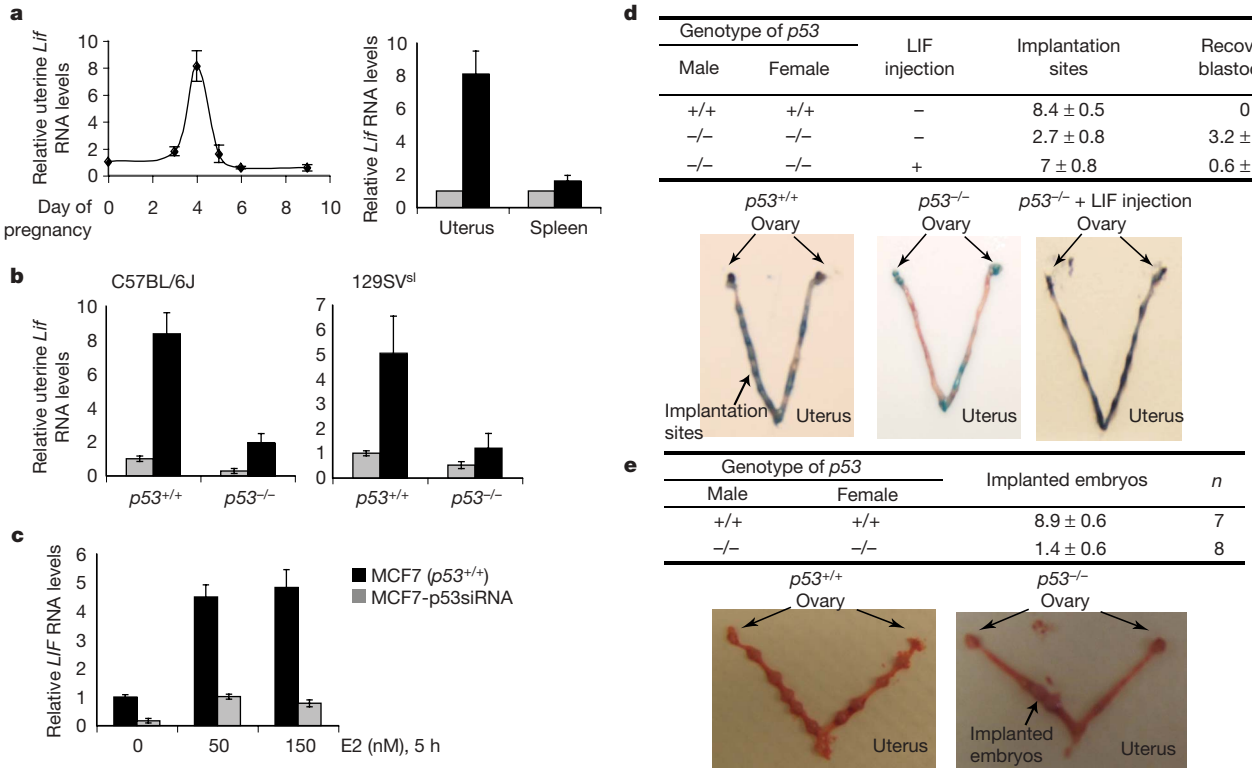


Figure 2 | p53 regulates *LIF* expression in mouse uterus during implantation. **a**, Transiently increased uterine *Lif* expression coincides with the onset of implantation in $p53^{+/+}$ C57BL/6J mice ($n = 4$). Grey bars, day 0; black bars, day 4. **b**, Lower uterine *Lif* expression in $p53^{-/-}$ mice, especially at the implantation stage ($n = 4$). Grey bars, day 0; black bars, day 4. **c**, The impact of p53 on regulation of *LIF* by oestrogen. **d**, **e**, Impaired implantation in $p53^{-/-}$ C57BL/6J mice and its restoration by injection of LIF. **d**, Early implantation examined at day 5 of pregnancy. LIF was injected at day 4 of pregnancy. **e**, Uteri were collected at day 7 of pregnancy and the number of implanted embryos was counted. Results are means and s.e.m.

none was recovered from $p53^{+/+}$ mice. A significantly smaller number of implanted embryos was consistently observed in $p53^{-/-}$ mice by day 7 of pregnancy (Fig. 2d, e). $p53^{-/-}$ mice have normal uterine morphology and ovarian oestrogen levels, and seem to have normal ovulation and fertilization functions (Supplementary Fig. 4), which were unlikely to be causes for the observed low uterine LIF levels and the small number of implanted embryos in $p53^{-/-}$ mice. Taken together, these data demonstrate impaired implantation in $p53^{-/-}$ mice.

Implantation can be restored in $Lif^{-/-}$ mice by the administration of exogenous LIF at day 4 after pregnancy[7]. To investigate whether low uterine LIF levels contributed to the reduced reproduction in $p53^{-/-}$ female mice, LIF was administered (by intraperitoneal injection) to the pregnant $p53^{-/-}$ mice at day 4. LIF injection did indeed improve maternal reproduction significantly in C57BL/6J $p53^{-/-}$ female mice, with 100% fertility and a significantly increased litter size (Table 2); 70% of the mice born were alive and normal. The rest of the mice had developmental abnormalities, mainly defects in the neural tube, which were predominantly associated with females, which is consistent with previous reports[10]. For the 129SV^{sl} strain, only those $p53^{-/-}$ female mice that failed to become pregnant after mating with $p53^{-/-}$ males for at least 1 month were employed for LIF injection. Injection of LIF significantly increased the pregnancy rate (from 0% to 100%) and litter size (Table 2). This effect was p53 dependent; LIF injection did not change the pregnancy rate or litter size in wild-type mice (Table 2). Injection of LIF seemed to act by improving implantation in $p53^{-/-}$ mice; it significantly increased the number of implantation sites and decreased the number of unimplanted blastocysts recovered at day 5 of pregnancy in $p53^{-/-}$ females (Fig. 2d, e). These results demonstrate that *p53* loss decreases the levels and function of uterine LIF, and administration of LIF to pregnant $p53^{-/-}$ mice restores maternal reproduction by improving implantation.

Table 2 | LIF injection increases pregnancy rate and litter size in $p53^{-/-}$ female mice

Strain and genotype	LIF injection	Litter size	Pregnancy rate (%)	Number of breeding pairs
C57BL/6J, $p53^{-/-} \times p53^{-/-}$	–	0.69 ± 0.16	27	69
	+	5.2 ± 0.57*	100†	10
129SV^{sl}, $p53^{-/-} \times p53^{-/-}$	–	2.35 ± 0.24	58	90
	+	3.8 ± 0.32*	100†	5
C57BL/6J, $p53^{+/+} \times p53^{+/+}$	–	6.71 ± 0.43	100	14
	+	7.4 ± 0.81	100	5

Recombinant LIF (5 μg) was administered to female mice twice, with 5 h between, by intraperitoneal injection at day 4 of pregnancy (the day on which a plug was observed was designated as day 1). For the 129SV^{sl} strain, only those $p53^{-/-}$ female mice that failed to get pregnant after mating with $p53^{-/-}$ male mice for at least 4 weeks were used for LIF injection. Where shown, errors are s.e.m.
* $P < 0.001$ versus untreated mice, *t*-test.
† $P < 0.001$ versus untreated mice, χ^2 test.

Implantation failure is the most frequent cause of lack of human pregnancy after *in vitro* fertilization and embryo transfer[11]. Sufficient uterine LIF protein is an essential condition for implantation, and low LIF levels have been reported in infertile women[12]. Considering the strict regulation of *LIF* by p53, the modulation of p53 function by single nucleotide polymorphisms (SNPs) in the *p53* pathway[4,13] may affect implantation. A recent study reported that the *p53* codon 72 polymorphism (an Arg to Pro change) is associated with recurrent implantation failure in humans, but its mechanism is unclear[11]. Because the proline allele of codon 72 seems to encode a p53 protein with a weaker activity than that from the arginine allele[4], our results suggest a possible mechanism that mediates the impact of this SNP. These studies may also help to explain the positive selection during human evolution of some alleles in the *MDM2* (SNP309)[3] and *p53* genes[4].

©2007 Nature Publishing Group

This study clearly demonstrates a new function for p53 in the regulation of maternal reproduction, in which *LIF*, a previously unidentified p53 target gene, functions as the downstream mediator of this p53 effect in mice (Supplementary Fig. 5).

METHODS SUMMARY

Wild-type and $p53^{-/-}$ mice were used for breeding assays, and reproduction parameters were measured. Early implantation sites were examined by the Chicago blue dye-staining method at day 5 of pregnancy, and the number of implanted embryos was also counted at day 7 of pregnancy. ChIP assays were performed with an Upstate ChIP assay kit. For reporter assays, putative p53-binding elements were cloned into the promoter region of a pGL2 luciferase reporter plasmid (Promega). The reporter plasmids were transfected into *p53*-null H1299, Soas2 and HCT116 $p53^{-/-}$ cells along with either pRC-wild-type-p53 or pRC-mutant-p53 p53 expression plasmid. The luciferase activities were determined 24 h after transfection. The mRNA levels of different genes were measured by Taqman quantitative real-time PCR. The LIF protein levels were measured by ELISA assay (R & D).

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

Received 24 July; accepted 10 October 2007.

1. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
2. Stewart, C. L. *et al.* Blastocyst implantation depends on maternal expression of leukaemia inhibitory factor. *Nature* **359**, 76–79 (1992).
3. Atwal, G. S. *et al.* Haplotype structure and selection of the *MDM2* oncogene in humans. *Proc. Natl Acad. Sci. USA* **104**, 4524–4529 (2007).
4. Murphy, M. E. Polymorphic variants in the p53 pathway. *Cell Death Differ.* **13**, 916–920 (2006).
5. el-Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W. & Vogelstein, B. Definition of a consensus binding site for p53. *Nature Genet.* **1**, 45–49 (1992).
6. Hoh, J. *et al.* The p53MH algorithm and its application in detecting p53-responsive genes. *Proc. Natl Acad. Sci. USA* **99**, 8467–8472 (2002).
7. Chen, J. R. *et al.* Leukemia inhibitory factor can substitute for nidatory estrogen and is essential to inducing a receptive uterus for implantation but is not essential for subsequent embryogenesis. *Endocrinology* **141**, 4365–4372 (2000).
8. Song, H., Lim, H., Das, S. K., Paria, B. C. & Dey, S. K. Dysregulation of EGF family of growth factors and COX-2 in the uterus during the preattachment and attachment reactions of the blastocyst with the luminal epithelium correlates with implantation failure in LIF-deficient mice. *Mol. Endocrinol.* **14**, 1147–1161 (2000).
9. Cullinan, E. B. *et al.* Leukemia inhibitory factor (LIF) and LIF receptor expression in human endometrium suggests a potential autocrine/paracrine function in regulating embryo implantation. *Proc. Natl Acad. Sci. USA* **93**, 3115–3120 (1996).
10. Armstrong, J. F., Kaufman, M. H., Harrison, D. J. & Clarke, A. R. High-frequency developmental abnormalities in p53-deficient mice. *Curr. Biol.* **5**, 931–936 (1995).
11. Kay, C., Jeyendran, R. S. & Coulam, C. B. p53 tumour suppressor gene polymorphism is associated with recurrent implantation failure. *Reprod. Biomed. Online* **13**, 492–496 (2006).
12. Hambartsoumian, E. Endometrial leukemia inhibitory factor (LIF) as a possible cause of unexplained infertility and multiple failures of implantation. *Am. J. Reprod. Immunol.* **39**, 137–143 (1998).
13. Bond, G. L. *et al.* A single nucleotide polymorphism in the *MDM2* promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell* **119**, 591–602 (2004).

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

Acknowledgements We thank S. Christen for critical review of the manuscript, and I. Snyder and the staff of the University of Medicine and Dentistry of New Jersey animal facility for providing animal care. This research was supported by grants from the Breast Cancer Research Foundation and the National Cancer Institute.

Author Contributions W.H., Z.F. and A.J.L. designed the research. W.H., Z.F. and A.K.T. performed the research. W.H., Z.F. and A.J.L. analysed data and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.J.L. (alevine@ias.edu).

©2007 Nature Publishing Group

METHODS

Mice and cell lines. The C57BL/6J and 129SV^{sl} strain mice were purchased from the Jackson Laboratory. Mice deficient in *p53* (C57BL/6J*129SV) were provided by T. Jacks. The *p53*$^{-/-}$ C57BL/6J and 129SV^{sl} strains were produced by backcrossing ten times to C57BL/6J and 129SV^{sl}, respectively. The wild-type and *p53*$^{-/-}$ MEFs were generated as described previously[14]. Mouse fibroblast Val5 and Vm10 cells, which contain a temperature-sensitive mutant *p53* plasmid (Ala 135 to Val) and express mutant *p53* at 39 °C but wild-type *p53* at 32 °C, were established as described previously[15,16]. HCT116 *p53*$^{+/+}$ and *p53*$^{-/-}$ human colon cancer cells were gifts from B. Vogelstein. Saos2 *p53*-null human osteosarcoma cells and MCF7 (*p53* wild-type) human breast cancer cells were obtained from the American Type Culture Collection (ATCC). The MCF7-p53siRNA cell line stably expressing *p53* siRNA, to knock down p53 protein[17], was a gift from A. Gartel. The human H1299-WTp53 cell line, which contains a *p53* expression plasmid under the control of tetracycline and expresses the wild-type p53 protein on tetracycline withdrawal, was established from the *p53*-null H1299 cell line as described previously[18].

Mouse breeding assay and examination of reproduction parameters. For breeding assays, mating pairs (8-week-old mice) were placed in cages for at least 6 weeks, and female mice were checked every morning for copulation plugs and signs of parturition. The day on which the plug was observed was considered to be day 1 of pregnancy. Pregnancy rate was calculated as the ratio of the number of females with confirmed pregnancy to the number of female mice housed with male mice. The oestrous cycle was evaluated by cytological analysis of daily vaginal smears for 2 weeks. The serum oestrogen levels were measured with an oestradiol enzyme immunoassay kit (Cayman). The number of corpora lutea on each ovary as an indicator of recent ovulation was recorded. Uterine horns were flushed with M2 medium (Sigma) to recover blastocysts[19]. Early implantation sites were examined by the Chicago blue dye-staining method[19] to reveal blue bands at day 5 of pregnancy. The number of implanted embryos was also counted in uteri collected from mice at day 7 of pregnancy.

Injection of LIF. Recombinant mouse LIF (5 μg; Millipore) was administered to female mice twice, with 5 h between, by intraperitoneal injection at day 4 of pregnancy.

Irradiation with γ-rays. Six-week-old C57Bl/6J wild-type and *p53*$^{-/-}$ mice were subjected to 5 Gy of total body irradiation with a ^{137}Cs γ-ray source. Mice were killed 6 h after irradiation, and different tissues were harvested for further experiments. For *in vitro* irradiation, uteri collected from *p53*$^{+/+}$ or *p53*$^{-/-}$ C57Bl/6J mice were irradiated (5 Gy) and cultured in DMEM medium supplemented with 10% fetal bovine serum for 6 h before being harvested. At least three mice were used in each group.

ChIP assay. ChIP assays were performed with an Upstate ChIP assay kit in accordance with the instructions of the manufacturer. Human H1299-WTp53 and *p53*-null H1299 cells were cultured in the absence of tetracycline for 24 h, followed by ChIP assay with DO-1 anti-p53 antibody. Mouse Val5 and *p53*-null 10(1) cells were cultured at 32 °C for 16 h followed by ChIP assay with FL393 anti-p53 antibody. The primer sets were designed to encompass the potential p53-binding elements in intron 1 of the human and mouse *LIF* genes. The sequences for the *LIF* genes are as follows: human, 5′-ACCCCGGCCC-CCACCATCTTC-3′ and 5′-ACTCATCCGCTGTCCCTTGTCC-3′; mouse, 5′-TGACCTGGGGGAGAGCAACCTAAC-3′ and 5′-AGCCAACAGCCCCA-GCATCAC-3′.

Construction of reporter plasmids and luciferase activity assay. The TOPO II vector (Invitrogen) was used to clone PCR fragments containing the putative p53-binding elements in intron 1 of human and mouse *LIF* genes by using PCR primers for ChIP assays. The sequence-confirmed clones were subcloned into pGL2 luciferase reporter plasmid (Promega). The pGL2 reporter plasmids containing one copy of each putative p53-binding element were transfected into *p53*-null H1299 cells, Saos2 and HCT116 *p53*$^{-/-}$ cells by using LipofectAMINE 2000 (Invitrogen) along with 1 μg of pRC-wtp53 (wild-type human *p53* expression plasmid) or pRC-273H (mutant human *p53* expression plasmid containing a substitution at R273H) and 0.5 ng of pRL-SV40 plasmid expressing *Renilla* luciferase as an internal control to normalize transfection efficiency. The luciferase activity was measured 24 h after transfection and normalized with the internal standard.

Treatment with oestrogen. For treatment with oestrogen, cells were cultured in Phenol red-free culture medium supplemented with 10% charcoal-stripped fetal bovine serum for 3 days before being treated with various concentrations of 17-β-oestradiol (Sigma) for various durations.

Quantitative real-time PCR. Total RNA was prepared from cells or mouse tissues with the RNeasy kit (Qiagen) and treated with DNase I to remove residual genomic DNA. The complementary DNA was generated with random primers using a TaqMan reverse transcription kit (Applied Biosystems). Real-time PCR was performed in triplicate with TaqMan PCR mixture (Applied Biosystems) in the 7000 ABI sequence detection system. All human and mouse primers were purchased from Applied Biosystems. The expression of genes was normalized to the housekeeping gene encoding β-actin.

Measurement of LIF protein levels. Val5 and Vm10 cells were cultured at 32 °C for various durations before cells and culture supernatants were collected. Cells were lysed in cell extraction buffer for enzyme-linked immunosorbent assay (ELISA; Biosource). Culture supernatants were concentrated with an Amicon Ultra-4 centrifugal filter device (Millipore) after a brief centrifugation to remove any cell debris. The LIF protein levels were measured with a mouse LIF ELISA kit (R & D) in accordance with the manufacturer's instructions.

14. Harvey, D. & Levine, A. J. p53 alteration is a common event in the spontaneous immortalization of primary BALB/c murine embryo fibroblasts. *Genes Dev.* **5,** 2375–2385 (1991).
15. Pochampally, R., Fodera, B., Chen, L., Lu, W. & Chen, J. Activation of an MDM2-specific caspase by p53 in the absence of apoptosis. *J. Biol. Chem.* **274,** 15271–15277 (1999).
16. Wu, X. & Levine, A. J. p53 and E2F-1 cooperate to mediate apoptosis. *Proc. Natl Acad. Sci. USA* **91,** 3602–3606 (1994).
17. Radhakrishnan, S. K., Gierut, J. & Gartel, A. L. Multiple alternate p21 transcripts are regulated by p53 in human cells. *Oncogene* **25,** 1812–1815 (2006).
18. Chen, X., Ko, L. J., Jayaraman, L. & Prives, C. p53 levels, functional domains, and DNA damage determine the extent of the apoptotic response of tumor cells. *Genes Dev.* **10,** 2438–2451 (1996).
19. Nagy, A., Gertsenstein, M., Vintersten, K. & Behringer, R. *Manipulating the Mouse Embryo, a Laboratory Manual* 3rd edn (Cold Spring Harbor Press, New York, 2003).

©2007 Nature Publishing Group

LETTERS

Intracellular bacterial growth is controlled by a kinase network around PKB/AKT1

Coenraad Kuijl[1], Nigel D. L. Savage[3], Marije Marsman[1], Adriaan W. Tuin[4], Lennert Janssen[1], David A. Egan[2], Mirjam Ketema[1], Rian van den Nieuwendijk[4], Susan J. F. van den Eeden[3], Annemieke Geluk[3], Alex Poot[4], Gijs van der Marel[4], Roderick L. Beijersbergen[2], Hermen Overkleeft[4], Tom H. M. Ottenhoff[3] & Jacques Neefjes[1]

With the emergence of multidrug resistant (MDR) bacteria, it is imperative to develop new intervention strategies. Current antibiotics typically target pathogen rather than host-specific biochemical pathways[1]. Here we have developed kinase inhibitors that prevent intracellular growth of unrelated pathogens such as *Salmonella typhimurium* and *Mycobacterium tuberculosis*. An RNA interference screen of the human kinome using automated microscopy revealed several host kinases capable of inhibiting intracellular growth of *S. typhimurium*. The kinases identified clustered in one network around AKT1 (also known as PKB). Inhibitors of AKT1 prevent intracellular growth of various bacteria including MDR-*M. tuberculosis*. AKT1 is activated by the *S. typhimurium* effector SopB, which promotes intracellular survival by controlling actin dynamics through PAK4, and phagosome–lysosome fusion through the AS160 (also known as TBC1D4)–RAB14 pathway. AKT1 inhibitors counteract the bacterial manipulation of host signalling processes, thus controlling intracellular growth of bacteria. By using a reciprocal chemical genetics approach, we identified kinase inhibitors with antibiotic properties and their host targets, and we determined host signalling networks that are activated by intracellular bacteria for survival.

A major bottleneck in the development of new antibiotics is our lack of knowledge of the intricate relationship between host and bacterial proteins, which usually results in intracellular survival of the pathogen. Intracellular bacteria persist in phagosomes where they avoid degradation in lysosomes[2,3]. *S. typhimurium* modulates microtubular kinesin motors on phagosomes[4] and manipulates dynein-motor-driven transport of phagosomes to lysosomes to rescue *Salmonella* growth[5]. *Salmonella* activates host kinases like protein kinase A (PKA)[6], which might inhibit dynein motors[7] to promote *Salmonella* replication. We therefore set out to identify host kinases as potential intervention targets in infectious diseases.

We tested selected kinase inhibitors for effects on intracellular *Salmonella typhimurium* growth in the human breast cancer cell line MCF7 (Fig. 1a). MCF7 cells were infected with Lux-expressing *S. typhimurium* SL12023 and cultured with inhibitors for 18 h. The effect on intracellular growth was monitored by luminescence (Fig. 1a), colony-forming units (c.f.u.) and fluorescence-activated cell sorting (FACS) assays (not shown). Only H-89 prevented intracellular growth of *S. typhimurium* without affecting proliferation of MCF7 cells. This was confirmed in various human and mouse cell lines (not shown), including primary human macrophages (Fig. 2b). Because two other inhibitors of PKA failed to affect *S. typhimurium* growth (Supplementary Fig. 1), H-89 probably inhibited kinases other than PKA. When MCF7 cells were infected with green fluorescent protein (GFP)-expressing *S. typhimurium* SL1344 for 18 h, large intracellular *S. typhimurium* structures in CD63-positive late endosomes were observed in control but not H-89-treated cells (Fig. 1b). H-89 promoted fusion of GFP-*S. typhimurium*-containing phagosomes with lysosomes (not shown). H-89 could have affected an *S. typhimurium* kinase, however bacterial growth was unaltered by H-89 in normal broth or under conditions mimicking the phagosomal environment (Fig. 1c).

To define host kinases involved in intracellular growth of *S. typhimurium* and other intracellular pathogens, H-89 analogues were synthesized for chemical profiling (Fig. 2a; synthesis, Supplementary Fig. 2). H-89 competes with ATP for access to kinases[8]. Enzyme specificity can be obtained by varying the linker in length and form and by modifying the bulky moiety. The effects of H-89 and nine variants were tested on intracellular *S. typhimurium* growth in primary human macrophages (Fig. 2b). *S. typhimurium* proliferation was inhibited by H-89, ETB067 and ETB275.

The activity of H-89-related variants was then examined on two mycobacterial strains, *M. smegmatis* (Supplementary Fig. 3) and *M. tuberculosis*, following human macrophage infection (Fig. 2c). Similar chemical profiles to *S. typhimurium* infection were obtained, suggesting similar host kinases control intracellular growth of different pathogens.

To identify these, a Dharmacon short interfering (si)RNA library containing 779 siRNA smartpools targeting the complete human kinome (658 candidates) and 121 kinase-associated or -regulatory proteins was tested in a high-throughput screen for inhibition of intracellular *S. typhimurium* growth using automated microscopy. Because H-89 and most analogues did not affect macrophage and MCF7 viability (Supplementary Fig. 4), silencing relevant kinase(s) by siRNA should neither affect cell viability. MCF7 cells were transfected with SMARTpool siRNA such that one defined kinase was targeted per well. Two days post transfection, cells were infected with GFP-expressing *S.typhimurium* SL1344, intracellular growth monitored 18 h post infection by automated microscopy (Fig. 3a) and 26,000 images analysed (Supplementary Figs 5 and 6). By quantifying GFP within cells (Supplementary Fig. 5), intracellular growth of *S. typhimurium* was determined, and plotted as a heat map (Supplementary Fig. 8c; for details on individual kinase targets and statistics see Methods). Data were visually verified and confirmed with a short hairpin (sh)RNA library specific to human kinases[9] (Supplementary Fig. 7). Downregulation of 11 kinases and 3 kinase-associated proteins reduced intracellular *S. typhimurium* growth (Fig. 3b). Ten of these kinases were validated with multiple siRNAs for the same kinase

[1]Division of Tumor Biology, and [2]Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands. [3]Department of Immunohematology and Blood Transfusion and Department of Infectious Diseases, Leiden Medical University Centre, Albinusdreef 2, 2333 ZA Leiden, The Netherlands. [4]Department of Bioorganic Chemistry, Leiden Institute of Chemistry, 2300 RA Leiden, The Netherlands.

©2007 Nature Publishing Group

(Fig. 3b, and Supplementary Fig. 8a, b). These targets were not identified in RNAi screens for host factors after intracellular infection of *Drosophila melanogaster* macrophage-like cells with *M. tuberculosis* or *M. fortuitum*[10,11]. AKT1 is involved in SimianVirus40 infection, but other kinases such as PAK4 were not identified in these RNAi studies[12]. Of the three AKT enzymes, our screen identified only AKT1, which is selectively targeted by H-89 (not shown). No apoptosis was observed after downregulation of AKT1 by siRNA or inactivation by H-89 (Fig. 3a), because this requires inactivation of AKT1 and AKT2 (R.L.B., unpublished observation). Autophagy was not observed (Supplementary Fig. 9).

Pathway analyses of data revealed that kinases controlling intracellular growth of *S. typhimurium* clustered around a single AKT1 network (Fig. 3c), suggesting this network is critical for intracellular *S. typhimurium* growth. To determine which kinases were targets of H-89 and its homologues, PKA and four other similar kinases (Supplementary Fig. 10) were tested for inhibition by H-89 and six variants (Fig. 4a). Only PKA and AKT1 activities were inhibited by H-89 and ETB067, in accordance with the chemical profile of intracellular bacterial growth inhibition (Fig. 2). Because PKA was not identified in the siRNA screen and other PKA inhibitors did not affect intracellular *S. typhimurium* growth (Fig. 1a; and Supplementary Fig. 1), AKT1 might be the target for H-89 and ETB067. An unrelated

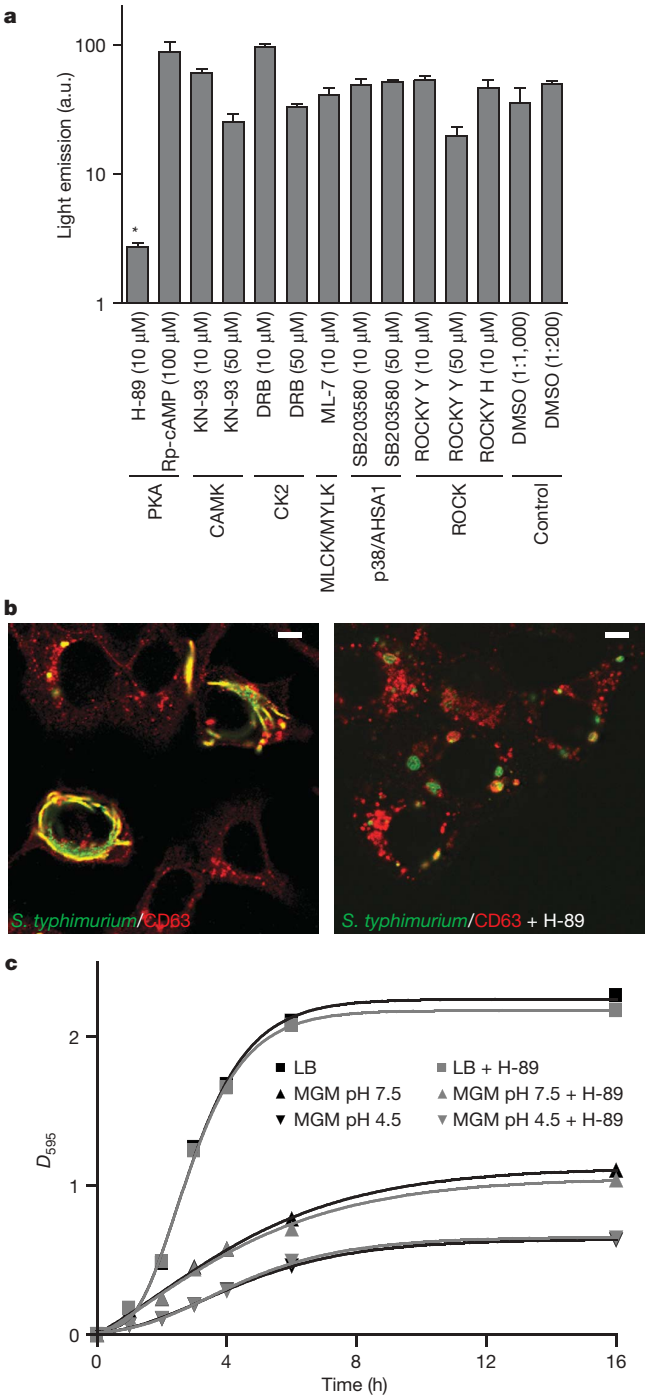


Figure 1 | The protein-kinase-A inhibitor H-89 inhibits host kinases that control intracellular *Salmonella typhimurium* growth. **a**, MCF7 cells infected with Lux-*S. typhimurium* were cultured for 18 h with the various chemicals. The intracellular *S. typhimurium* was quantified by luminometry. Inhibitors and kinase targets are indicated. Experiments were in triplicate (±s.d.) (*$P = 0.017$). CK2 is also known as CSNK2A1. **b**, MCF7 cells were infected with GFP-*S.typhimurium* SL1344 and cultured with H-89 for 18 h before staining for the late endosomal marker CD63 ($n > 100$). Scale bar, 10 μm. **c**, *S. typhimurium* growth in LB or a medium mimicking endosomal conditions (MGM) in the presence or absence of 10 μM H-89 at pH 7.5 or pH 4.5 was measured at a wavelength of 595 nm.

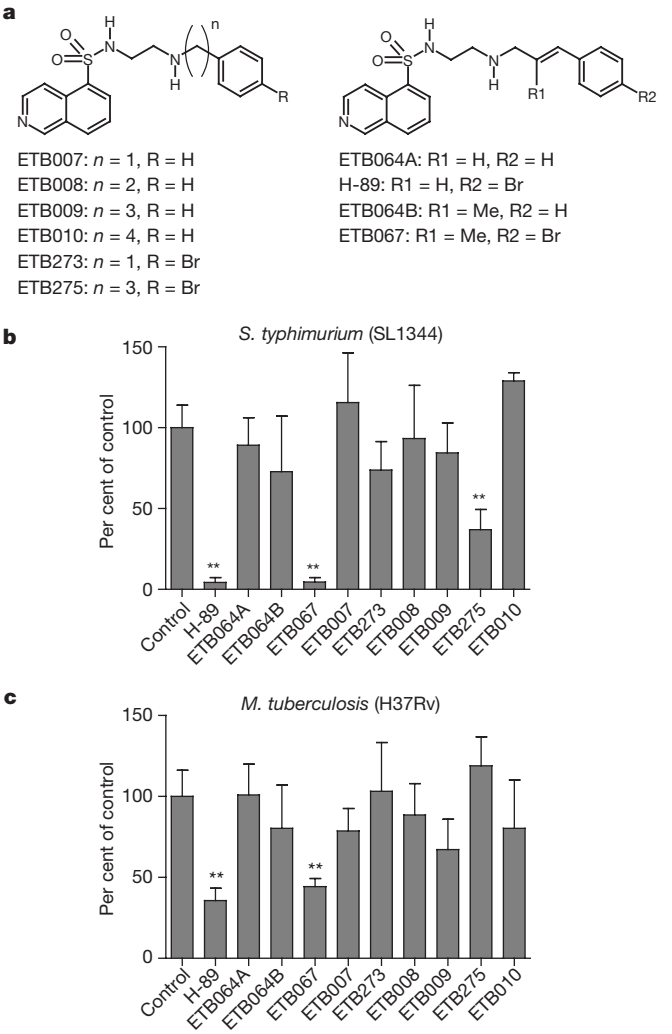


Figure 2 | Chemical profiling for antibiotic activity of kinase inhibitors. **a**, The building block H-89 was selectively modified. For synthesis see Supplementary Fig. 2. **b**, Chemical profiling for *S. typhimurium* infection. Primary human macrophages infected with *S. typhimurium* SL1344 were cultured in the presence of 10 μM of the compounds for 18 h. Intracellular growth was determined in c.f.u. assays. Shown is the mean of quadruplicate c.f.u. counts ± s.e.m. (**$P < 0.001$) **c**, Chemical profiling for *M. tuberculosis* infection. Primary human macrophages infected with *M. tuberculosis* were cultured as in **b** for 6 days with daily renewal of medium, and intracellular growth determined. Shown is the mean of quadruplicate c.f.u. counts ± s.e.m. (**$P < 0.001$)

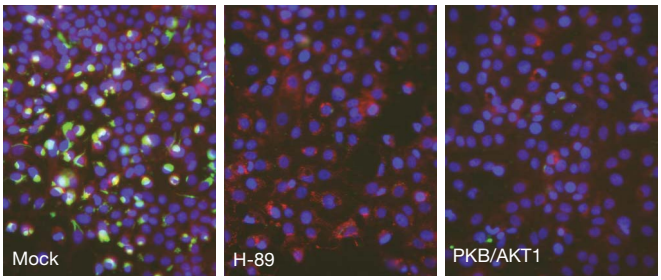
©2007 Nature Publishing Group

PKB inhibitor AKTi1/2 (which does not inhibit PKA; Fig. 4a) also inhibited intracellular propagation of *S. typhimurium* and *M. smegmatis* in human macrophages and MCF7 cells (Supplementary Fig. 11). AKTi1/2 (which inhibits AKT1 and AKT2) induced apoptosis in long-term cultures, unlike H-89 and analogues (Supplementary Fig. 4), and therefore the effect on slowly replicating pathogen *M. tuberculosis* could not be determined.

Host kinase AKT1 represents a new target for combating MDR-pathogens like MDR-*M. tuberculosis*, which should be sensitive to ETB067 and H-89. Indeed, human macrophages infected with MDR-*M. tuberculosis* (Beijing/W family) and cultured with various H-89 analogues showed similar chemical profiles (Fig. 4b) to *S. typhimurium*, *M. smegmatis* and *M. tuberculosis* (Fig. 2b, c, and Supplementary Fig. 3)

The various compounds may determine sites in H-89 to increase the specificity for AKT1. ATP-binding sites of PKA and AKT1 are highly similar and H-89 fills this in a 'bended' conformation (Supplementary Fig. 12)[13]. The H-89 isoquinoline group mimics the ATP adenine group and the bromo-phenyl group fills a pocket that is identical in PKA and AKT1[13]. Changing the halogen probably does not improve selectivity of the compounds, but, at best, affinity. Manipulating the linker length may position the isoquinoline-group and the bulky moiety incorrectly for an optimal fit into the ATP-binding site thus reducing affinity, as observed (Figs 2, 4a). The reduced bond in the linker bends the inhibitor in the active site of PKA and AKT1 (Supplementary Fig. 12) and improves affinity but

a



b

Gene	Method	Validated	Confirmation
ABL1	siRNA	+	qPCR
CAMK2B	siRNA	+	
FASTK	siRNA	+	
LYN	siRNA	–	
MAP3K14	siRNA	+	
PAK4	shRNAi	+	qPCR
PCTK1	siRNA	+	
PIK3C2G	siRNA	+	qPCR
PIK3CA	siRNA	+	qPCR
PKB/AKT	siRNA	+	WB
PRKD2	siRNA	+	qPCR
CALM3	siRNA		
CINP	siRNA		
MAPK8IP2	siRNA		

c

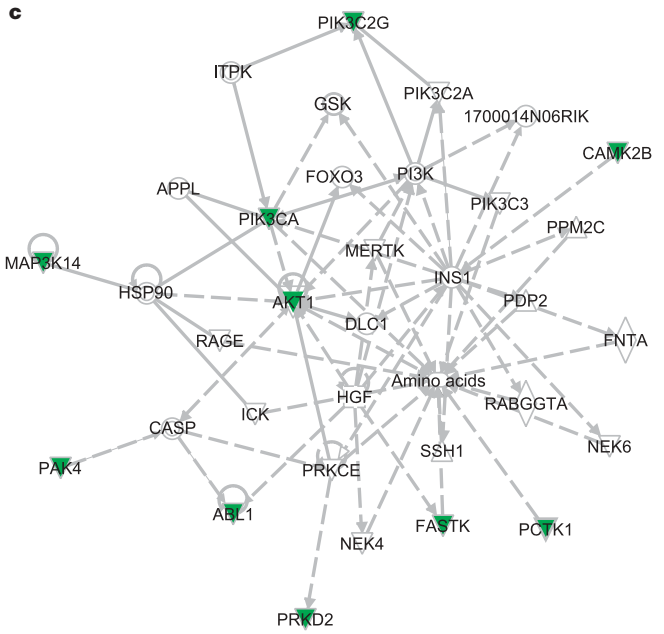
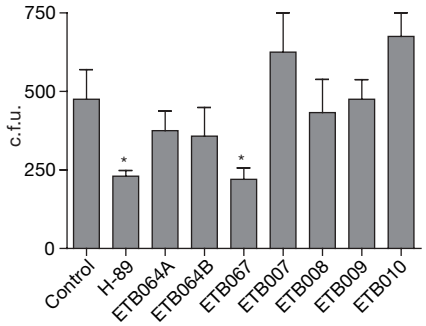


Figure 3 | Identification of host kinases controlling intracellular growth of *S. typhimurium*. **a**, Automated-microscopy images from mock- and H-89-treated, as well as *AKT1*-siRNA-transfected MCF7 cells (magnification, ×40). Blue, nuclei; red, cell boundaries; green, GFP-*S. typhimurium*. Intracellular growth (amount of GFP-*S. typhimurium* per infected cell) was calculated from these images (Supplementary Figs 5 and 6). For details on individual siRNA targets and statistics, see Supplementary Fig. 8. **b**, Identified targets: CALM3, CINP and MAPK8IP2 are kinase-associated hits. Hits were confirmed in an independent shRNAi screen and validated by two or more independent siRNAs indicated (Supplementary Fig. 8). Downregulation of targets was determined by quantitative qPCR or western blotting (WB) (Supplementary Fig. 8). **c**, Ingenuity Pathway Analysis of the kinase candidates affecting intracellular growth of *S. typhimurium*. In the network, the ten validated kinases are coded green. Open symbols, connecting proteins. PI3K, series of PI3Ks (PIK3C2G, PIK3CA, PIK3CB, PIK3CD and PIK3CG).

a

IC_{50} (μM)	H-89	ETB067	ETB064A	ETB064B	ETB007	ETB008	ETB009	ETB010	AKTi1/2
PKA	2	10	70	80	-	-	-	-	-
PKB/AKT	2	10	-	80	-	-	-	-	3
CAMK2B	10	-	-	-	ND	ND	ND	ND	ND
PRKD2	75	-	-	-	-	-	-	-	ND
PAK4	-	-	-	-	-	-	-	-	ND
PCTK1	-	-	-	-	-	-	-	-	-

b



c

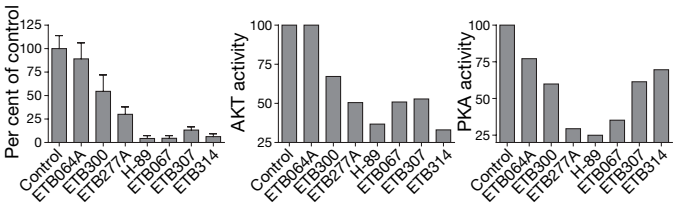


Figure 4 | Identification of host kinase targets for the chemical compounds inhibiting *S. typhimurium* and (MDR-)*M. tuberculosis*. **a**, *In vitro* kinase reaction with purified PKA, AKT1 and 4 GFP-tagged kinase domains. The half-maximal inhibitory concentration (IC_{50}) of kinase activity is determined. -, $IC_{50} > 100$ μM; ND, not determined. **b**, Human primary macrophages were infected with MDR-*M. tuberculosis* MDR16319 and cultured for 6 days in the presence of various kinase inhibitors. Shown is the mean of quadruplicate c.f.u. counts ± s.e.m. (*$P > 0.05$). **c**, The effects of H-89 variants (Supplementary Fig. 13) on intracellular *S. typhimurium* growth in human macrophages was tested 18 h post infection, determined by c.f.u. assays and related to untreated cells, set at 100% (left panel; shown is the mean of quadruplicate experiments ± s.e.m.). In addition, inhibition of 10 μM of compounds on isolated AKT1 (middle panel) or PKA (right panel) is shown as a mean of triplicate experiments.

©2007 Nature Publishing Group

not selectivity (Figs 2 and 4). The methyl group in the linker of ETB067 generated specificity, because CAMK2B is no longer inhibited by ETB067 (Fig. 4a). We increased the size of our chemical library and extended the methyl group in the linker and/or altered the halogen (Supplementary Fig. 13). Human macrophages were infected with *S. typhimurium* and the compounds tested (Fig. 4c; left panel). Replacing Br in H-89 for smaller halogens or hydrogen attenuated inhibition of intracellular *S. typhimurium* growth and decreased inhibition of both AKT1 and PKA (Fig. 4c). Extending the methyl group in the linker region of ETB067 did not affect bactericidal activity but selectively decreased PKA inhibition (Fig. 4c, and Supplementary Fig. 13). All combinations (halogens and extensions in the linker side chain) were tested for their effects on intracellular growth of *S. typhimurium* (Supplementary Fig. 13). The effect on intracellular *S. typhimurium* replication of the 16 H-89 variants was cross-correlated with the rate of AKT1 or PKA inhibition. An obvious correlation was observed only for AKT1 (Supplementary Fig. 13).

Application of compound ETB067 to *S. typhimurium*-infected mice following the first signs of illness prolonged survival (Supplementary Fig. 14). Notably, ETB067 neither affected viability of mice nor caused any detectable tissue damage at the dose given (Supplementary Fig. 15). Although better AKT inhibitors have entered clinical trials as anti-cancer drugs, application of our compounds illustrates the feasibility of using AKT inhibitors as antibiotics.

We next studied the mechanism of AKT inhibitors as bacteriostatic or bactericidal agents. AKT contains a PH domain, which binds phosphoinositides. Phosphoinositides are frequently manipulated and AKT often activated by intracellular bacteria[14]. *S. typhimurium* secretes SopB, an effector protein with phosphoinositide phosphatase activity into the host cytosol, and analogous effectors were identified for *Shighella flexneri*[15]. The resulting activation of AKT may prevent phagosomal maturation[16] and promote survival under nutrient-poor phagosomal conditions[17]. Other bacterial effectors activate the small GTPase Rac1, which controls actin dynamics during and following phagocytosis[18]. To show AKT activation by SopB, we infected MCF7 cells with control or SopB-deficient *S. typhimurium* before staining for activated S473-phosphorylated AKT (Fig. 5a). SopB activates AKT at the plasma membrane but not on phagosomes.

Because PAK4 activity is controlled by phosphatidylinositol-3-OH kinases[19], PAK4 (also identified in our screen) could be downstream of AKT. PAK4 binds to, but is not activated by CDC42 and phosphorylates the RHO/RAC GEF-H1 (also known as ARHGEF2), which then fails to activate RHOA, instead activating RAC1[20] and resulting in different and local effects on actin. We first showed that PAK4 was phosphorylated in cells expressing activated AKT1 and then showed that the phosphorylated PAK4 was more active (Supplementary Fig. 16a). AKT1 then lies upstream of PAK4, thus controlling RHOA, RAC1 and actin remodelling.

To test whether activation of the AKT–PAK4 axis aids survival of intracellular pathogens, MCF7 cells expressing control or constitutively active PAK4(S445N)[21] were infected with *S. typhimurium* and cultured in the presence or absence of AKT inhibitor AKTi1/2. Intracellular growth of DsRed-*S. typhimurium* was determined 18 h post infection by FACS (Fig. 5b). GFP–PAK4(S445N) only partially rescued intracellular expansion of DsRed-*S. typhimurium* in the presence of AKTi1/2, implying that other pathways downstream of AKT1 should also be involved.

Phagosome–lysosome fusion effects were reported in *Dictyostelium discoideum*[22], in which activation of AKT delays fusion of phagosomes to lysosomes to promote pathogen survival. Our inhibitors may counteract this block in phagosomal maturation, thus inducing lysosome fusion. Preventing lysosomal acidification therefore obscured effects of AKT inhibitors (Supplementary Fig. 16b). At least 18 Rab GTPases are implicated in phagosomal maturation[23]. RAB14 is involved in the control of *M. tuberculosis* phagosomal

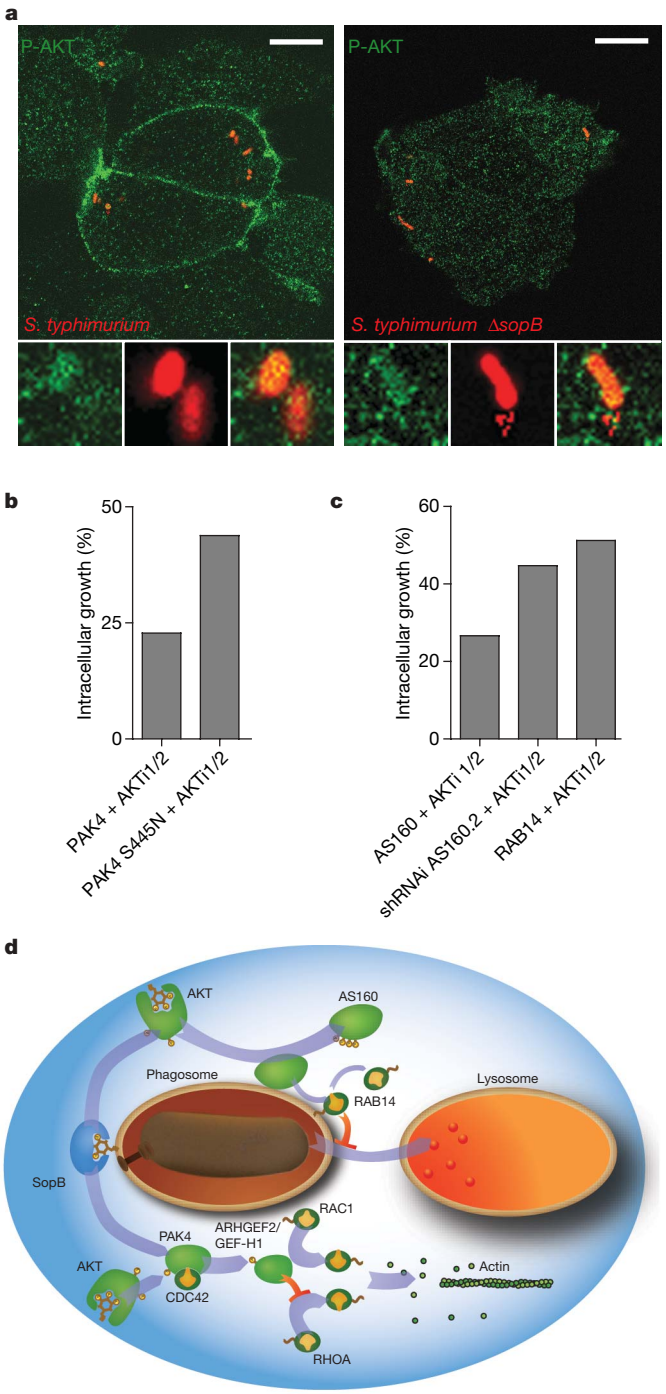


Figure 5 | Mechanisms of AKT1 control of intracellular *S. typhimurium* infections. **a**, MCF7 cells infected with control or SopB-deficient DsRed-expressing *S. typhimurium* before staining for activated AKT (P-AKT). Insets, higher magnification ($n > 100$). Scale bar, 10 μm. **b**, GFP-labelled PAK4 or constitutively active PAK4(S445N) expressed in MCF7 cells were infected and cultured with DsRed-*S.typhimurium* for 18 h with AKTi1/2. DsRed-*S. typhimurium* were quantified by FACS and related to the infection of GFP–PAK4-transfected cells, set at 100%. Shown is the mean of triplicate experiments. **c**, MCF7 cells were transfected with Myc–RAB14, AS160 or the shRNAi AS160.2 construct for AS160, and infected with GFP-*S. typhimurium* in the presence or absence of AKTi1/2. Intracellular *S. typhimurium* was quantified by FACS 18 h later and compared to untreated AS160-transfected MCF7 cells (set at 100%). Shown is the mean of triplicate experiments. **d**, *S. typhimurium* effector SopB activates AKT. AKT1 targets PAK4, which phosphorylates GEF-H1, thus controlling RHOA, RAC1 and actin. AKT1 also phosphorylates the RAB14-GAP, AS160. This prevents AS160 binding to phagosomal membranes, thus activating RAB14 and inhibiting phagosomal maturation.

©2007 Nature Publishing Group

maturation[24]. RabD, the *D. discoideum* RAB14, regulates phagocytosis and homotypic phagosome and lysosome fusion[25]. We followed GFP–RAB14 during infection with control or SopB-deficient DsRed-*S. typhimurium* (Supplementary Fig. 16c). GFP–RAB14 localized on *trans*-Golgi and endosomal structures and was recruited to phagosomes only with SopB-expressing *S. typhimurium*. Because Rab GTPases cycle between active membrane-bound GTP and inactive cytosolic GDP forms, the Rab cycle can be monitored in living cells using photo-activation protocols (Supplementary Fig. 16d). The $t_{1/2}$ of phagosomal membrane dissociation of GTP-loaded PAGFP–RAB14 was 70 s, and this markedly decreased following inhibition of AKT in *S. typhimurium*-infected cells (Supplementary Fig. 16e).

The RAB14 cycle may be accelerated by AS160 (ref. 26) (the AKT substrate of 160 kDa; TBC1D4), a potential RAB14-GTPase activator protein (GAP). This may explain how AKT affects the RAB14 cycle. Silencing *AS160* by shRNA constructs—but also *in vitro* GTP hydrolysis experiments—identified AS160 as the GAP for RAB14 (Supplementary Fig. 16f). AS160 becomes inactivated following phosphorylation by AKT and fails to bind membranes[26]. Chemically inhibiting AKT after activation by *S. typhimurium* accelerates the RAB14 cycle (Supplementary Fig. 16e), suggesting bacterial effectors like SopB activate AKT (at the plasma membrane; Fig. 5a) to phosphorylate cytosolic AS160. Phosphorylated AS160 then fails to bind to phagosomal membranes and RAB14 remains activated. Active RAB14 then recruits (unidentified) effectors that inhibit phagosomal fusion with lysosomes[24]. AKT1 inhibitors counteract these bacterial manipulations of the host machinery, resulting in bacterial growth inhibition. To test whether AKT activity on AS160 and RAB14 affects intracellular growth of bacteria, we introduced an shRNA for *AS160* or an expression vector for *AS160* or *RAB14* in MCF7 cells before infection with DsRed-*S. typhimurium*. Cells were cultured following infection with inhibitor AKTi1/2 for 18 h before FACS analyses to determine intracellular growth (Fig. 5c). Inhibition of AKT1 restricted intracellular growth of *S. typhimurium*, which was partially reversed by downregulating AS160 or overexpressing its target, RAB14. This suggests that AKT inhibitors control RAB14 via AS160, inducing phagosomal maturation to phagolysosomes, and thus controlling intracellular bacterial growth. How other kinases identified in the screen (Fig. 3b, c) contribute to bacterial survival remains to be determined.

Thus, AKT1 acts as a master regulator by controlling at least three GTPases and two essential host pathways, notably PAK4–RAC1/RHOA and AS160–RAB14, which can be exploited by various intracellular pathogens for intracellular survival[17,27] (Fig. 5d). Our AKT inhibitors counteract this bacterial manipulation of host processes, resulting in antibiotic activities.

We have used reciprocal chemical genetics to identify host kinases that are essential in controlling intracellular infections and can be targeted by novel chemical compounds to control infection *in vitro* and delay infection-induced mortality *in vivo*. This procedure reverses the normal route followed in drug discovery in which first a target and then a compound is defined. Our rationale was to identify a class of proteins (kinases) using chemical inhibitors with relevant biological activities. Subsequent synthesis of inhibitor variants for chemical profiling, and RNAi screening of the relevant protein family, identified potential host targets. AKT1 was identified by this approach and implicated in intracellular survival of *S. typhimurium* and other pathogens. We identified an inhibitor of AKT with limited specificity that still showed therapeutic potential as an antibiotic without noticeable side effects in mice, and we identified new mechanisms of bacterial manipulation of host cells. Other AKT inhibitors are already in phase II trials as anti-cancer drugs[28] because AKT is found activated in many human tumours. Chronic infections are linked to increased cancer incidences. Patients with chronic *Salmonella typhi* infections, which usually reside in the biliary tract, have a reported tenfold higher incidence in gallbladder cancer[29]. Sustained activation of the AKT pathway by SopB for the survival of the intracellular *S. typhi* may well contribute towards cell transformation and gallbladder cancer. Many anti-cancer drugs were a spin-off from the search for antibiotics in the 1950s. Conversely, the current development of inhibitors of AKT, PAK4 and other kinases as anti-cancer drugs may yield novel antibiotics that counteract host pathways activated by intracellular pathogens for survival.

METHODS SUMMARY

Molecular methods, reagents and chemistry. The cloning of various constructs, generation of various *S. typhimurium* strains, growth and infection conditions, and *in vitro* kinase and GTPase assays are described in detail in the Methods. The details on the chemistry of the H-89 variants with nuclear mass resonance and mass spectrometry information are supplied in Supplementary Fig. 2. The details of the large-scale synthesis of ETB067 for the *in vivo* mouse experiments are given in Supplementary Fig. 2. The Student's *t*-test was applied where indicated.

Confocal microscopy and live-cell imaging. Confocal microscopy was performed with a Leica AOBS LSCM (Leica Microsystems) equipped with a 405 nm laser for photo-activation experiments in a 37 °C climate chamber. For details, see Methods. The decay of phagosomal photo-activatable-GFP–RAB14 was calculated using MATLab software.

Automated microscopy, siRNA screening of the human kinome and statistical analysis. MCF7 cells were transfected with pooled siRNAs for every individual kinase (Dharmacon) present in the human genome before infection with GFP-*S. typhimurium* and analysis by automated microscopy (BD Pathway Bioimager). Images were analysed with CellProfiler (open-source cell image analysis software; Supplementary Figs 3, 5 and 6). The image statistics are shown in Methods. The heat map was generated using MATLAB software. Targets were validated with multiple individual siRNAs for the same target using Lux-*S. typhimurium*, and were quantified by luminometry. Pathway analysis was performed using Ingenuity Pathway Analysis programs. For details, see Methods.

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

Received 4 July; accepted 4 October 2007.

1. Walsh, C. Molecular mechanisms that confer antibacterial drug resistance. *Nature* **406,** 775–781 (2000).
2. Meresse, S. *et al.* Controlling the maturation of pathogen-containing vacuoles: a matter of life and death. *Nature Cell Biol.* **1,** E183–E188 (1999).
3. Vergne, I., Chua, J., Singh, S. B. & Deretic, V. Cell biology of mycobacterium tuberculosis phagosome. *Annu. Rev. Cell Dev. Biol.* **20,** 367–394 (2004).
4. Boucrot, E., Henry, T., Borg, J. P., Gorvel, J. P. & Meresse, S. The intracellular fate of *Salmonella* depends on the recruitment of kinesin. *Science* **308,** 1174–1178 (2005).
5. Harrison, R. E., Bucci, C., Vieira, O. V., Schroer, T. A. & Grinstein, S. Phagosomes fuse with late endosomes and/or lysosomes by extension of membrane protrusions along microtubules: role of Rab7 and RILP. *Mol. Cell. Biol.* **23,** 6494–6506 (2003).
6. Uchiya, K. & Nikai, T. *Salmonella enterica* serovar Typhimurium infection induces cyclooxygenase 2 expression in macrophages: involvement of *Salmonella* pathogenicity island 2. *Infect. Immun.* **72,** 6860–6869 (2004).
7. Kashina, A. S. *et al.* Protein kinase A, which regulates intracellular transport, forms complexes with molecular motors on organelles. *Curr. Biol.* **14,** 1877–1881 (2004).
8. Engh, R. A., Girod, A., Kinzel, V., Huber, R. & Bossemeyer, D. Crystal structures of catalytic subunit of cAMP-dependent protein kinase in complex with isoquinolinesulfonyl protein kinase inhibitors H7, H8, and H89. Structural implications for selectivity. *J. Biol. Chem.* **271,** 26157–26164 (1996).
9. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428,** 431–437 (2004).
10. Agaisse, H. *et al.* Genome-wide RNAi screen for host factors required for intracellular bacterial infection. *Science* **309,** 1248–1251 (2005).
11. Philips, J. A., Rubin, E. J. & Perrimon, N. *Drosophila* RNAi screen reveals CD36 family member required for mycobacterial infection. *Science* **309,** 1251–1253 (2005).
12. Pelkmans, L. *et al.* Genome-wide analysis of human kinases in clathrin- and caveolae/raft-mediated endocytosis. *Nature* **436,** 78–86 (2005).
13. Davies, T. G. *et al.* A structural comparison of inhibitor binding to PKB, PKA and PKA–PKB chimera. *J. Mol. Biol.* **367,** 882–894 (2007).
14. Chua, J., Vergne, I., Master, S. & Deretic, V. A tale of two lipids: *Mycobacterium tuberculosis* phagosome maturation arrest. *Curr. Opin. Microbiol.* **7,** 71–77 (2004).
15. Pendaries, C. *et al.* PtdIns5P activates the host cell PI3-kinase/Akt pathway during *Shigella flexneri* infection. *EMBO J.* **25,** 1024–1034 (2006).
16. Hernandez, L. D., Hueffer, K., Wenk, M. R. & Galan, J. E. *Salmonella* modulates vesicular traffic by altering phosphoinositide metabolism. *Science* **304,** 1805–1807 (2004).
17. Knodler, L. A., Finlay, B. B. & Steele-Mortimer, O. The *Salmonella* effector protein SopB protects epithelial cells from apoptosis by sustained activation of Akt. *J. Biol. Chem.* **280,** 9058–9064 (2005).

©2007 Nature Publishing Group

18. Scott, C. C. *et al.* Phosphatidylinositol-4,5-bisphosphate hydrolysis directs actin remodeling during phagocytosis. *J. Cell Biol.* **169,** 139–149 (2005).
19. Wells, C. M., Abo, A. & Ridley, A. J. PAK4 is activated via PI3K in HGF-stimulated epithelial cells. *J. Cell Sci.* **115,** 3947–3956 (2002).
20. Callow, M. G., Zozulya, S., Gishizky, M. L., Jallal, B. & Smeal, T. PAK4 mediates morphological changes through the regulation of GEF-H1. *J. Cell Sci.* **118,** 1861–1872 (2005).
21. Qu, J. *et al.* Activated PAK4 regulates cell adhesion and anchorage-independent growth. *Mol. Cell. Biol.* **21,** 3523–3533 (2001).
22. Rupper, A. C., Rodriguez-Paris, J. M., Grove, B. D. & Cardelli, J. A. p110-related PI 3-kinases regulate phagosome-phagosome fusion and phagosomal pH through a PKB/Akt dependent pathway in *Dictyostelium*. *J. Cell Sci.* **114,** 1283–1295 (2001).
23. Smith, A. C. *et al.* A network of Rab GTPases controls phagosome maturation and is modulated by *Salmonella enterica* serovar Typhimurium. *J. Cell Biol.* **176,** 263–268 (2007).
24. Kyei, G. B. *et al.* Rab14 is critical for maintenance of *Mycobacterium tuberculosis* phagosome maturation arrest. *EMBO J.* **25,** 5250–5259 (2006).
25. Harris, E. & Cardelli, J. RabD, a *Dictyostelium* Rab14-related GTPase, regulates phagocytosis and homotypic phagosome and lysosome fusion. *J. Cell Sci.* **115,** 3703–3713 (2002).
26. Miinea, C. P. *et al.* AS160, the Akt substrate regulating GLUT4 translocation, has a functional Rab GTPase-activating protein domain. *Biochem. J.* **391,** 87–93 (2005).
27. Pizarro-Cerda, J. & Cossart, P. Subversion of phosphoinositide metabolism by intracellular bacterial pathogens. *Nature Cell Biol.* **6,** 1026–1033 (2004).
28. Granville, C. A., Memmott, R. M., Gills, J. J. & Dennis, P. A. Handicapping the race to develop inhibitors of the phosphoinositide 3-kinase/Akt/mammalian target of rapamycin pathway. *Clin. Cancer Res.* **12,** 679–689 (2006).
29. Strom, B. L. *et al.* Risk factors for gallbladder cancer. An international collaborative case-control study. *Cancer* **76,** 1747–1756 (1995).

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

Acknowledgements This work was supported by a program grant from N.W.O. and the Dutch Cancer Society KWF to J.N. and T.H.M.O., and an EEC network grant (Microban). We thank J. Sung for the histochemical analyses of mouse tissues, W. Mooi and W. Zwart for kinase-compound modelling, S. Commandeur and K. Walburg for support in the chemical-profiling experiments, D. Holden for SopB-deficient *S. typhimurium*, L. Wilson for help with culturing *M. tuberculosis* strains, H. v. d. Elst for assistance in purification and LCMS analysis of the chemical inhibitors, K. Kremer for providing *M. tuberculosis* clinical isolates, K. Nealson and S. Vesterlund for Lux-*S. typhimurium*, C. Wells and A. Ridley for haemagglutinin-tagged *PAK4* contructs, and G. Lienhard for *AS160* constructs.

Author Contributions C.K. performed the kinase assays (with J.N.) and experiments with RAB14, AS160 (with M.K.) and PAK4. M.M. and C.K. performed the shRNAi screen with a library made by R.L.B. and obtained from D.A.E. C.K. performed the siRNA screening, automated microscopy and data analyses, supported by the NKI robotics facility (R.L.B., D.E.). Subcloning and sequencing was by L.J. A.T., R.v.d.N. and A.P. synthesized the kinase inhibitors under the supervision of H.O. and G.v.d.M. N.D.L.S., S.J.F.v.d.E. and A.G. performed *in vivo* and chemical-profiling experiments (supervised by T.H.M.O.). J.N. supervised the study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.N. (j.neefjes@nki.nl).

©2007 Nature Publishing Group

METHODS

Constructs and reagents. For the *in vitro* kinase assay the following *Image* clones encoding kinases were ordered at the MGC or ORIGENE and their kinase domains cloned by PCR into the pEGFP-C1 mammalian expression vector (Clontech): full-length *AKT1* (T308D, S473D) and the DNA encoding the catalytically active forms were a gift of M. Voorhoeve. Full-length PAK4 (amino acids 2–591, Origene, NM_005884.3), PRKD2 (438–721, MGC4127433), full-length PCTK1 (5153A) and CAMK2B (10–573, MGC29528) were used as templates for generating GFP-tagged kinase domains using specific PCR primers. HA–PAK4 constructs were a gift of C. Wells and A.Ridley[30]. HA-tagged AS160 constructs[31] were a gift of G. Lienhard. The CD8 expression construct is described in ref. 32. The construct encoding CaPAK4(S445N)[33] was made by PCR.

For GFP–RAB14, full-length human *RAB14* complementary DNA (IMAGE: 2963119) was amplified by PCR. The PCR fragment was cloned into pGEMTeasy (Promega), and then pEGFP C1 (Clontech), to obtain constructs encoding amino-terminal enhanced-GFP-tagged RAB14. EGFP was exchanged for photo-activatable-GFP[34] using the PA-GFP construct.

RNAi against *AS160* was performed according to a procedure described previously[35]. In short, a primer directing the synthesis of shRNA complementary to human *AS160*, was cloned 5′ of the *H1-RNA* promoter in the pSUPER vector. The target sequence shRNAi AS160.1 corresponds to *AS160* 394–416 nucleotides (sequence: 5′-AACAGCCACGACCTCACCTACTT), and shRNAi AS160.2 to an already characterized sequence[36]. The specificity of the target sequences was verified by BLAST search against the gene data bank.

PKI was a gift from J. Zhang[37] and subcloned into pIRES2-DsRed2 (Clontech). All DNA constructs were verified by sequencing.

Antibodies. Antibodies used were: rabbit polyclonal anti-CD63 (ref. 38) and mouse monoclonal anti-human-CD63 (Cymbus Biotechnology), mouse monoclonal anti-human-CD8 (BD Biosciences), mouse monoclonal anti-human-AKT(2H10) and rabbit monoclonal phospho-AKT(193H12) (Cell Signalling), and mouse monoclonal anti-human LC3 (nanoTOOLS). In addition: rabbit polyclonal anti-*S. typhimurium*-LPS (Difco laboratories) and mouse monoclonal 1E6 anti-*S. typhimurium*-LPS (Biodesign International). Secondary antibodies were: Alexa 488- or Texas Red/Alexa 568-conjugated mouse or rabbit (Molecular Probes).

Cell lines. MCF7 cells were maintained in Iscoves medium (GIBCO-BRL) supplemented with 7.5% FCS, 2 mM glutamine, 100 U ml^{-1} penicillin, 100 μg ml^{-1} streptomycin at 37 °C in 5% CO_2. Human primary macrophages (mf1) were generated by differentiating human monocytes[39].

Animals. Female 6–8-week-old BALB/c, C57BL/6 and C3H/HeNHsd mice were purchased from Charles River Laboratories and maintained in specific pathogen-free conditions within the animal facility at the LUMC. All experiments were ethically approved (Leiden LUMC UDEC 05103).

Inhibitors. All inhibitors were used at concentrations reported to inhibit their kinase target. The PKA inhibitor H-89 (Biomol or Sigma) was used at a concentration of 10 μM, Rp-cAMP (Biomol) was used at a concentration of 100 μM. The following RHO kinase inhibiters were used: H-1152 and Y-27632 (Calbiochem) at concentrations of 10 and 50 μM. SB203580 (Invitrogen) was used to inhibit p38MAPK, and LY 294002 (Calbiochem) was used to inhibit PI3K (concentrations 10 and 50 μM). Casein kinase was inhibited by DRB (5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole, Calbiochem) and calmodulin kinase by KN-93 (Calbiochem) at concentrations of 10 and 50 μM. Myosin light-chain kinase was inhibited by ML-7 (Biomol) at 20 μM concentration. The PKB inhibitor AKTi1/2 (Calbiochem) was used at a concentration of 10 μM. Inhibitors were added 1 h post infection along with (10 μg ml^{-1}) gentamycin to eliminate extracellularly growing *S. typhimurium* to determine the effects on intracellular growth of *S. typhimurium*.

Construction of new chemical compounds based on H-89 is described in detail in Supplementary Fig. 2. These compounds were dissolved in 100% ethanol.

Bacterial strains, growth conditions and infections. The *Salmonella* strains *S. typhimurium* SL1344 and GFP-*S. typhimurium* SL1344[40] were a gift from S. Meresse. Control and SopB-deficient *S. typhimurium* 12023 were a gift from D. Holden. DsRED (pMW266, a gift from D. Bumann) was expressed in *S. typhimurium* SL1344 (ref. 41). This construct was isolated and expressed in the various *S. typhimurium* 12023 strains. The *lux* operon from *P. luminescence* (*luxCD–ABE*) was isolated from *S. typhimurium* 14028 (a gift of S. Vesterlund)[42] and expressed in *S.typhimurium* 12023. MCF-7 cells were cultured at 37 °C in 5% CO_2 for 48 h in Iscoves medium without antibiotics. After overnight growth at 37 °C with shaking, bacteria were subcultured at a dilution of 1:33 in fresh LB medium and incubated for 3.5 h at 37 °C with shaking. MCF-7 cells were infected at multiplicity of infection (m.o.i.) 100 (100 bacteria per cell unless indicated otherwise) in Iscoves without antibiotics for 30 min at 37 °C in 5% CO_2. Infected cells were washed 4–6 times and incubated for 1 h in tissue culture medium containing 100 μg ml^{-1} gentamycin (GIBCO) to eliminate extracellular bacteria. Infected cells were subsequently incubated for the indicated time points in medium containing 10 μg ml^{-1} gentamycin.

Primary macrophages were infected in quadruplicate wells with log phase cultures of *S. typhimurium* or mycobacteria (*M. smegmatis*, H37Rv *M. tuberculosis* or MDR-*M. tuberculosis* (Beijing/W family resistant to rifampacin, isoniazide, ethambutol and pyrazinamide[43]) at an m.o.i. of 5 for 1 h at 37 °C. Inoculum was verified by c.f.u. assays on 7H10 plates. Extracellular bacteria were physically removed and the medium replaced with gentamycin-containing medium (3 μg ml^{-1}) and 10 μM of the appropriate kinase inhibitor. For experiments lasting more than 24 h, medium containing gentamycin and 10 μM kinase inhibitor was refreshed daily. To recover the bacteria from the infected macrophages, macrophages were washed 3 times with PBS before lysis in H_2O, and serial dilutions of lysates were plated out on 7H10 plates. Colony-forming units were counted after 2 days for *M. smegmatis*, 1 day for *S. typhimurium*, and after 3 weeks for *M. tuberculosis*. Lux activity could be directly (in the absence of exogenous substrates) detected through the activity of proteins encoded in the *P. luminescence* (*luxCD–ABE*) operon. The *luxCDE* gene products are required for the synthesis of long-chain fatty aldehydes that were used as substrate for light production by the *luxAB* gene products[42]. Light was detected by chemoluminescence (Lumat LB 9507 EG&G BERTHOLD). Alternatively, internalization and expansion of GFP-*S. typhimurium* was determined by two-colour FACS, in which CD8 was detected as a transfection marker. Two-colour FACS was also used when the effect of GFP-tagged proteins on expansion of DsRED-*S. typhimurium* was determined.

***In vivo* experiments to determine the activity of ETB067 on *S. typhimurium* survival.** *S. typhimurium* (SL1344) were grown statically and 2×10^3 cells (in a volume of 100μl of PBS) were injected intraperitoneally in female 8–12-week-old C3H/HeNHsd mice. Mice were given a single dose of 4 mg ETB067 in 8% ethanol/PBS solution (vehicle control, 8% ethanol only), intraperitoneally, in a volume of 100 μl per injection at day 6 and 8 post infection, when the animal showed visible signs of infection on the basis of weight loss. Mice were maintained in SPF conditions and the experiments were approved by the local ethical committee of the LUMC, Leiden (DEC protocol 05103).

siRNA screen of the human kinome, validation and qPCR. A synthetic siRNA library was obtained from Dharmacon, targeting 779 kinases and kinase-related proteins (Human siARRAY SMARTpool Reagent - Protein Kinases; Dharmacon). MCF-7 cells were seeded in 384 well μCLEAR plates (Greiner bio-one) at 2,000 cells per well. Reversed transfection was done with 0.1 μl Dharmafect4 and 50 nM siRNA in a total volume of 50 μl. Transfection efficiency (tested with siRNA for *GAPDH*) exceeded 95% under these conditions. The transfected cells were cultured for 48 h and infected with GFP-*S. typhimurium* SL1344 (50 μl of 2.67×10^7 bacteria per ml), according to the protocol described above. Eighteen hours post infection, the cells were fixed with 3.75% formaldehyde in PBS for 1 h. The nuclei were stained with Hoechst 33342 (2 μg ml^{-1}) (Molecular Probes), and the cell membranes with Wheat Germ Agglutinin (WGA) Alexa Fluor 647 (Molecular Probes). The microplates were imaged with the BD Pathway Bioimager (Becton Dickinson) with the following filters: Hoechst 33342 excitation 380/10, epifluorescence 400 DCLP, emission 435 LP; GFP excitation 470/10, epifluorescence Fura/FITC, emission 515 LP; WGA Alexa 647 excitation 635/20, epifluorescence 84,000, emission 84,101. The images were analysed with open-source cell image analysis software (www.cellprofiler.org)[44]. First, nuclei were identified, based on the Hoechst 33342 stain. Cell boundaries were then identified using the nuclear stain as a reference point and determining the WGA Alexa Fluor 647 fluorescence. From the data, the rate of intracellular proliferation of GFP-*S. typhimurium* was calculated, as depicted in Supplementary Fig. 3.

Data analysis, programming and statistical analyses are discussed in Supplementary Figs 5 and 6, and below.

Using MATLAB, the results were plotted as heat maps containing the individual kinases and kinase-related siRNAs. SiRNAs reducing intracellular growth of *S. typhimurium* were visually inspected to exclude any image artefacts. The siRNA screen of the human kinome was performed twice in triplicate with four images made of every individual well (36 images in total).

The mean GFP fluorescence per cell was measured and plotted as a histogram per image. For intracellular growth of *S. typhimurium*, H-89-treated samples were taken as a reference. Median (triplicate) GFP fluorescence per infected cells was compared between siRNA transfected and H-89-treated MCF7 cells and plotted in a heat map or as a bar diagram.

To validate the hits further, multiple individual siRNA sequences per target (Dharmacon) were tested for their effect on intracellular growth of Lux-*S. typhimurium*. Effects on intracellular growth were determined by luminometry. A hit was considered validated when confirmed by at least two independent siRNAs. The siRNA sequences validating the hits were: Supplementary Fig. 8a. The effects

©2007 Nature Publishing Group

of the different siRNAs on messenger RNA expression were determined by qPCR. For primer (Sigma) sequences used, see Supplementary Fig. 8a.

Role of PKB/AKT and downstream effectors in control intracellular *S. typhimurium* growth. MCF7 cells were transfected with expression constructs for PAK4, PAK4(S445N), GFP-RAB14 or the shRNA for AS160 and infected 48hr later with lux-*S. typhimurium*. 1 h post-infection, cells were cultured further in the presence or absence of the PKB/Akt inhibitor AKTi1/2 for 18 h before intracellular growth of lux-*S. typhimurium* was measured by luminometry.

***In vitro* kinase activity assays.** GFP-tagged full-length kinases or kinase domains were expressed in HEK293 cells by DEAE-dextran transfection. Proteins were isolated using anti-GFP pre-coupled IgG Dynabeads (Invitrogen). AKT (Upstate) contained additional kinase activities and we used immunopurified AKT1 instead. PKA (Sigma) was used in *in vitro* kinase assays. Kinase assays were performed in triplicate in a 40 μl volume containing: 20 mM MOPS; 25 mM β-glycerol phosphate, pH 7.2; 1 mM EGTA; 1 mM sodium orthovanadate; 1 mM dithiothreitol; 20 μM non-radioactive ATP; 15 mM $MgCl_2$; 10 μg Myelin Basic Protein (MBP; Sigma); 0.5 μl [γ-^{32}P] ATP (2 mCi ml^{-1}; GE Healthcare); chemical inhibitors (dissolved in DMSO) or vehicle control (DMSO). To detect AKT1 activity, a Crosstide synthetic peptide with two additional carboxy-terminal lysine residues (GRPRTSSFAEGKK) was used. This peptide is suitable as a substrate for kinases, including AKT, SGK, p70S6 kinase and MST1. The double lysine modification enables binding to P81 paper in radioactive kinase assays. The reactions were started by addition of the isolated beads with GFP-kinase domains or isolated proteins. In all reactions, DMSO constitutes 8% of the volume. Inhibitors were tested at 0, 1.25, 2.5, 5, 10, 20, 40 and 80 μM final concentrations and from the inhibition curves the IC_{50} values were determined. *In vitro* kinase reactions were incubated at 37 °C for 20–120 min (depending on the kinase) and terminated by spotting 20 μl of the reaction volume on P81 paper (Whatman). Free [$\gamma-^{32}$P] ATP was removed by extensive washing in 0.65 (v/v)% phosphoric acid, followed by one acetone wash and air-drying. Phosphorylation of MBP was detected using phosphorimager plates read by phosphoimaging (Fujix BAS 2000) and quantified by AIDA software.

***In vitro* GTPase assay.** The GTPase assay was performed as described previously[45]. Briefly, GST–RAB14 (amino acids 1–212) was produced in *Escherichia coli*, purified and loaded with [γ-^{32}P] GTP (10 mCi ml^{-1}, > 5,000 Ci $mmol^{-1}$; GE Healthcare) in the presence of 10 mM EDTA. Subsequently, $MgCl_2$ was added to a final concentration of 20 mM, followed by gel filtration on a Bio-Spin 6 column (BioRad) equilibrated with buffer (0.1 M Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 2 mM dithiothreitol, 0.5 M NaCl) to remove free [γ-^{32}P] GTP. Reaction mixtures containing the different fusion proteins in reaction buffer (40 mM Tris-HCl, pH 8.0, 50 mM NaCl, 8 mM $MgCl_2$, 1 mM dithiothreitol, 0.5 mM non-radioactive GTP, 0.1 mg ml^{-1} BSA, 1 mM phosphate, 1% glycerol) were assembled on ice. To investigate the GAP activity of AS160 towards RAB14, GST–RAB14 was incubated with different concentrations of GST–AS160 (GAP-domain, amino acids 856–1182). Individual reactions were stopped by adding 1 ml of charcoal suspension (7% w/v charcoal, 10% v/v ethanol, 0.1 M HCl, 10 mM KH_2PO_4)[46], and the mixture was centrifuged for 5 min in an Eppendorf centrifuge. Release of [^{32}P]-orthophosphate was quantified by liquid-scintillation-counting of the supernatant.

Cells were fixed and immuno-stained before analysis by confocal laser-scanning microscopy. Fixation was either in 3.7% formaldehyde for 15 min at room temperature followed by permeabilization in 0.1% Triton-X100 or in methanol (−20 °C) for 2 min.

Confocal analyses were performed using a Leica TCS SP or AOBS confocal laser-scanning microscope (CLSM; Leica Microsystems). Green fluorescence was detected at $\lambda > 515$ nm after excitation at $\lambda = 488$ nm. For dual analyses, green fluorescence was detected at $\lambda = 520$–560 nm. Red fluorochromes were excited at $\lambda = 568$ nm and were detected at $\lambda > 585$ nm.

To detect active phosphorylated AKT or GFP–RAB7 in *S. typhimurium*-infected MCF7 cells, these cells were infected with SopB- or control *S. typhimurium* 12023 for 45 min before fixation. Activated AKT was stained with anti-phosphoAKT (Cell Signalling) and secondary ALEXA-488-coupled antibodies before image analyses by CLSM.

For uptake experiments, GFP–RAB14-expressing MCF7 cells were infected with DsRED-labelled control or SopB-deficient *S. typhimurium* 12023 for 45 min before analyses by CLSM.

For photo-activation experiments, PA-GFP–RAB14-transfected MCF7 cells were infected with DsRED-labelled control or SopB-deficient *S. typhimurium* 12023 for 45 min, and living cells analysed in a 37 °C culture hood on a Leica AOBS CLSM (Leica Microsystems). The PA-GFP–RAB14 cycle was determined by photoactivation analysis[47, 48], in which the loss of membrane-associated PA-GFP–RAB14 was monitored after photo-activation with a 405 nm laser light using a point-bleach mode, operational with Leica Software. The decay curves for loss of membrane-associated fluorescence were determined using MATLab software. Alternatively, the PA-GFP–RAB14 cycle, reflected by loss of membrane-bound fluorescence was determined in MCF7 cells expressing PA-GFP–RAB14 in the presence or absence of the pSUPER shRNA constructs for AS160.

Statistical analysis of the data generated with Cellprofiler. Two independent siRNA Screens were performed to identify kinases involved in intracellular replication of *S. typhimurium*. Data acquisition is described in Supplementary Figs 5 and 6.

Interpretation of the data requires computational and statistical analysis techniques. We generated a scored phenotype (reduced intracellular *S. typhimurium* growth only considering infected cells) list from the raw data with the open-source Bioconductor/R package and cellHTS[49] (http://www.dkfz.de/signaling/cellHTS). The results are formatted in a HTML format that enables viewing of all the results (http://nature.labstore.nl). Owing to the edge effect the data from the first screen could not be B-score[50]-normalized and therefore not ranked (although ranking is included). Only by considering a sample's direct neighbours, was determination of whether a particular siRNA reduced intracellular growth possible. In the second screen, the data could be B-score-normalized and ranked. Data from the first and second screen were both visually inspected to exclude any artefacts in the automated image analysis or artefacts that could not be resolved by the computational analyses. The heat map in Supplementary Fig. 8c is based on the results of computational analysis only. The hit table in Fig. 3b is based on ranking, visual inspection and independent validation.

For statistics on the individual kinases see http://nature.labstore.nl.

30. Wells, C. M., Abo, A. & Ridley, A. J. PAK4 is activated via PI3K in HGF-stimulated epithelial cells. *J. Cell Sci.* **115**, 3947–3956 (2002).
31. Sano, H. *et al.* Insulin-stimulated phosphorylation of a Rab GTPase-activating protein regulates GLUT4 translocation. *J. Biol. Chem.* **278**, 14599–14602 (2003).
32. Nilsson, T., Jackson, M. & Peterson, P. A. Short cytoplasmic sequences serve as retention signals for transmembrane proteins in the endoplasmic reticulum. *Cell* **58**, 707–718 (1989).
33. Qu, J. *et al.* Activated PAK4 regulates cell adhesion and anchorage-independent growth. *Mol. Cell. Biol.* **21**, 3523–3533 (2001).
34. Patterson, G. H. & Lippincott-Schwartz, J. Selective photolabeling of proteins using photoactivatable GFP. *Methods* **32**, 445–450 (2004).
35. Brummelkamp, T. R., Bernards, R. & Agami, R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
36. Larance, M. *et al.* Characterization of the role of the Rab GTPase-activating protein AS160 in insulin-regulated GLUT4 trafficking. *J. Biol. Chem.* **280**, 37803–37813 (2005).
37. Zhang, J., Hupfeld, C. J., Taylor, S. S., Olefsky, J. M. & Tsien, R. Y. Insulin disrupts β-adrenergic signalling to protein kinase A in adipocytes. *Nature* **437**, 569–573 (2005).
38. Vennegoor, C. *et al.* Biochemical characterization and cellular localization of a formalin-resistant melanoma-associated antigen reacting with monoclonal antibody NKI/C-3. *Int. J. Cancer* **35**, 287–295 (1985).
39. Verreck, F. A. *et al.* Human IL-23-producing type 1 macrophages promote but IL-10-producing type 2 macrophages subvert immunity to (myco)bacteria. *Proc. Natl Acad. Sci. USA* **101**, 4560–4565 (2004).
40. Meresse, S., Steele-Mortimer, O., Finlay, B. B. & Gorvel, J. P. The rab7 GTPase controls the maturation of *Salmonella typhimurium*-containing vacuoles in HeLa cells. *EMBO J.* **18**, 4394–4403 (1999).
41. Zwart, W. *et al.* Spatial separation of HLA-DM/HLA-DR interactions within MIIC and phagosome-induced immune escape. *Immunity* **22**, 221–233 (2005).
42. Vesterlund, S., Paltta, J., Laukova, A., Karp, M. & Ouwehand, A. C. Rapid screening method for the detection of antimicrobial substances. *J. Microbiol. Methods* **57**, 23–31 (2004).
43. Kremer, K. *et al.* Comparison of methods based on different molecular epidemiological markers for typing of *Mycobacterium tuberculosis* complex strains: interlaboratory study of discriminatory power and reproducibility. *J. Clin. Microbiol.* **37**, 2607–2618 (1999).
44. Carpenter, A. E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
45. Askjaer, P. *et al.* RanGTP-regulated interactions of CRM1 with nucleoporins and a shuttling DEAD-box helicase. *Mol. Cell. Biol.* **19**, 6276–6285 (1999).
46. Bischoff, F. R. & Ponstingl, H. Catalysis of guanine nucleotide exchange of Ran by RCC1 and stimulation of hydrolysis of Ran-bound GTP by Ran-GAP1. *Methods Enzymol.* **257**, 135–144 (1995).
47. Dantuma, N. P., Groothuis, T. A., Salomons, F. A. & Neefjes, J. A dynamic ubiquitin equilibrium couples proteasomal activity to chromatin remodeling. *J. Cell Biol.* **173**, 19–26 (2006).
48. Jordens, I. *et al.* The Rab7 effector protein RILP controls lysosomal transport by inducing the recruitment of dynein–dynactin motors. *Curr. Biol.* **11**, 1680–1685 (2001).
49. Boutros, M., Bras, L. P. & Huber, W. Analysis of cell-based RNAi screens. *Genome Biol.* **7**, R66 (2006).
50. Malo, N., Hanley, J. A., Cerquozzi, S., Pelletier, J. & Nadon, R. Statistical practice in high-throughput screening data analysis. *Nature Biotechnol.* **24**, 167–175 (2006).

©2007 Nature Publishing Group

LETTERS

Calcineurin sets the bandwidth for discrimination of signals during thymocyte development

Elena M. Gallo[1,2], Monte M.Winslow[1,2]†, Kirsten Canté-Barrett[1,2]†, Amy N. Radermacher[1,2], Lena Ho[1,2], Lisa McGinnis[1,2], Brian Iritani[3], Joel R. Neilson[1,4]† & Gerald R. Crabtree[1,2]

At critical times in development, cells are able to convert graded signals into discrete developmental outcomes; however, the mechanisms involved are poorly understood. During thymocyte development, cell fate is determined by signals originating from the αβ T-cell receptor. Low-affinity/avidity interactions between the T-cell receptor and peptide–MHC complexes direct differentiation to the single-positive stage (positive selection), whereas high-affinity/avidity interactions induce death by apoptosis (negative selection)[1,2]. Here we show that mice deficient in both calcineurin and nuclear factor of activated T cells (NFAT)c2/c3 lack a population of preselection thymocytes with enhanced ability to activate the mitogen-activated protein kinase (Raf–MEK–ERK) pathway, and fail to undergo positive selection. This defect can be partially rescued with constitutively active Raf, indicating that calcineurin controls MAPK signalling. Analysis of mice deficient in both Bim (which is required for negative selection) and calcineurin revealed that calcineurin-induced ERK (extracellular signal-regulated kinase) sensitization is required for differentiation in response to 'weak' positive selecting signals but not in response to 'strong' negative selecting signals (which normally induce apoptosis). These results indicate that early calcineurin/NFAT signalling produces a developmental period of ERK hypersensitivity, allowing very weak signals to induce positive selection. This mechanism might be generally useful in the discrimination of graded signals that induce different cell fates.

The calcineurin/NFAT[3,4] and the Raf–MEK–ERK[5–7] pathways have been shown to be required for positive selection of thymocytes but not for their negative selection. Calcineurin B1 (*Cnb1*)-deficient $CD4^+$ $CD8^+$ double-positive thymocytes lack calcineurin activity, fail to dephosphorylate NFATc transcription factors and are not positively selected (Fig. 1a and ref. 3). *Cnb1*-deficient thymocytes have normal phosphorylation of JNK (c-Jun N-terminal kinase), p38, protein kinase C-θ, protein kinase D and glycogen synthase kinase 3α after crosslinking of the T-cell receptor (TCR) (Fig. 1b and ref. 3). Actin polymerization and Ca^{2+} influx were also normal in these cells (Supplementary Fig. 2a, b). Collectively, these results indicated that *Cnb1*-deficient thymocytes did not have widespread signalling defects downstream of the TCR. However, *Cnb1*-deficient thymocytes showed a specific and severe defect in ERK1/2 phosphorylation (Fig. 1c), resulting in almost undetectable induction of the ERK/Elk4 target gene Egr1 (ref. 6) after engagement of the TCR (Fig. 1d). In addition, both MAP-kinase kinase (MEK) 1/2 and Raf activation were defective in *Cnb1*-deficient double-positive thymocytes (Fig. 1c, e, f) after stimulation by crosslinking of the TCR.

To test whether calcineurin activity was directly required for Raf activation, we stimulated thymocytes in the presence of the calcineurin inhibitor cyclosporin A (CsA) or the MEK1/2 inhibitor UO126. Acute *in vitro* inhibition of MEK1/2 with UO126 but not inhibition of calcineurin activity with CsA impaired ERK1/2 phosphorylation and Egr1 induction (Fig. 2a, b). In contrast, 10-day treatment of mice *in vivo* with CsA consistently recapitulated both the block in positive selection and defective ERK1/2 phosphorylation observed in *Cnb1*-deficient thymocytes (Fig. 2a, b and Supplementary Fig. 3a). These data indicated that calcineurin activity was not directly required for ERK1/2 phosphorylation but was instead

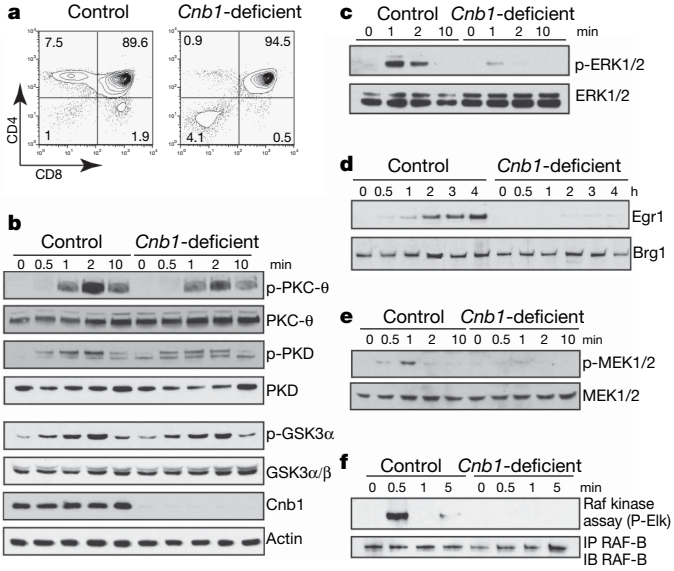


Figure 1 | Specific and severe defect in Raf–MEK–ERK activation in *Cnb1*-deficient thymocytes. a, Expression of CD4 and CD8 on *Cnb1*-deficient and control thymocytes. The numbers in the corners of the panels represent the percentage of cells in each quadrant. **b**, Immunoblot analysis of phosphorylated and total proteins in *Cnb1*-deficient and control double-positive thymocytes after CD3ε crosslinking. GSK, glycogen synthase kinase; PKC, protein kinase C; PKD, protein kinase D. **c**, Immunoblot analysis of phosphorylated ERK1/2 in *Cnb1*-deficient and control double-positive thymocytes after CD3ε crosslinking. **d**, Immunoblot analysis of Egr1 induction in double-positive thymocytes from *Cnb1*-deficient and control littermates. Brg1 shows equal loading. **e**, Immunoblot analysis of phosphorylated MEK1/2 in *Cnb1*-deficient and control double-positive thymocytes after CD3ε crosslinking. **f**, Raf-B kinase activity in *Cnb1*-deficient and control double-positive thymocytes after CD3ε crosslinking. IP, immunoprecipitation; IB, immunoblotting.

[1]Howard Hughes Medical Institute and the Departments of Pathology and of Developmental Biology, Stanford University, Stanford, California 94305, USA. [2]Program in Immunology, Stanford University, Stanford University, Stanford, California 94305, USA. [3]Department of Comparative Medicine, School of Medicine, University of Washington, Seattle, Washington 98195, USA. [4]Department of Microbiology and Immunology, Stanford University, Stanford, California 94305, USA. †Present addresses: Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (M.M.W., J.R.N.); Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden 2300 RC, The Netherlands (K.C.-B.).

©2007 Nature Publishing Group

required during development to acquire the ability to activate ERK1/2 properly in response to subsequent TCR signalling.

When ERK1/2 phosphorylation was assayed by intracellular staining, two populations with different levels of ERK1/2 phosphorylation were observed in control double-positive CD69-negative thymocytes after crosslinking with anti-CD3ε antibody. The population of thymocytes with the higher level of ERK1/2 phosphorylation was absent in *Cnb1*-deficient and CsA-treated mice (Fig. 2c and Supplementary Fig. 3b). The specificity of the staining and uniformity of crosslinking was confirmed by pretreating the cells with UO126 and by counterstaining for anti-CD3ε, respectively (Supplementary Fig. 3c, d). We also examined the expression of TCR-β on *Cnb1*-deficient and control thymocytes and found no difference in the percentage of thymocytes with intermediate levels of TCR-β expression (double-positive TCR-β^{int}), a population that is absent from mice that fail to rearrange the TCR-α chain[8–10] (Supplementary Fig. 3e). We concluded that two thymocyte populations existed, within the double-positive CD69-negative population, with different abilities to phosphorylate ERK1/2 and that calcineurin activity was required for the presence of the population with an increased ability to phosphorylate ERK1/2. We refer to these populations as 'ERK low competence' and 'ERK high competence' populations, respectively.

We used a single bromodeoxyuridine (BrdU) pulse to mark developing thymocytes[11] and establish the precursor/progeny relationship of these two populations. At early time points, most BrdU-positive, double-positive thymocytes were in the 'ERK low competence' state, with the percentage of BrdU-positive 'ERK high competence' double-positive thymocytes increasing over time (Fig. 2d, upper panel). The transition to the 'ERK high competence' state was not observed in *Cnb1*-deficient thymocytes (Fig. 2d, lower panel). These data indicated that the 'ERK low competence' state preceded the 'ERK high competence' state during normal thymocyte development and that calcineurin activity was required for this transition. The lag time for the development of the 'ERK high competence' population was consistent with a requirement for transcription.

Because the calcineurin phosphatase complex regulates the NFATc family of transcription factors[12,13], we analysed *NFATc2/NFATc3* double-knockout thymocytes, which also have impaired positive selection[4]. The development of the 'ERK high competence' population was impaired in *NFATc2/NFATc3* double-knockout thymocytes (Fig. 2e), the incomplete block probably reflecting a partly redundant function of NFATc1. These data indicated that calcineurin/NFAT signalling was required for the transition to the 'ERK high competence' state. The developmental requirement for calcineurin and NFAT activity for normal activation of the ERK pathway was specific to thymocytes because short-term or long-term treatment with cyclosporin had no effect on peripheral T and B lymphocytes (Supplementary Fig. 3f and data not shown). Analysis of MHCI/MHCII double-knockout mice revealed that the development of the 'high ERK competence' double-positive thymocyte population did not require TCR–MHC (major histocompatibility complex) interaction or positive selection (Supplementary Fig. 4a). Microarray analysis showed that 312 transcripts were differentially expressed in double-positive thymocytes from untreated and CsA-treated MHCI/MHCII double-knockout mice, indicating that calcineurin activity is required for preconditioning of double-positive

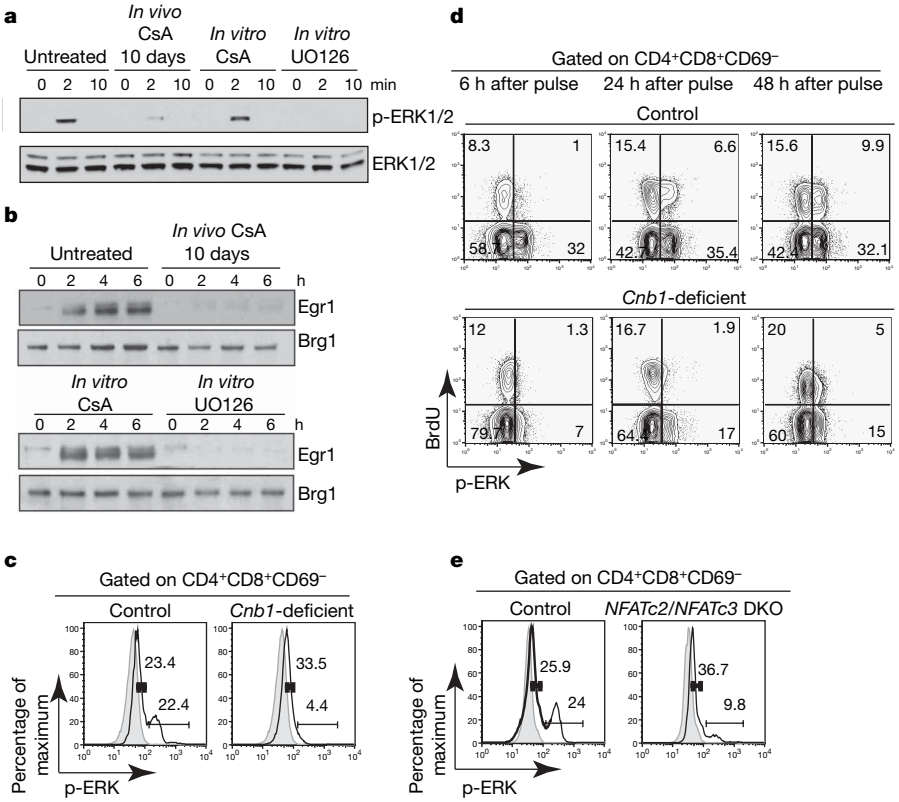


Figure 2 | Developmental but not direct requirement for calcineurin/NFAT activity for proper activation of ERK. **a**, ERK1/2 phosphorylation in double-positive cells from untreated and CsA-treated mice in the presence of CsA (200 ng ml^{-1}) or UO126 (10 μM) after CD3ε crosslinking. **b**, Erg1 induction in double-positive cells from untreated or CsA-treated mice and in double-positive cells stimulated in the presence of CsA (200 ng ml^{-1}) or UO126 (10 μM) after CD3ε crosslinking. Brg1 shows equal loading. **c**, ERK1/2 phosphorylation in double-positive CD69-negative *Cnb1*-deficient and control thymocytes after CD3ε crosslinking for 2 min (solid lines). Grey areas, unstimulated. **d**, BrdU incorporation and ERK1/2 phosphorylation in double-positive CD69-negative thymocytes from *Cnb1*-deficient mice and control littermates injected once with BrdU after CD3ε crosslinking for 2 min. The numbers in the corners of the panels represent the percentage of cells in each quadrant. **e**, ERK1/2 phosphorylation in *NFATc2/NFATc3* double knockout (DKO) and control double-positive CD69-negative thymocytes after CD3ε crosslinking for 2 min (solid lines). Grey lines, unstimulated. The numbers in graphs **c** and **e** represent the percentage of cells in the indicated interval.

©2007 Nature Publishing Group

thymocytes independently of positive selection (Supplementary Fig. 4b, c, and Supplementary Table 1). We speculate that the development of the 'ERK high competence' population depends on pre-TCR signalling and/or ligand-independent tonic TCR-αβ signalling and that the modulation of Raf activity could be due to a program of gene expression rather than a single modulator working at a single point in the pathway.

If a main mechanism by which calcineurin controlled positive selection were by modulating the sensitivity of the Raf–MEK–ERK pathway, restoring Raf signalling in the absence of calcineurin activity should lead to at least a partial rescue of positive selection. To test this, we obtained mice whose thymocytes express a constitutively active Raf-1 mutant protein (*Raf-CAAX* transgenic mice[14]), which does not induce positive selection in the absence of TCR signalling[14]. Mice were analysed between four and six weeks of age, before they developed any sign of thymic lymphoma. Long-term treatment with CsA completely blocked positive selection in control mice (Fig. 3a). However, positive selection is rescued in CsA-treated Raf-CAAX mice as assessed by the upregulation of CD69 and TCR-β and the development of mature $CD24^{low}Qa2^{high}$ single-positive CD4 and CD8 thymocytes (Fig. 3a–d). Moreover, CD4 single-positive and CD8 single-positive thymocytes that are rescued in CsA-treated Raf-CAAX mice were able to respond functionally when stimulated with anti-CD3 and anti-CD28 antibodies (Supplementary Fig. 5). Analogous results were obtained when Raf-CAAX was expressed in *Cnb1*-deficient thymocytes (Supplementary Fig. 6). As expected, the decreased thymic cellularity that is observed in the absence of calcineurin activity as a consequence of impaired transition from double negative to double positive was not rescued by the *Raf-CAAX* transgene (Fig. 3e and Supplementary Fig. 6d). Raf-CAAX only partly restores ERK phosphorylation in thymocytes from CsA-treated mice (Supplementary Fig. 6e). The fact that the *Raf-CAAX* transgene does not rescue the double-negative to double-positive developmental block and does not restore appropriate timing or intensity of ERK signalling, together with additional ERK-independent Cnb1/NFAT targets (such as the transcription factor TOX) (ref. 15), might explain the incomplete rescue in *Cnb1*-deficient Raf-CAAX mice.

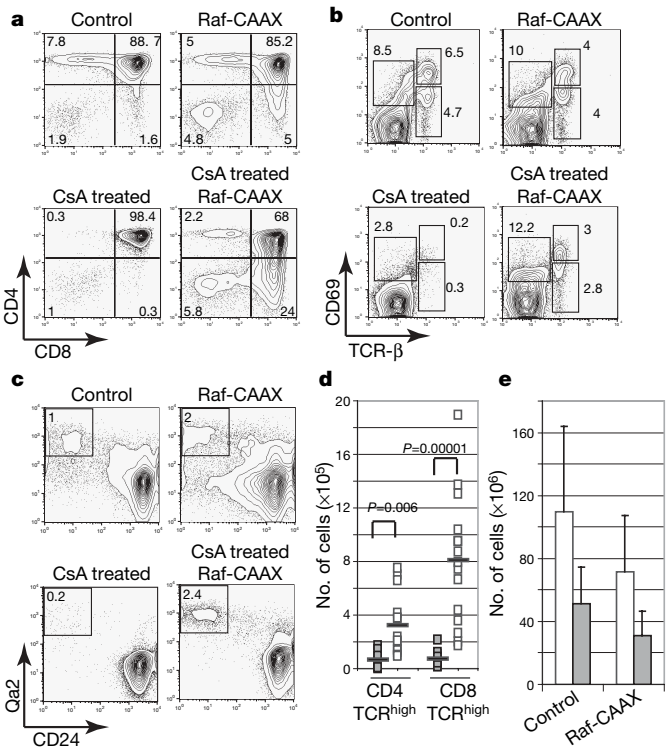


Figure 3 | Reconstitution of Raf–MEK–ERK signalling partly rescues positive selection in the absence of calcineurin activity. **a**, Expression of CD4 and CD8 in Raf-CAAX transgenic and control mice treated with CsA or left untreated. **b**, Analysis of CD69 and TCR-β expression on thymocytes from Raf-CAAX transgenic and control mice treated with CsA or left untreated. **c**, Analysis of CD24 and Qa2 expression on thymocytes from Raf-CAAX transgenic and control mice treated or not with CsA. **d**, Absolute numbers of TCR^{high} CD4 and CD8 single-positive cells ($n \geq 8$, each square represents an individual mouse; bar equals mean value). *P* values refer to a one-tailed *t*-test. Open symbols, CsA, Raf-CAAX; filled symbols, CsA. **e**, Absolute numbers of thymocytes for mice of indicated genotype ($n \geq 8$; error bars show s.d.). Open bars, control; filled bars, CsA-treated. The numbers in the corners of the panels represent the percentage of cells in each quadrant.

We proposed that this developmental window of ERK hypersensitivity might transiently increase the dynamic range (or bandwidth) of TCR signals, thereby enabling thymocytes to respond to 'weak' positively selecting ligands. To test this, we used the OT-I TAP (transporter associated with antigen processing)-null mice mouse model[16,17] (OT-I TAP^0). We compared the ability of double-positive thymocytes from control and CsA-treated OT-I TAP^0 mice to respond to the negatively selecting peptide SIINFEKL and the positively selecting peptide RTYTYEKL (ref. 17). OT-I-positive double-positive thymocytes from control mice responded to both peptides by phosphorylating ERK1/2, inducing Egr1 and upregulating CD69 (Fig. 4a–c and Supplementary Fig. 7). The response to the positively selecting peptide is weaker and delayed, as has been previously reported[18,19]. In contrast, double-positive thymocytes from CsA-treated mice were able to respond partly to the negatively selecting peptide SIINFEKL but failed to respond to the positively selecting peptide RTYTYEKL (Fig. 4a–c and Supplementary Fig. 7). These data suggested that calcineurin-dependent ERK sensitization is required for a response to weaker positively selecting ligands, whereas stronger negative selecting signals are able to activate the ERK pathway to a certain extent even when the transition to the 'ERK high competence' state has not occurred.

If the role of calcineurin were to increase the 'signalling bandwidth' and allow effective discrimination of graded signals, one would predict that *Cnb1*-deficient thymocytes that were prevented from dying should be positively selected in response to stronger signals, which would normally induce cell death. To test this, we used mice deficient in *Bim*, which is necessary for negative selection[20]. In control mice, *in vivo* treatment with CsA completely blocked the development of single-positive thymocytes. However, differentiation occurred in *Bim*-deficient mice even in the absence of calcineurin activity as assessed by the development of TCR-β^{high} CD4 single-positive and CD8 single-positive cells and upregulation of TCR-β, CD69 and CCR7 on double-positive cells (Fig. 4d and Supplementary Fig. 8a–c). The double-negative to double-positive developmental block that is observed in CsA-treated and *Cnb1*-deficient thymocytes[3], which results in an overall decrease in thymus cellularity in these mice, was not rescued by Bim deficiency (Supplementary Fig. 8c). Analogous results were obtained when *Bim*-deficient mice were crossed with conditional *Cnb1*-deficient mice (Supplementary Fig. 8d, e). Because Bim deficiency did not rescue development of the 'ERK high competence' population itself (Supplementary Fig. 8f), we postulated that in *Bim*-deficient mice 'strong' negatively selecting ligands could trigger the activation of ERK in thymocytes that are in the 'ERK low competence' state. According to this hypothesis, in *Bim*-deficient mice negatively selecting ligands could circumvent the need for the calcineurin-dependent sensitization and could induce the differentiation of single-positive thymocytes that in normal circumstances are negatively selected. Indeed, analysis of *Bim*-deficient mice in a background (Balb/c) that expresses the MHC class II molecule I-E and allows the superantigen-mediated deletion of T cells[2] revealed that in CsA-treated *Bim*-deficient mice a higher percentage of CD4 single-positive thymocytes expressed TCR Vβ chains that are reactive to endogenous superantigens in comparison with controls,

©2007 Nature Publishing Group

suggesting that single-positive thymocytes that develop in these mice are in fact those that received negatively selecting signals (Fig. 4e).

As a second approach to testing whether the role of ERK competence is to expand the signalling bandwidth, allowing the effective discrimination of weak signals, we examined thymocyte selection in the HY TCR transgenic mouse model. In this model most thymocytes express the HY TCR and are negatively selected by antigens expressed in male but not female mice[21]. As expected, the deletion of HY thymocytes in male mice (Supplementary Fig. 9a) was impaired by *Bim* deficiency[20]. Because CsA treatment resulted in the accumulation of immature CD8 single-positive cells (Supplementary Fig. 9a), we evaluated positive selection by monitoring CD69 upregulation and the development of $CD24^{low}$ and $Qa2^{high}$ CD8 single-positive cells. Treatment with CsA impaired CD69 upregulation and the development of mature CD8 single-positive thymocytes in control and *Bim*-deficient female mice, whereas it did not block the development of mature CD8 single-positive cells in male mice (Fig. 4f and Supplementary Fig. 9b, c). Similar results were obtained in the HY^{CD4} model, in which TCR-αβ expression is properly timed[22] (data not shown). These results suggested that the amount of ERK signalling provided by 'strong', negatively selecting signals in CsA-treated mice is able to induce differentiation and that negatively selecting signals do not require the development of 'ERK competence'.

Cell fate determination often occurs within morphogenic gradients that produce different cell fates at different points in the gradient, apparently as a result of signals of different intensities. A similar analogue-to-digital switch occurs in T-cell development (Supplementary Fig. 1), where signal intensity determines the outcome of TCR–MHC interactions. Our studies indicate that an early calcineurin–NFAT signal sensitizes the Raf–MEK–ERK pathway, allowing responses to weak TCR signals that would otherwise not be detected (Supplementary Fig. 1). In the absence of this calcineurin-dependent preconditioning, the signal intensity needed for positive selection overlaps with that needed for negative selection,

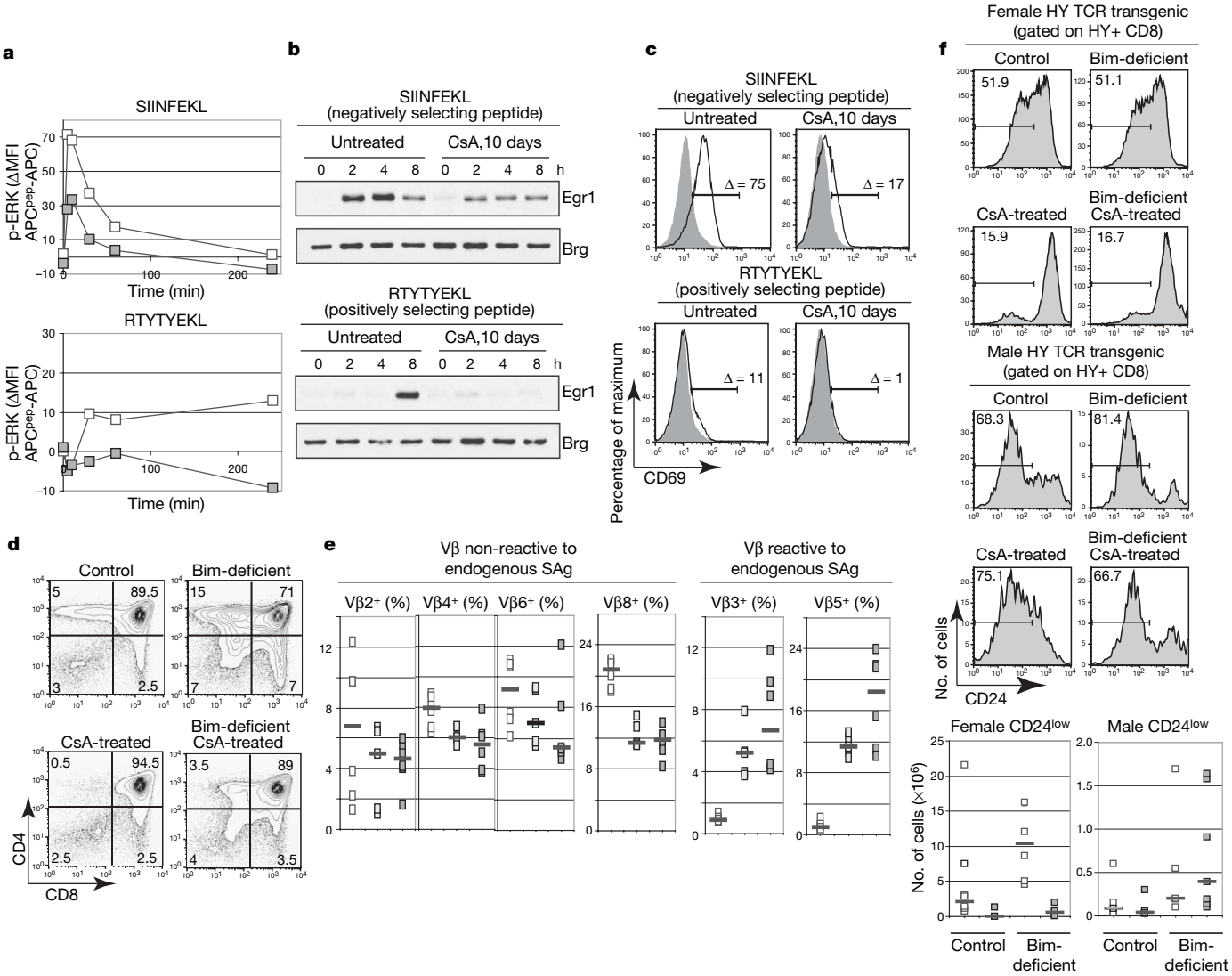


Figure 4 | Transition to the 'high ERK competence' state is required to respond functionally to positively selecting ligands. **a**, Phospho-ERK1/2 in OT-I double-positive thymocytes from mice stimulated with SIINFEKL or RTYTYEKL. White symbols, untreated; grey symbols, CsA-treated. **b**, Egr1 upregulation in OT-I double-positive thymocytes stimulated with SIINFEKL or RTYTYEKL. **c**, CD69 upregulation in OT-I double-positive thymocytes after 4 h of stimulation. Δ indicates the percentage difference in CD69 thymocytes stimulated by APCs only or by peptide-pulsed APCs. Solid lines, APCs plus peptide; grey areas, APCs only. **d**, CD4 and CD8 expression in thymocytes from mice of the indicated genotypes. **e**, Percentage of CD4 single-positive expressing Vβ chains reactive or not to endogenous superantigens (squares show results for individual mice; bars show medians; $n \geq 3$). White symbols, untreated control; light grey symbols, *Bim*-deficient; dark grey symbols, *Bim*-deficient and CsA-treated. **f**, Percentages (numbers shown in each panel refer to the percentage of cells in the indicated interval) and absolute numbers of HY^{+} CD8 single-positive $CD24^{low}$ cells in mice of the indicated genotypes (squares show results for individual mice, bars show means; $n \geq 4$). White symbols, untreated; grey symbols, CsA-treated.

©2007 Nature Publishing Group

and effective discrimination of graded developmental signals cannot occur.

METHODS SUMMARY

Mice. Mice were maintained in the animal facility of Stanford University in accordance with federal and institutional guidelines and were used between 4 and 12 weeks of age unless otherwise noted.

Cell isolation and flow cytometry. *Cnb1*-deficient and control double-positive thymocytes were routinely purified by anti-CD8α magnetic beads (Miltenyi Biotech) (purity more than 90%). Antibodies against CD markers, Q2a and CCR7 were from eBioscience. Phospho-ERK (clone 197G2) antibody was purchased from Cell Signalling and staining was performed in accordance with the manufacturer's recommendations.

BrdU labelling. Mice were injected intraperitoneally with 1 mg of BrdU (BD Pharmingen) and killed at the indicated time points. BrdU staining was performed in accordance with the manufacturer's instructions.

Biochemical analysis. For all biochemical analyses, *Cnb1*-deficient and control thymocytes were processed, stimulated and analysed as described in ref. 23. For stimulations of OT-I TAP^0 transgenic mice, CH27 cells transduced with H-$2K^b$ (CH27-H-2 K^b; provided by M. Davis's laboratory[24]) were used as antigen-presenting cells (APCs). Stimulation was performed as described[25]. For phospho-ERK staining, T2-H-2 K^b cells (ref. 26) (provided by M. Davis's laboratory) were used as APCs.

Microarray. Total RNA was harvested from sorted thymocytes with Trizol, followed by purification and on-column DNase digestion with an RNeasy mini kit (Qiagen) in accordance with the manufacturer's recommendations. Total RNA (200 ng) was used to generate biotinylated RNA with two rounds of amplification in accordance with Affymetrix's recommendation. All amplifications, antisense RNA labelling, hybridizations, washings, and data collection were performed by the Affymetrix Core Facility at the Stanford Protein and Nucleic Acid Facility. Data were analysed with GENESIS software[27] and GCOS and DMT software (Affymetrix) in accordance with the manufacturer's recommendations, leaving default settings unchanged.

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

Received 27 August; accepted 26 September 2007.

1. Palmer, E. Negative selection–clearing out the bad apples from the T-cell repertoire. *Nature Rev. Immunol.* **3,** 383–391 (2003).
2. Starr, T. K., Jameson, S. C. & Hogquist, K. A. Positive and negative selection of T cells. *Annu. Rev. Immunol.* **21,** 139–176 (2003).
3. Neilson, J. R., Winslow, M. M., Hur, E. M. & Crabtree, G. R. Calcineurin B1 is essential for positive but not negative selection during thymocyte development. *Immunity* **20,** 255–266 (2004).
4. Cante-Barrett, K., Winslow, M. M. & Crabtree, G. R. Selective role of NFATc3 in positive selection of thymocytes. *J. Immunol.* **179,** 103–110 (2007).
5. Alberola-Ila, J., Forbush, K. A., Seger, R., Krebs, E. G. & Perlmutter, R. M. Selective requirement for MAP kinase activation in thymocyte differentiation. *Nature* **373,** 620–623 (1995).
6. Costello, P. S., Nicolas, R. H., Watanabe, Y., Rosewell, I. & Treisman, R. Ternary complex factor SAP-1 is required for Erk-mediated thymocyte positive selection. *Nature Immunol.* **5,** 289–298 (2004).
7. Fischer, A. M., Katayama, C. D., Pages, G., Pouyssegur, J. & Hedrick, S. M. The role of erk1 and erk2 in multiple stages of T cell development. *Immunity* **23,** 431–443 (2005).
8. Shortman, K., Vremec, D. & Egerton, M. The kinetics of T cell antigen receptor expression by subgroups of $CD4^+8^+$ thymocytes: delineation of $CD4^+8^+3(2+)$ thymocytes as post-selection intermediates leading to mature T cells. *J. Exp. Med.* **173,** 323–332 (1991).
9. Fehling, H. J., Krotkova, A., Saint-Ruf, C. & von Boehmer, H. Crucial role of the pre-T-cell receptor α gene in development of αβ but not γδ T cells. *Nature* **375,** 795–798 (1995).
10. Levelt, C. N., Carsetti, R. & Eichmann, K. Regulation of thymocyte development through CD3. II. Expression of T cell receptor β CD3 epsilon and maturation to the $CD4^+8^+$ stage are highly correlated in individual thymocytes. *J. Exp. Med.* **178,** 1867–1875 (1993).
11. Penit, C. *In vivo* thymocyte maturation. BUdR labeling of cycling thymocytes and phenotypic analysis of their progeny support the single lineage model. *J. Immunol.* **137,** 2115–2121 (1986).
12. Macian, F. *et al.* Transcriptional mechanisms underlying lymphocyte tolerance. *Cell* **109,** 719–731 (2002).
13. Gallo, E. M., Cante-Barrett, K. & Crabtree, G. R. Lymphocyte calcium signaling from membrane to nucleus. *Nature Immunol.* **7,** 25–32 (2006).
14. Iritani, B. M., Alberola-Ila, J., Forbush, K. A. & Perimutter, R. M. Distinct signals mediate maturation and allelic exclusion in lymphocyte progenitors. *Immunity* **10,** 713–722 (1999).
15. Aliahmad, P. *et al.* TOX provides a link between calcineurin activation and CD8 lineage commitment. *J. Exp. Med.* **199,** 1089–1099 (2004).
16. Hogquist, K. A. *et al.* T cell receptor antagonist peptides induce positive selection. *Cell* **76,** 17–27 (1994).
17. Hogquist, K. A. *et al.* Identification of a naturally occurring ligand for thymic positive selection. *Immunity* **6,** 389–399 (1997).
18. Werlen, G., Hausmann, B. & Palmer, E. A motif in the αβ T-cell receptor controls positive selection by modulating ERK activity. *Nature* **406,** 422–426 (2000).
19. McNeil, L. K., Starr, T. K. & Hogquist, K. A. A requirement for sustained ERK signaling during thymocyte positive selection *in vivo*. *Proc. Natl Acad. Sci. USA* **102,** 13574–13579 (2005).
20. Bouillet, P. *et al.* BH3-only Bcl-2 family member Bim is required for apoptosis of autoreactive thymocytes. *Nature* **415,** 922–926 (2002).
21. Kisielow, P., Bluthmann, H., Staerz, U. D., Steinmetz, M. & von Boehmer, H. Tolerance in T-cell-receptor transgenic mice involves deletion of nonmature $CD4^+8^+$ thymocytes. *Nature* **333,** 742–746 (1988).
22. Baldwin, T. A., Sandau, M. M., Jameson, S. C. & Hogquist, K. A. The timing of TCR α expression critically influences T cell development and selection. *J. Exp. Med.* **202,** 111–121 (2005).
23. Reynolds, L. F. *et al.* Vav1 transduces T cell receptor signals to the activation of the Ras/ERK pathway via LAT, Sos, and RasGRP1. *J. Biol. Chem.* **279,** 18239–18246 (2004).
24. Purbhoo, M. A., Irvine, D. J., Huppa, J. B. & Davis, M. M. T cell killing does not require the formation of a stable mature immunological synapse. *Nature Immunol.* **5,** 524–530 (2004).
25. Cante-Barrett, K., Gallo, E. M., Winslow, M. M. & Crabtree, G. R. Thymocyte negative selection is mediated by protein kinase C- and Ca^{2+}-dependent transcriptional induction of bim of cell death. *J. Immunol.* **176,** 2299–2306 (2006).
26. Wei, M. L. & Cresswell, P. HLA-A2 molecules in an antigen-processing mutant cell contain signal sequence-derived peptides. *Nature* **356,** 443–446 (1992).
27. Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics* **18,** 207–208 (2002).

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

Acknowledgements We thank P. Ebert for helping with calcium flux studies, and K. A. Hogquist and C. Wilson for providing mice and reagents. E.M.G., M.M.W. and A.N.R. were supported by Stanford Graduate Fellowships. M.M.W. was additionally supported by a Howard Hughes Medical Institute predoctoral fellowship. A.N.R. was also supported by a National Science Foundation Graduate Research Fellowship. K.C.B. was supported by the Boehringer Ingelheim Fonds. L.H. was also supported by Agency for Science, Technology and Research Singapore. This work was supported by grants from Howard Hughes Medical Institute and the National Institute of Heath to G.R.C.

Author Contributions E.M.G., M.M.W. and G.R.C. generated the hypotheses, designed the experiments and wrote the manuscript. E.M.G. performed the experiments and generated the figures. K.C.-B. generated the *NFATc3* conditional knockout mice, maintained this line in the *NFATc2*-null background and contributed to the experiments in Fig. 4. A.N.R. and L.H. contributed to pilot experiments and experiments shown in Fig. 2 and Supplementary Fig. 8. J.R.N. generated the *Cnb1* conditional knockout mice, conducted pilot experiments and contributed to experimental rationale. L.M. contributed to experiments shown in Supplementary Fig. 5. B.I. provided the *Raf-CAAX* transgenic mice.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.R.C. (crabtree@stanford.edu).

©2007 Nature Publishing Group

METHODS

Mice. *Cnb1* conditional knockout and Raf-CAAX mice have been previously characterized[3,14]. *Lck-cre* mice were a gift from C. Wilson. Because no consistent differences were observed among mice of the genotypes $Cnb1^{+/+}$-lck^{Cre};$Cnb1^{fl/+}$-lck^{Cre}, $Cnb1^{\Delta/+}$-lck^{Cre};$Cnb1^{fl/f}$ or $Cnb1^{fl/\Delta}$, they are collectively referred to as 'control mice' throughout this manuscript. In addition, no differences were observed between $Cnb1^{fl/f}$-lck^{Cre} and $Cnb1^{fl/\Delta}$-lck^{Cre} mice, and these animals were used interchangeably for the experiments described in this study and compared with age-matched control mice. OT-I transgenic mice[16] on the TAP^0 background were a gift from K. A. Hogquist. C57/B6 $H2\text{-}A^b/\beta 2m$ double knockout mice, C57BL/6 and HY TCR transgenic mice were purchased from Taconic. *Bim*-deficient mice were purchased from the Jackson Laboratory. For *in vivo* treatment with CsA, mice were treated with daily intraperitoneal injections of CsA ($30\ mg\ kg^{-1}\ d^{-1}$) for ten or more days. Mice in which exon 3 of the *NFATc3* gene is flanked by loxP sites ($NFATc3^{fl/f}$) were generated in our laboratory[4] and crossed to *Lck-cre* mice and to *NFATc2*-deficient mice[28] to obtain *NFATc2/NFATc3*-deficient thymocytes.

Phospho-ERK intracellular staining. The staining for phospho-ERK was performed as indicated in the staining protocol provided by Cell Signalling. For co-staining with anti-BrdU and anti-phospho-ERK antibodies, the staining was performed in accordance with instructions from Cell Signalling for the anti-phospho-ERK antibody with the following modifications: first, DNase treatment was performed in accordance with the BD Pharmingen protocol after the blocking step, and second, anti-BrdU antibody was added at the same time as the anti-phospho-ERK antibody. For phospho-ERK staining in OT-I TAP^0 thymocytes, T2-H-2 K^b APCs were pulsed overnight at a concentration of $10^6\ ml^{-1}$ with 2 μM SIINFEKL and 100 μM RTYTYEKL in RPMI medium. APCs were washed twice in PBS and then incubated with thymocytes at a 1:1 ratio at 37 °C for the indicated duration. Reaction was started with a 2-min centrifugation at 300*g*. Conjugates were disrupted by the addition of ice-cold PBS, 10 mM EDTA, 2% paraformaldehyde and by vigorous pipetting. Cells were then left on ice for 5 min and then incubated at 25 °C for 15 min. Staining was performed in accordance with instructions from Cell Signalling, except that the incubation in ice-cold methanol was followed by an additional incubation overnight in methanol at −20 °C.

Biochemical analysis. SDS–PAGE and immunoblotting were conducted with standard procedures. All phospho-specific antibodies and total protein antibody used for immunoblotting were purchased from Cell Signalling with the exception of anti-B-Raf and anti-Egr1 (Santa Cruz Biotechnology) and anti-ERK1/2 (Upstate Biotechnology).

Microarray. The indicated thymocyte populations were sorted on ARIA and the purity was assessed to be at least 98% by reanalysis. Three independent RNA samples from thymocytes sorted from individual mice were analysed for each experimental group. Probe sets were first filtered with DMT software (Affymetrix) by eliminating those that did not have a 'present call' in all control samples (for increased calls) or all experimental samples (for decreased calls). Nine pairwise comparisons of the three experimental versus three control samples were performed with DMT software. To be considered significant, probe sets had to receive a 'change call' in 100% of comparisons, had to have an absolute log ratio of 1 or more, and had to be considered significantly changed by one-way analysis of variance (ANOVA) ($P \leq 0.01$). One-way ANOVA was run with GENESIS software[27] on all probe sets, after filtering for absent calls. Probe sets were annotated by submitting them to the Affymetrix analysis website (www.affymetrix.com/analysis/index.affx).

28. Xanthoudakis, S. *et al.* An enhanced immune response in mice lacking the transcription factor NFAT1. *Science* **272**, 892–895 (1996).

©2007 Nature Publishing Group

LETTERS

mTOR controls mitochondrial oxidative function through a YY1–PGC-1α transcriptional complex

John T. Cunningham[1,2], Joseph T. Rodgers[1], Daniel H. Arlow[3], Francisca Vazquez[1], Vamsi K. Mootha[3] & Pere Puigserver[1,2]

Transcriptional complexes that contain peroxisome-proliferator-activated receptor coactivator (PGC)-1α control mitochondrial oxidative function to maintain energy homeostasis in response to nutrient and hormonal signals[1,2]. An important component in the energy and nutrient pathways is mammalian target of rapamycin (mTOR), a kinase that regulates cell growth, size and survival[3–5]. However, it is unknown whether and how mTOR controls mitochondrial oxidative activities. Here we show that mTOR is necessary for the maintenance of mitochondrial oxidative function. In skeletal muscle tissues and cells, the mTOR inhibitor rapamycin decreased the gene expression of the mitochondrial transcriptional regulators PGC-1α, oestrogen-related receptor α and nuclear respiratory factors, resulting in a decrease in mitochondrial gene expression and oxygen consumption. Using computational genomics, we identified the transcription factor yin-yang 1 (YY1) as a common target of mTOR and PGC-1α. Knockdown of YY1 caused a significant decrease in mitochondrial gene expression and in respiration, and YY1 was required for rapamycin-dependent repression of those genes. Moreover, mTOR and raptor interacted with YY1, and inhibition of mTOR resulted in a failure of YY1 to interact with and be coactivated by PGC-1α. We have therefore identified a mechanism by which a nutrient sensor (mTOR) balances energy metabolism by means of the transcriptional control of mitochondrial oxidative function. These results have important implications for our understanding of how these pathways might be altered in metabolic diseases and cancer.

Growth factors including insulin induce phosphatidylinositol-3-OH kinase, which activates protein kinase B/Akt, leading to phosphorylation of the tuberous sclerosis (TSC)1–TSC2 complex. This in turn releases Rheb from its inhibition, which then activates the mTOR kinase[4]. mTOR also responds to nutrients through a less well characterized mechanism through Rheb and the class III PI(3)K Vps34 (refs 6–8). mTOR controls the transcription of many genes[9] and positively regulates mitochondrial activity[10]; however, the mechanisms mediating this response remain unknown. PGC-1α controls mitochondrial function through interaction with transcription factors such as oestrogen-related receptor α (ERR-α) and nuclear respiratory factors (NRFs)[11]. Decreases in PGC-1α and PGC-1β and downregulation of oxidative phosphorylation target genes in skeletal muscle have been associated with insulin resistance and type 2 diabetes[12,13].

Taking these results together, we proposed that mTOR might influence nutrient oxidative rates to maintain energy levels in mammalian cells by controlling mitochondrial transcriptional regulators. Figure 1a and Supplementary Fig. 1 show that treatment of C2C12 myotubes with rapamycin (a specific mTOR inhibitor) decreased the messenger RNAs encoding PGC-1α, ERR-α, NRF-1 and Gabpa/b. Mitochondrial gene targets of PGC-1α involved in oxidative phosphorylation, the tricarboxylic acid cycle and uncoupling respiration were also downregulated by rapamycin (Fig. 1b and Supplementary Fig. 1). These effects on gene expression translated into a 32% decrease in mitochondrial DNA content (Fig. 1d) and a 12% decrease in oxygen consumption by rapamycin (Fig. 1e). Similar results were also observed in primary skeletal muscle cells treated with rapamycin (Supplementary Fig. 2). To demonstrate further that mTOR controls mitochondrial genes, we used $TSC2^{-/-}$ cells, in which mTOR is constitutively active[14]. As shown in Fig. 1a–c and Supplementary Fig. 1, $TSC2^{-/-}$ cells showed an elevated expression of mitochondrial genes that were downregulated by rapamycin. Moreover, knockdown of components of the mTORC1 complex, mTOR and raptor, but not S6K1, resulted in decreased expression of mitochondrial genes, and the cells became insensitive to rapamycin (Fig. 1f, 1g, and Supplementary Fig. 3). Mitochondrial DNA content (Fig. 1d) and oxygen consumption (Fig. 1e) followed the same pattern. Next, we investigated whether the effects of rapamycin were also observed in mice. Skeletal muscle from mice treated with rapamycin showed a decrease in mitochondrial gene expression and oxygen consumption (Fig. 1h, i). Mice treated with rapamycin developed symptoms of diabetes, with hyperglycaemia, glucose intolerance, higher levels of free fatty acids in skeletal muscle, hypertriglyceraemia and hypercholesterolaemia (Supplementary Fig. 4). Further analysis of the metabolic effects of mTOR in cells treated with rapamycin revealed a decrease in lactate production. Despite lower rates of oxygen consumption and glycolysis, ATP levels were preserved (Supplementary Fig. 5). Because rapamycin can lead to changes in Akt activity through feedback mechanisms[15,16], we determined whether Akt mediated the effects on mitochondrial function. Whereas mitochondrial gene expression and oxygen consumption were decreased after 12 hours of treatment with rapamycin, phospho-Akt was either increased in $TSC2^{-/-}$ cells or unchanged in C2C12 cells. Furthermore, whereas rapamycin decreased expression of the genes encoding PGC-1α and cytochrome *c*, an Akt inhibitor did not (Supplementary Fig. 6). Taken together, these results support the finding that mTOR positively controls mitochondrial gene expression and oxygen consumption in an Akt-independent manner.

To systematically identify the transcription factors downstream of mTOR that are specifically involved in regulating mitochondrial function, we used a computational genomics approach. We performed microarray analysis of $TSC2^{-/-}$ cells treated with rapamycin (Supplementary Table 1; rapamycin versus vehicle). Using SAM[17], we identified 1,907 differentially expressed genes at an estimated false

[1]Dana-Farber Cancer Institute and Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [3]Departments of Systems Biology and Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA and Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts 02139, USA.

©2007 Nature Publishing Group

discovery rate (FDR) of 6.3%. Roughly twice as many transcripts were downregulated as were upregulated, and the downregulated transcripts were highly enriched (12.9% of downregulated transcripts versus 3.7% overall, $P < 8 \times 10^{-45}$) for transcripts predicted to encode mitochondrial proteins[18]. Because PGC-1α is known to regulate the expression of many mitochondrial genes, we also profiled *TSC2*$^{-/-}$ cells overexpressing *PGC-1α* (Supplementary Table 2). As seen in Fig. 2a, many of the mitochondrial genes induced by PGC-1α are suppressed by rapamycin; whereas 27% of all genes are in the upper left quadrant, 50% of mitochondrial genes[18] are found in the same quadrant ($\chi^2 = 463$, $P < 9 \times 10^{-103}$). Together, these microarray analyses demonstrate that treatment with rapamycin suppresses the expression of many mitochondrial genes that are induced by PGC-1α. We next used motifADE[11] to identify *cis*-regulatory elements that might mediate the transcriptional response to rapamycin. Among all possible hexamer, heptamer and octamer motifs, the hexamer ATGGCG was the highest scoring element (adjusted $P < 6 \times 10^{-9}$) (Fig. 2b). This hexameric sequence is known to be bound by the transcription factor YY1 (ref. 19). The identified YY1 motif was enriched upstream of predicted mitochondrial genes[18] (24% versus 15%, $P < 3 \times 10^{-11}$) and highly enriched upstream of oxidative phosphorylation genes (36% versus 15%, $P < 5 \times 10^{-5}$). Because PGC-1α upregulates mitochondrial genes and rapamycin downregulates a large set of these genes (Fig. 2a), we used motifADE to test whether genes with the YY1-binding site are induced during PGC-1α-mediated mitochondrial biogenesis. We found that the hexameric motif was highly significant in the *TSC2*$^{-/-}$ cells ($P < 2 \times 10^{-7}$) and in C2C12 myotubes ($P < 4 \times 10^{-6}$) overexpressing *PGC-1α*.

YY1-binding motifs were highly enriched in mitochondrial genes regulated by rapamycin and PGC-1α, suggesting that YY1 might control their expression. Figure 3 shows that knockdown of YY1 decreased the expression of *PGC-1α* and *PGC-1β*, mitochondrial genes, mitochondrial DNA and oxygen consumption. Conversely, ectopic expression of YY1 increased the expression of mitochondrial genes (Fig. 3e). Rapamycin-dependent downregulation of those genes was largely abolished by knockdown of YY1. This indicates that YY1 is required for conferring rapamycin-dependent downregulation.

To gain insights into the mechanisms by which mTOR controls the YY1–PGC-1α function, we performed chromatin immunoprecipitation (ChIP) analysis and found that YY1 is recruited to promoter regions of the genes encoding PGC-1α and cytochrome *c* (Fig. 4a). Moreover, overexpression of YY1 increased, but knockdown of YY1 decreased, the activity of both promoters (Fig. 4b). Because PGC-1α

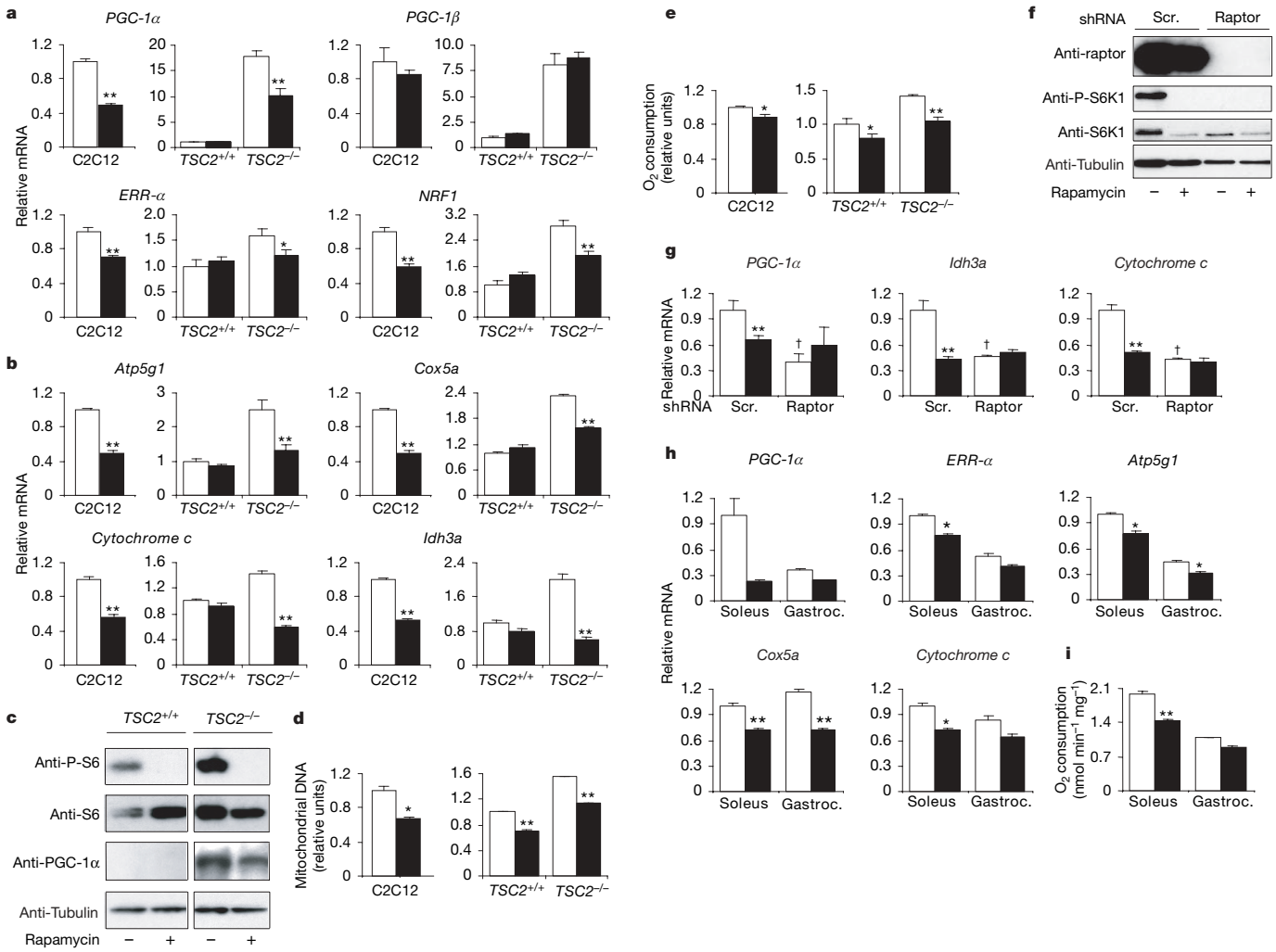
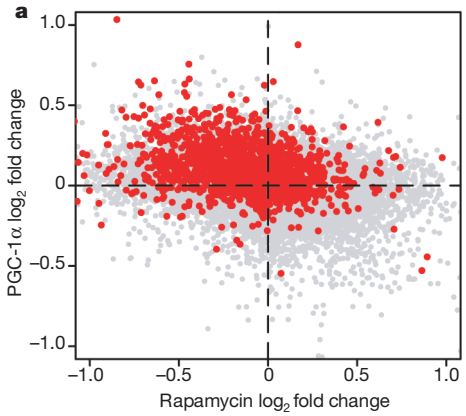


Figure 1 | mTOR controls mitochondrial gene expression and oxygen consumption. **a**, **b**, Transcriptional regulators (**a**) and mitochondrial genes (**b**) are regulated by mTOR. Open columns, vehicle; filled columns, rapamycin treatment. **c**–**e**, RNA and protein (**c**), mitochondrial DNA (**d**) and oxygen consumption (**e**) are decreased by rapamycin. **f**, **g**, Raptor knockdown decreases PGC-1α and mitochondrial genes. Lentiviral shRNAs were used to infect *TSC2*$^{-/-}$ cells. Scr., scrambled. **f**, Western blot analysis of indicated proteins. **g**, Quantitative real-time PCR analysis of indicated genes. **h**, **i**, Rapamycin decreases mitochondrial genes (**h**) and oxygen consumption (**i**) in mouse skeletal muscle. Gastroc., gastrocnemius. Measurements are described in Methods. Error bars represent s.e.m.; $n = 6$; asterisk, $P < 0.05$; two asterisks, $P < 0.01$ for vehicle versus rapamycin treatment; dagger, $P < 0.05$ scrambled versus raptor shRNA.

©2007 Nature Publishing Group



b

Motif	Frequency	Adj. *P* value	Annotation
ATGGCG	0.136	5.14×10^{-9}	YY1
CGGAAG	0.188	4.11×10^{-8}	ETS
GGCGGC	0.397	1.39×10^{-7}	YY1
GGCCGG	0.366	2.10×10^{-7}	
CGGGCG	0.294	3.11×10^{-7}	
TGGCGG	0.273	3.67×10^{-7}	YY1
CCGGAAG	0.076	1.65×10^{-6}	ETS
GCGCGC	0.292	1.81×10^{-6}	
CGCGCC	0.289	2.57×10^{-6}	
CAGGCG	0.260	4.82×10^{-6}	

Figure 2 | Genomic analysis reveals that mitochondrial genes are regulated by PGC-1α and mTOR pathways by means of the transcription factor YY1. **a**, Relationship between the change in expression of all genes in *TSC2*$^{-/-}$ cells treated with rapamycin (*x* axis) and overexpressing PGC-1α (*y* axis). Mitochondrial genes[18] are shown in red; 50% of mitochondrial genes are found in the upper left quadrant, whereas only 27% of all genes are found in the quadrant ($P < 9 \times 10^{-103}$). **b**, MotifADE output showing the top ten most significant motifs for rapamycin versus vehicle in *TSC2*$^{-/-}$ cells. The YY1-binding site is the highest-scoring motif in cells treated with rapamycin. Adj., adjusted.

regulates both genes[11], we analysed whether YY1 was coactivated by PGC-1α. As shown in Fig. 4c, PGC-1α increased the transcriptional activity of GAL4-YY1. PGC-1α coactivation of YY1, but not of ERR-α, was repressed by rapamycin. These results indicate that YY1 binds directly to mitochondrial gene promoters and that PGC-1α functions as a transcriptional coactivator for YY1 in an mTOR-dependent manner. To test how mTOR might control YY1, we performed immunoprecipitation assays. As shown in Fig. 4d, mTOR is present in the YY1 protein complex. This interaction is direct and requires both the YY1 amino and carboxy termini (Supplementary Fig. 7b). Raptor (a component of mTORC1), but not rictor (a component of mTORC2) also interacted with YY1 and was bound to promoters of the genes encoding PGC-1α and cytochrome *c* (Fig. 4a, e). In addition, mapping studies indicated that raptor binds to the YY1 region that lies between amino-acid residues 203 and 235, which has been shown to be necessary for co-repressor recruitment (Supplementary Fig. 7e)[20]. In binding assays *in vitro*, glutathione *S*-transferase (GST)–YY1 interacted strongly with full-length PGC-1α through the C-terminal domain. In addition, PGC-1α interacted with YY1 primarily through the third C-terminal zinc finger (Supplementary Fig. 7h). Finally, the fact that coactivation of YY1 by PGC-1α was decreased by rapamycin while YY1 remained bound to the

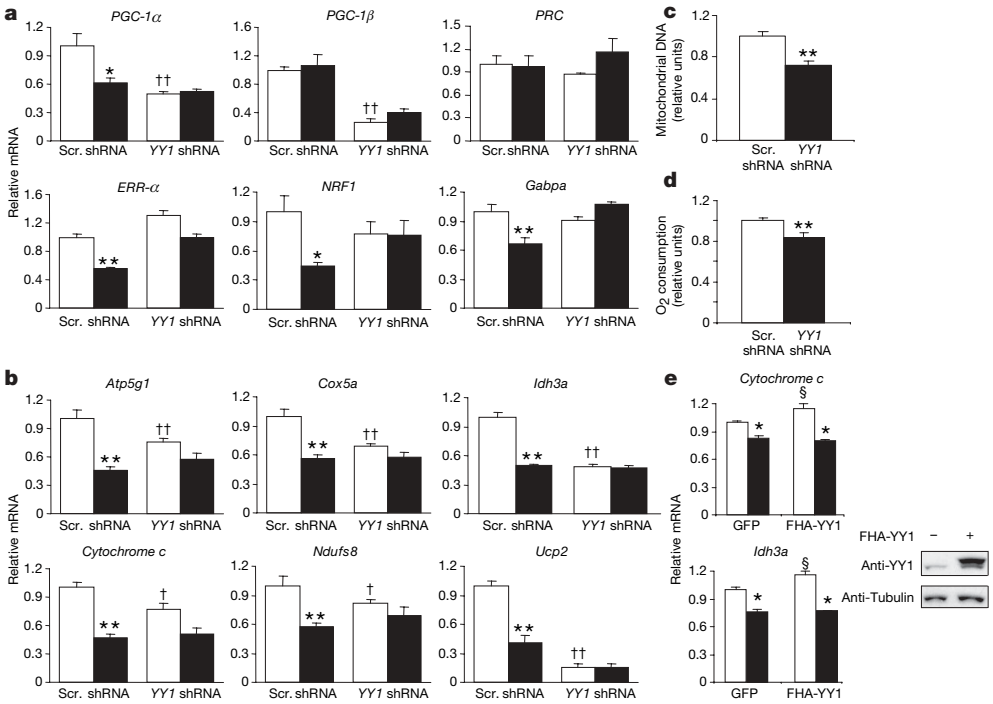


Figure 3 | YY1 regulates mitochondrial gene expression and oxygen consumption. **a**, PGC-1α and PGC-1β are controlled by YY1. C2C12 myotubes were infected with adenoviruses encoding scrambled (scr.) or *YY1* shRNA. RNA was quantified as described in Methods. Open columns, vehicle; filled columns, rapamycin treatment. PRC, PGC-1-related coactivator. **b–d**, Mitochondrial gene expression (**b**), mitochondrial DNA (**c**) and oxygen consumption (**d**) are decreased by *YY1* shRNA. In **b**, open columns, vehicle; filled columns, rapamycin treatment. **e**, Ectopic expression of *YY1* increased mitochondrial genes. HEK-293 cells were infected with lentiviruses encoding YY1 as described in Methods. FHA, Flag–HA epitope tag. Open columns, vehicle; filled columns, rapamycin treatment. Error bars represent s.e.m.; $n = 6$; asterisk, $P < 0.05$; two asterisks, $P < 0.01$ for vehicle versus rapamycin; dagger, $P < 0.05$; two daggers, $P < 0.01$ for scrambled versus *YY1* shRNA; section sign, $P < 0.05$ for GFP versus FHA-YY1.

©2007 Nature Publishing Group

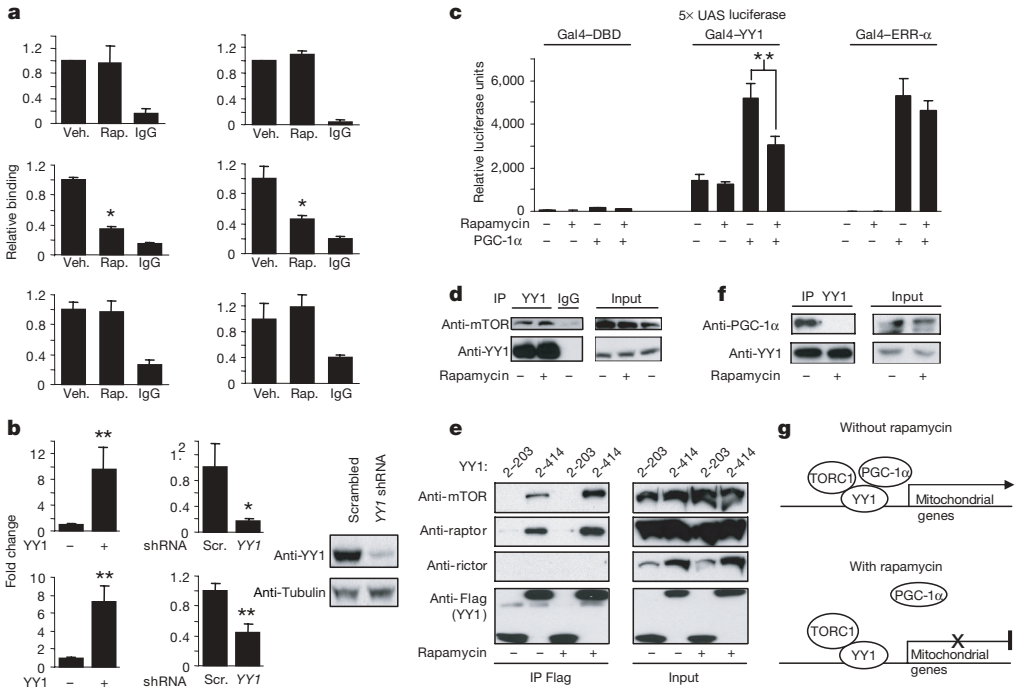


Figure 4 | Rapamycin-dependent coactivation and interaction between PGC-1α, YY1 and mTORC1. **a**, YY1, PGC-1α and raptor bind the promoters of the genes encoding PGC-1α (left panels) and cytochrome *c* (right panels). ChIP analysis with YY1 (top row), Flag–HA–PGC-1α (middle row) and raptor (bottom row) was performed on $TSC2^{-/-}$ cells. Bars represent s.e.m.; $n = 6$. **b**, YY1 activates the promoters of the genes encoding cytochrome *c* (top graphs) and PGC-1α (bottom graphs). Luciferase reporter assay. **c**, YY1–PGC-1α transcriptional activity is repressed by rapamycin. Error bars represent s.e.m.; $n = 8$; asterisk, $P < 0.05$; two asterisks, $P < 0.01$ for vehicle versus rapamycin treatment in **c**, and for versus vector controls in **b**. **d**, **e**, mTOR (**d**) and raptor (**e**) interact with YY1. IP, immunoprecipitation. **f**, Rapamycin-dependent interaction between YY1 and PGC-1α. YY1 immunoprecipitates were analysed by western blot. **g**, Model of how mTOR controls mitochondrial respiration.

promoters, and that mTOR and raptor interacted with YY1 independently of rapamycin, prompted us to test whether the interaction between YY1 and PGC-1α was the key regulated step that controlled YY1-dependent mitochondrial gene transcription in response to mTOR activity. Figure 4f and Supplementary Fig. 7j show that PGC-1α and YY1 interacted in cells; more significantly, the interaction between these two proteins was disrupted by rapamycin. ChIP analysis showed that PGC-1α is bound to promoters of the genes encoding cytochrome *c* and PGC-1α; however, rapamycin prevented this recruitment (Fig. 4a). These results indicate that mTOR regulates the transcriptional function of YY1–PGC-1α by directly altering their physical interaction.

Transcription factors such as ERR-α and Gabpa/b are important for PGC-1α-induced mitochondrial biogenesis[11,21]; however, mTOR did not directly regulate the transcriptional activity of PGC-1α on these factors. Instead, mTOR controlled mitochondrial gene expression through the direct modulation of YY1–PGC-1α activity. This regulation would allow the cell to connect nutrient pathways to activate mitochondrial function and ensure energy supply for cellular activities. PGC-1α and oxidative phosphorylation genes are modestly downregulated in skeletal muscle of humans with type 2 diabetes[12,13]. Our motif analysis also identified the YY1 motif as being significantly associated with this pattern of mitochondrial gene expression. In fact, the YY1 motif is also significantly ($P < 0.05$) associated with differential expression between diabetic and control patients[12]. However, it is not clear how mTOR is regulated in skeletal muscle in type 2 diabetes and insulin resistance. We show that mice treated with rapamcyin developed several symptoms of diabetes, which is consistent with diminished oxidative function in skeletal muscle. Treatment with rapamycin in humans receiving organ transplants and in rodents has been associated with increased levels of blood triacylglycerols and cholesterol, which might contribute to insulin resistance[22,23]. Our model (Fig. 4g) predicts that a decrease in mTOR activity would inhibit YY1–PGC-1α function, resulting in a decreased expression of mitochondrial genes that might be linked to increased circulating and intracellular lipids. Finally, the fact that the YY1–PGC-1α interaction can be directly modulated by mTOR activity opens new possibilities for potential pharmacological intervention to increase mitochondrial activity in metabolic diseases in which it is compromised.

METHODS SUMMARY

Experiments on mice. Male Balb/c mice 10 weeks old were daily injected intraperitoneally with 2.5 mg kg^{-1} rapamycin or vehicle. Mice were sacrificed after 11 days of injection and soleus and gastrocnemius muscles were dissected. Oxygen consumption was measured from saponin-permeabilized muscle slices using 5 mM glutamate and 2 mM malate as substrates, as described, with a Rank Brothers oxygen electrode[24,25].

Molecular biology. Cells were cultured in DMEM containing 10% cosmic calf serum. C2C12 myoblasts were differentiated in DMEM with 2% horse serum. YY1 short hairpin RNA (shRNA) was cloned into pSUPER by using the sequence 5′-GGGAGCAGAAGCAGGTGCAGAT-3′. Other expression and shRNA plasmids have been described previously or were generated by PCR cloning of described constructs (see Methods). RNA was isolated with Trizol, reverse transcribed using Superscript II Reverse Transcriptase with oligo(dT) primers and analysed by quantitative real-time (qRT)–PCR by SYBR green fluorescence. Mitochondrial DNA was quantified by determining the ratio of mitochondrial cyclo-oxygenase (Cox)2 to nuclear intron of β-globin by qRT–PCR of isolated DNA. ChIP was performed by using the Upstate protocol with the indicated antibodies. Luciferase assays were performed on lysates after transfection of cells with the indicated plasmids by using Polyfect reagent.

Microarray and *cis*-regulatory motif analysis. Expression profiles of $TSC2^{-/-}$ cells were obtained with Affymetrix MOE430v2 GeneChips. Expression summary values were calculated by using RMA[26] with default settings in one batch for the rapamycin/vehicle-treated cells and separately for the PGC-1α/green fluorescent protein (GFP)-transduced cells. A modified version of motifADE[11] was used to identify *cis* elements associated with differential expression between treatment groups. Promoter regions consisting of 2 kilobases of genomic sequence centred on the transcription start site were compiled, and a gene was recorded as having a particular motif only if both mouse and human promoters contained the motif, although the position need not be conserved. A Bonferroni

©2007 Nature Publishing Group

correction was applied when testing all *k*-mer motifs. The hypergeometric distribution was used to test whether the YY1 motif was enriched upstream of nuclear-encoded oxidative phosphorylation and mitochondrial genes[18].

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

Received 5 August; accepted 25 September 2007.

1. Finck, B. N. & Kelly, D. P. PGC-1 coactivators: inducible regulators of energy metabolism in health and disease. *J. Clin. Invest.* **116,** 615–622 (2006).
2. Lin, J., Handschin, C. & Spiegelman, B. M. Metabolic control through the PGC-1 family of transcription coactivators. *Cell Metab.* **1,** 361–370 (2005).
3. Wullschleger, S., Loewith, R. & Hall, M. N. TOR signaling in growth and metabolism. *Cell* **124,** 471–484 (2006).
4. Dann, S. G. & Thomas, G. The amino acid sensitive TOR pathway from yeast to mammals. *FEBS Lett.* **580,** 2821–2829 (2006).
5. Sarbassov, D. D., Ali, S. M. & Sabatini, D. M. Growing roles for the mTOR pathway. *Curr. Opin. Cell Biol.* **17,** 596–603 (2005).
6. Inoki, K., Li, Y., Xu, T. & Guan, K. L. Rheb GTPase is a direct target of TSC2 GAP activity and regulates mTOR signaling. *Genes Dev.* **17,** 1829–1834 (2003).
7. Long, X., Lin, Y., Ortiz-Vega, S., Yonezawa, K. & Avruch, J. Rheb binds and regulates the mTOR kinase. *Curr. Biol.* **15,** 702–713 (2005).
8. Nobukini, T. *et al.* Amino acids mediate mTOR/raptor signaling through activation of class 3 phosphatidylinositol 3OH-kinase. *Proc. Natl Acad. Sci. USA* **102,** 14238–14243 (2005).
9. Peng, T., Golub, T. R. & Sabatini, D. M. The immunosuppressant rapamycin mimics a starvation-like signal distinct from amino acid and glucose deprivation. *Mol. Cell. Biol.* **22,** 5575–5584 (2002).
10. Schieke, S. M. *et al.* The mammalian target of rapamycin (mTOR) pathway regulates mitochondrial oxygen consumption and oxidative capacity. *J. Biol. Chem.* **281,** 27643–27652 (2006).
11. Mootha, V. K. *et al.* Errα and Gabpa/b specify PGC-1α-dependent oxidative phosphorylation gene expression that is altered in diabetic muscle. *Proc. Natl Acad. Sci. USA* **101,** 6570–6575 (2004).
12. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34,** 267–273 (2003).
13. Patti, M. E. *et al.* Coordinated reduction of genes of oxidative metabolism in humans with insulin resistance and diabetes: Potential role of PGC1 and NRF1. *Proc. Natl Acad. Sci. USA* **100,** 8466–8471 (2003).
14. Zhang, H. *et al.* Loss of Tsc1/Tsc2 activates mTOR and disrupts PI3K–Akt signaling through downregulation of PDGFR. *J. Clin. Invest.* **112,** 1223–1233 (2003).
15. Shah, O. J., Wang, Z. & Hunter, T. Inappropriate activation of the TSC/Rheb/mTOR/S6K cassette induces IRS1/2 depletion, insulin resistance, and cell survival deficiencies. *Curr. Biol.* **14,** 1650–1656 (2004).
16. Sarbassov, D. D. *et al.* Prolonged rapamycin treatment inhibits mTORC2 assembly and Akt/PKB. *Mol. Cell* **22,** 159–168 (2006).
17. Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl Acad. Sci. USA* **98,** 5116–5121 (2001).
18. Calvo, S. *et al.* Systematic identification of human mitochondrial disease genes through integrative genomics. *Nature Genet.* **38,** 576–582 (2006).
19. Yant, S. R. *et al.* High affinity YY1 binding motifs: identification of two core types (ACAT and CCAT) and distribution of potential binding sites within the human beta globin cluster. *Nucleic Acids Res.* **23,** 4353–4362 (1995).
20. Wilkinson, F. H., Park, K. & Atchison, M. L. Polycomb recruitment to DNA *in vivo* by the YY1 REPO domain. *Proc. Natl Acad. Sci. USA* **103,** 19296–19301 (2006).
21. Schreiber, S. N. *et al.* The estrogen-related receptor α (ERRα) functions in PPARγ coactivator 1α (PGC-1α)-induced mitochondrial biogenesis. *Proc. Natl Acad. Sci. USA* **101,** 6472–6477 (2004).
22. Ribes, D., Kamar, N., Esposito, L. & Rostaing, L. Combined use of tacrolimus and sirolimus in *de novo* renal transplant patients: current data. *Transplant. Proc.* **37,** 2813–2816 (2005).
23. Morrisett, J. D. *et al.* Effects of sirolimus on plasma lipids, lipoprotein levels, and fatty acid metabolism in renal transplant patients. *J. Lipid Res.* **43,** 1170–1180 (2002).
24. Leone, T. C. *et al.* PGC-1α deficiency causes multi-system energy metabolic derangements: muscle dysfunction, abnormal weight control and hepatic steatosis. *PLoS Biol.* **3,** e101 (2005).
25. Saks, V. A. *et al.* Permeabilized cell and skinned fiber techniques in studies of mitochondrial function *in vivo*. *Mol. Cell. Biochem.* **184,** 81–100 (1998).
26. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31,** e15 (2003).

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

Acknowledgements We thank members of the Puigserver laboratory for helpful comments and discussions on this work; S.-H. Kim for technical assistance; M. Montminy for the anti-PGC-1α polyclonal antibody; D. Kwiatkowski for the *TSC2*$^{-/-}$ and *TSC2*$^{+/+}$ murine embryonic fibroblasts; R. Abraham for the AU1-mTOR expression plasmid; and D. Sabatini for HA–raptor and Myc–rictor expression constructs. These studies were supported by a National Institutes of Health R21 grant (P.P.), a grant from the American Diabetes Association/Smith Family Foundation (V.K.M.) and a Burroughs Wellcome Career Award in the Biomedical Sciences (V.K.M.).

Author Information Microarray data is available online through the Gene Expression Omnibus (GEO accession number GSE5332). Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to P.P. (pere_puigserver@dfci.harvard.edu) or V.K.M. (vamsi@hms.harvard.edu).

©2007 Nature Publishing Group

METHODS

Constructs and reagents. Adenoviruses were constructed using the pAdEASY system (Stratagene). The cytochrome *c* and 2-kilobase PGC-1α pGL3 luciferase reporter plasmids have been described previously[11,27]. Gal4–ERR-α and Gal4–YY1 constructs were made by cloning full-length complementary DNAs into pCMX-R *Eco*RI/*Bam*HI and *Bam*HI/*Nhe*I sites, respectively. Flag–haemagglutinin (HA)–YY1 was cloned into VVCW/BE for lentiviral expression by using *Eco*RI/*Not*I sites. Antibodies used were: phospho-S6K1 (catalogue no. 9205; Cell Signaling), S6K1 (catalogue no. 9202; Cell Signaling), phospho-S6 (catalogue no. 2211; Cell Signaling), S6 (catalogue no. 2212; Cell Signaling), tubulin (catalogue no. 05-661; Upstate), mTOR (catalogue no. 2972; Cell Signaling), raptor (catalogue no. 4978; Cell Signaling) and rictor (catalogue no. 2140; Cell Signaling).

Cell culture and treatments. C2C12 cells were differentiated using DMEM containing 2% horse serum for 72 h before treatments or infections. Infected C2C12 cells were analysed or treated with vehicle or rapamycin (14 h at 20 nM unless otherwise indicated) 72 h after infection.

ChIP. Chromatin:protein complexes were prepared from cells using the Upstate protocol. For Raptor and Flag–HA–PGC-1α experiments, 1.5 mM DSP[28] was used to crosslink protein:protein/DNA complexes in addition to 1% formaldehyde. DNA was sheared to 200–800-base-pair fragments and immunoprecipitated with rabbit anti-YY1 antibody (catalogue no. sc-281; Santa Cruz), HA-affinity matrix (Roche) or anti-raptor antibodies (catalogue nos 2280 and 4978; Cell Signaling). Primer sequences are provided in Supplementary Information. Ratios of input DNA to bound DNA were calculated and data were normalized to vehicle-treated samples.

Mitochondrial DNA quantification. Cells were treated with dimethylsulphoxide or 20 nM rapamycin for 48 h for *TSC2*$^{-/-}$ cells and for 72 h for C2C12 cells. Mitochondrial DNA was quantified by quantitative real-time PCR by measuring the ratio of mitochondrially encoded Cox2 to an intron of the nuclear-encoded β-globin gene. Similar results were obtained with mitochondrial cytochrome *b* and an intron of nuclear glucagon. Primer sequences are provided in Supplementary Information.

Protein interaction analysis. For immunoprecipitation, nuclear extracts were prepared with the Dignam protocol except that nuclei were lysed in buffer containing 20 mM HEPES pH 7.9, 125 mM NaCl, 0.5% Nonidet P40, 1 mM EDTA, protease and phosphatase inhibitors. YY1 protein was immunoprecipitated using anti-YY1 antibody (catalogue no. sc-281; Santa Cruz). YY1 was western blotted using anti-YY1 antibody (catalogue no. sc-7341; Santa Cruz) and anti-PGC-1α antibody (gift from B. Spiegelman). YY1 and mTORC1 immunoprecipitations were performed with the protocol described[29].

Oxygen consumption analysis. Cells were trypsinized, washed in PBS and then resuspended in Dulbecco's phosphate-buffered saline (dPBS) supplemented with 25 mM glucose, 1 mM pyruvate and 2% BSA. The rate of oxygen consumption was measured with a Rank Brothers oxygen electrode. An equal number of cells under each condition were counted and placed in the electrode chamber for measurement. Data are expressed relative to the rate of respiration of control samples.

Transient transfection, luciferase reporter assays and lentivirus infection. Transient transfections were performed in HEK-293 or *TSC2*$^{-/-}$ murine embryonic fibroblasts using Polyfect (Qiagen) at a DNA:Polyfect ratio of 1:2. Cell culture medium was changed after 12 h. Cells were Lysed with Reporter Lysis Buffer (Promega) and luciferase assays were performed. For lentivirus encoding shRNA experiments, we used the same protocol as described in ref. 30; sequences are provided in Supplementary Information.

Gene expression analysis. Total RNA was extracted with Trizol (Invitrogen). cDNA generated by Superscript II enzyme (Invitrogen) with oligo(dT) primer was analysed by qRT−PCR with an iQ SYBR Green Supermix (Bio-Rad). Cell culture data were normalized to β-actin expression and mouse data were normalized to 36B4 expression. Primer sequences are provided in Supplementary Information.

Microarray analysis. *TSC2*$^{-/-}$ murine embryonic fibroblasts were either (1) treated with vehicle or 20 nM rapamycin for 14 h or (2) infected with adenovirus expressing GFP or PGC-1α for 24 h. RNA was isolated with Trizol reagent. Using Affymetrix MOE430 v2 gene chips, biological triplicates of each condition were analysed: vehicle-treated, rapamycin-treated, GFP-infected and PGC-1α-infected. Expression summary values were computed by using RMA[26] in Bioconductor with default settings (background correction version 1.2, quantile normalization) in one batch for the rapamycin/vehicle-treated cells and in another for the PGC-1α/GFP-infected cells.

***Cis*-regulatory motif analysis.** Databases of promoter regions consisting of 2,000 bases centred on the annotated transcription start site (TSS) of all mouse and human RefSeq genes were prepared using the UCSC genome browser and the MM4 and HG16 builds of the mouse and human genomes. Using an orthology mapping from the Jackson Laboratory, 10,503 mouse/human pairs of promoter regions of orthologous genes were compiled. For all sequence analyses, a gene was recorded as having a particular motif in its promoter region only if the mouse and corresponding human promoter regions both contained an instance of the motif, although the position relative to the TSS need not be conserved. The cumulative hypergeometric distribution was used to test whether the identified YY1 motif ATGGCG was significantly enriched upstream of nuclear-encoded mitochondrial genes and upstream of nuclear-encoded genes involved in oxidative phosphorylation. The Maestro predicted mitochondrial genes were used as a comprehensive list of mitochondrial genes[18]. Of the 10,503 mouse/human pairs in the promoter region database, 712 were Maestro genes and 64 were oxidative phosphorylation genes. Using the annotations from Affymetrix, 4,955 mouse/human promoter region pairs were compiled for which the mouse gene was measured by a single probe set on the Mouse 430A chip. These probe sets were ordered by using Welch's *t* statistic as a measure of differential expression between rapamycin and vehicle, and all hexamer, heptamer and octamer DNA sequences were tested for association with differential expression by using motifADE[11]. To correct for multiple testing, a flat Bonferroni correction of 4^k was applied to the nominal *P* values when testing all *k*-mer motifs. MotifADE was also used to test whether the identified YY1 motif ATGGCG was associated with differential expression in several other data sets[12,18]. In each case, all genes were ordered by differential expression by using Welch's *t* statistic. When testing a single motif (instead of, for example, 4^6 motifs) it is unnecessary to apply the Bonferroni correction.

27. Andersson, U. & Scarpulla, R. C. Pgc-1-related coactivator, a novel, serum-inducible coactivator of nuclear respiratory factor 1-dependent transcription in mammalian cells. *Mol. Cell. Biol.* **21**, 3738–3749 (2001).
28. Pierce, S. B. *et al.* Regulation of DAF-2 receptor signaling by human insulin and *ins-1*, a member of the unusually large and diverse *C. elegans* insulin gene family. *Genes Dev.* **15**, 672–686 (2001).
29. Kim, D. H. *et al.* mTOR interacts with raptor to form a nutrient-sensitive complex that signals to the cell growth machinery. *Cell* **110**, 163–175 (2002).
30. Sancak, Y. *et al.* PRAS40 is an insulin-regulated inhibitor of the mTORC1 protein kinase. *Mol. Cell* **25**, 903–915 (2007).

©2007 Nature Publishing Group

LETTERS

Inhibition of the EGF receptor by binding of MIG6 to an activating kinase domain interface

Xuewu Zhang[1]†, Kerry A. Pickin[2], Ron Bose[2]†, Natalia Jura[1], Philip A. Cole[2] & John Kuriyan[1,3]

Members of the epidermal growth factor receptor family (EGFR/ERBB1, ERBB2/HER2, ERBB3/HER3 and ERBB4/HER4) are key targets for inhibition in cancer therapy[1]. Critical for activation is the formation of an asymmetric dimer by the intracellular kinase domains, in which the carboxy-terminal lobe (C lobe) of one kinase domain induces an active conformation in the other[2]. The cytoplasmic protein MIG6 (mitogen-induced gene 6; also known as ERRFI1) interacts with and inhibits the kinase domains of EGFR and ERBB2 (refs 3–5). Crystal structures of complexes between the EGFR kinase domain and a fragment of MIG6 show that a ~25-residue epitope (segment 1) from MIG6 binds to the distal surface of the C lobe of the kinase domain. Biochemical and cell-based analyses confirm that this interaction contributes to EGFR inhibition by blocking the formation of the activating dimer interface. A longer MIG6 peptide that is extended C terminal to segment 1 has increased potency as an inhibitor of the activated EGFR kinase domain, while retaining a critical dependence on segment 1. We show that signalling by EGFR molecules that contain constitutively active kinase domains still requires formation of the asymmetric dimer, underscoring the importance of dimer interface blockage in MIG6-mediated inhibition.

Before activation, the EGFR kinase domain is in an autoinhibited conformation that resembles that of inactive cyclin-dependent kinases (CDKs) and the Src family kinases[2,6]. Conversion to the active form requires interactions between the distal surface of the C lobe of one kinase domain and the amino-terminal lobe (N lobe) of the other in the asymmetric activating dimer[2]. This conformational change resembles closely the activation switch induced in CDKs by cyclins[7], even though the C lobe of the EGFR kinase domain is structurally unrelated to cyclins.

If the cyclin/CDK-like asymmetric dimer is indeed critical for EGFR activation, then the modulation of this interaction might underlie naturally occurring mechanisms of EGFR regulation. We looked for protein inhibitors of EGFR that are known to function by interacting with the intracellular portions of the receptor. One such protein is MIG6 (or receptor-associated late transducer, RALT, the gene for which is also named gene 33), which is a feedback inhibitor of both EGFR and ERBB2 (refs 3 and 5). MIG6 inhibits EGFR-mediated signals in mouse skin[8], and deletion of the *MIG6* gene leads to hyper-activation of EGFR[9,10].

The N-terminal region of MIG6 is not implicated in EGFR inhibition (Fig. 1a). The C-terminal region shows sequence similarity to only a non-catalytic region of the ACK1 tyrosine kinase (Fig. 1a), which also binds to the EGFR cytoplasmic domain[11]. A segment within this region of MIG6 (residues 323–372) is critical for EGFR and ERBB2 binding (Fig. 1a)[12,13]. We determined the crystal structure of a 60-residue fragment spanning this segment (residues 315–374) bound to the EGFR kinase domain (Supplementary Information). This structure and structures of EGFR complexed to two overlapping 40- and 25-residue fragments (residues 325–364 and 340–364) define a 25-residue epitope of MIG6 that binds to the EGFR kinase domain (residues 337–361, denoted MIG6(segment 1). The structure of the 40-residue peptide complex has been determined at 2.9 Å resolution.

The EGFR kinase domain bound to MIG6(segment 1) adopts the Src/CDK-like inactive conformation, and not the active conformation normally seen in crystals of the kinase domain (Fig. 1b)[2,6]. The interface, which buries 1,800 Å^2 of surface area, involves an extended conformation of the MIG6 peptide and disparate binding elements on the kinase domain (Fig. 1b and c; Supplementary Information). MIG6(segment 1) lies within a shallow depression on the distal surface of the C lobe of the kinase domain, formed by helices αG and αH and the loops connecting helices αF–αG, αG–αH and αH–αI. The interactions are mainly polar, although a few hydrophobic residues from helix αH contribute to the interface.

The footprint of MIG6(segment 1) on the kinase domain overlaps the cyclin-like face of the kinase domain in the asymmetric kinase domain dimer, and so binding of MIG6 to an EGFR kinase domain will prevent it from acting as a cyclin-like activator for other kinase domains (Fig. 1, and see later). Residues in EGFR located at the MIG6(segment 1)-binding interface are conserved[2], suggesting that MIG6 will also bind to other EGFR family members.

MIG6(segment 1) binds to the EGFR kinase domain with micromolar affinity. The dissociation constant for a 30-residue fluorescein-labelled MIG6 peptide (residues 334–363, spanning the entire binding epitope of segment 1) is 13.0 ± 1.3 μM (Fig. 2a, and Supplementary Table 1). Val 924 in the C lobe of the kinase domain is located in the centre of the asymmetric kinase domain dimer interface and also participates in the interaction between the kinase domain and MIG6(segment 1)[2] (Fig. 1b, c). A V924R mutation in the kinase domain abolishes peptide binding (Fig. 2a). Met 346, Phe 352 and Tyr 358 in MIG6 are within the kinase/MIG6(segment 1) interface (Fig. 1c), and mutation of any of these residues also abrogates binding (Fig. 2b).

The EGFR kinase domain has very low activity in solution, but is activated on increasing its local concentration by tethering it to lipid vesicles, which promotes the formation of the asymmetric dimer[2]. Various MIG6 peptides that contain segment 1 inhibit the activity of the kinase domain attached to lipid vesicles, with half maximal inhibitory concentration (IC_{50}) values of ~10 μM (Fig. 2c). A 25-residue peptide (residues 340–364) that lacks 3 residues in the N-terminal portion of MIG6(segment 1), is much less potent (Fig. 2c, and Supplementary Information). Peptides that contain mutations that disrupt the binding interface (M346A, F352A and Y358A) do not inhibit kinase activity markedly (Fig. 2c). An EGFR

[1]Department of Molecular and Cell Biology, and Department of Chemistry, and Howard Hughes Medical Institute, California Institute for Quantitative Sciences, University of California, Berkeley, California 94720, USA. [2]Department of Pharmacology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [3]Physical Biosciences Division, Lawrence Berkeley National Laboratory Berkeley, California 94720, USA. †Present addresses: Department of Pharmacology and Department of Biochemistry, UT Southwestern Medical Center, Dallas, Texas 75390, USA (X.Z.); Division of Oncology, Department of Medicine, Washington University School of Medicine, St Louis, Missouri 63110, USA (R.B.).

©2007 Nature Publishing Group

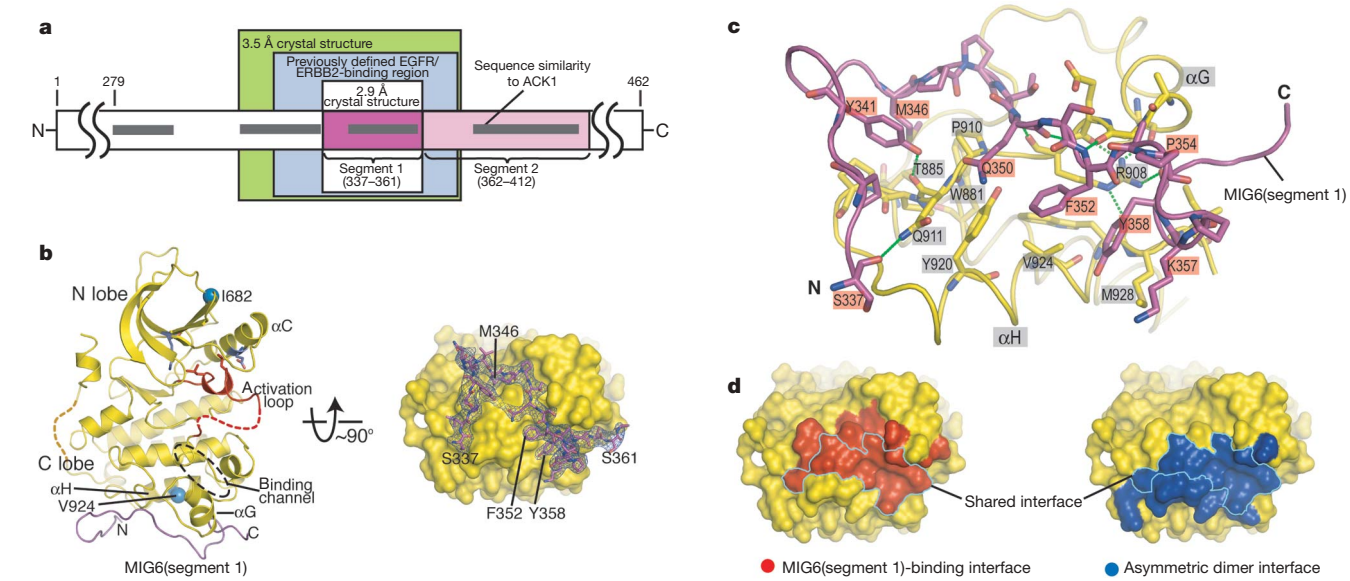


Figure 1 | Structure of the EGFR kinase domain/MIG6(segment 1). **a**, Schematic diagram of human MIG6 primary structure. Regions of interest, including the previously defined EGFR/ERBB2 binding region[4,5,12], are boxed and labelled. **b**, Two orthogonal views of the EGFR kinase domain/MIG6(segment 1) complex. A channel to which peptide inhibitors of some other kinases docked is indicated[15,16]. The electron density around MIG6(segment 1) in the right panel is contoured at 3σ and is from a simulated annealing omit map with coefficients $(|F_O|-|F_C|)e^{i\alpha_C}$, where the calculated structure factors are generated from a model that does not contain MIG6. **c**, Detailed view of the interface between the EGFR kinase domain and MIG6(segment 1). Hydrogen bonds are represented by dashed lines. **d**, Comparison of the MIG6(segment 1) binding interface and the kinase domain asymmetric dimer interface on the distal surface of the kinase C lobe. A large portion of the surface is shared by the two interfaces (outlined), and it is clear that binding of the EGFR kinase domain by MIG6(segment 1) would block the formation of the asymmetric activating dimer. **c** and **d** are in similar orientations to that in the right panel of **b**.

kinase domain bearing an I682Q mutation is not stimulated by concentration at the membrane because it is unable to form the asymmetric dimer[2]. The basal activity of this mutant in solution is not inhibited by MIG6(segment 1), which has the same binding affinity for this mutation as for the wild-type kinase domain (Fig. 2a and Supplementary Fig. 3). Thus, MIG6(segment 1) is only able to inhibit the kinase domain in the context of asymmetric dimer formation.

We tested the inhibition of EGFR autophosphorylation by full-length MIG6 in a cell-based assay. Co-expression of the wild-type MIG6 with EGFR decreases the EGF-induced autophosphorylation of EGFR, whereas introduction of individual mutations in MIG6(segment 1) (M346A, F352A or Y358A) abolishes this effect (Fig. 2d), confirming that segment 1 is important for inhibition of EGFR by full-length MIG6.

An intriguing property of MIG6 is its ability to bind more tightly to activated EGFR than to the unliganded receptor[3,5,12]. MIG6(segment 1) alone cannot confer this property, because the kinase residues that interact with it do not change conformation on activation[2,6,14]. The

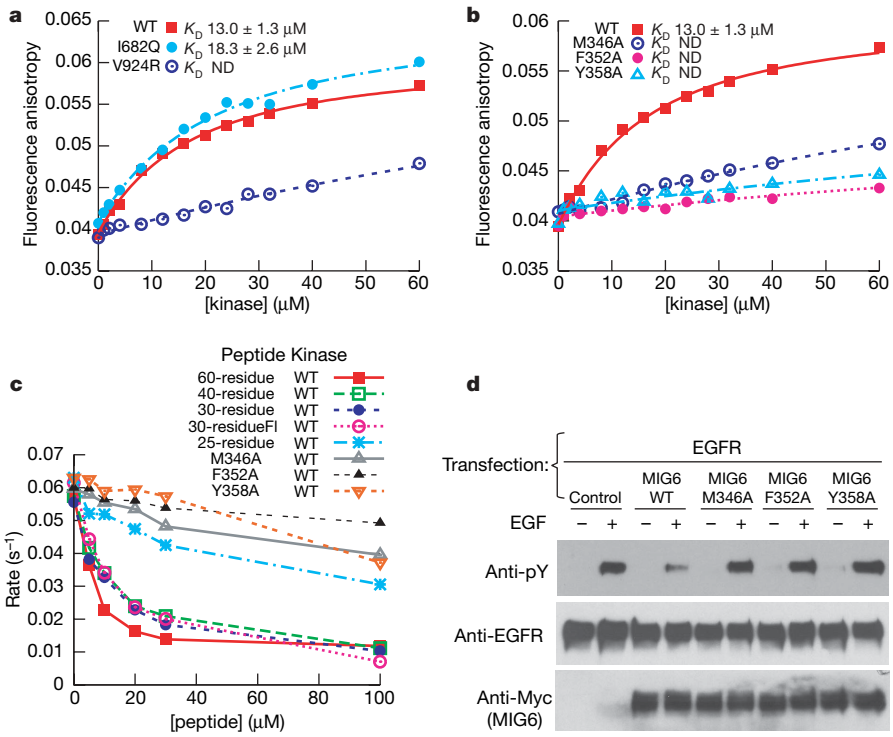


Figure 2 | Binding and inhibition of EGFR by MIG6(segment 1). **a**, Titrations of the wild-type EGFR kinase domain and the V924R and I682Q mutants to the 30-residue (residues 334–363) fluorescein-labelled MIG6 peptide. **b**, Titrations of the wild-type EGFR kinase domain to the wild-type and three mutant 30-residue fluorescein-labelled peptides. ND denotes 'not determined' (K_D values cannot be determined reliably). **c**, Inhibition of the activity of the EGFR kinase domain by peptides spanning MIG6(segment 1) in the vesicle-based kinase assay. The 60-, 40- and 30-residue peptides contain the entire binding epitope of segment 1, whereas the 25-residue peptide lacks the N-terminal 3 residues. The mutations were introduced in the 30-residue peptide. See Supplementary Table 1 for the residue boundaries. Fl, fluorescein. **d**, A cell-based assay showing that MIG6 inhibits full-length EGFR autophosphorylation, whereas mutations in segment 1 abolish the inhibition.

©2007 Nature Publishing Group

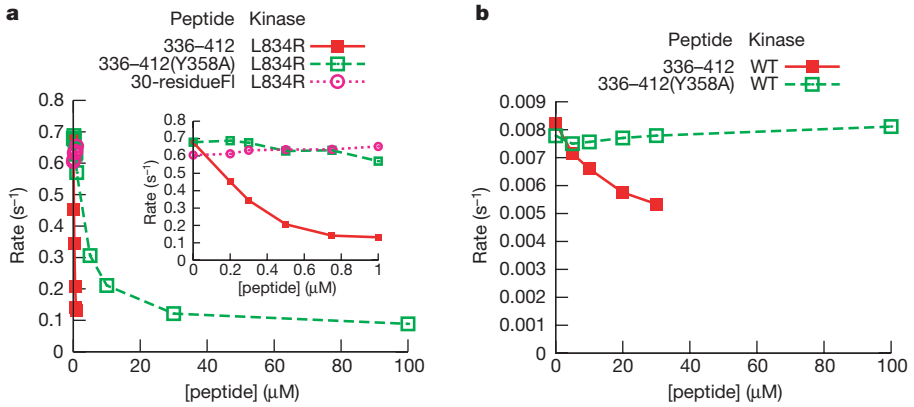


Figure 3 | Inhibition of EGFR kinase activity by MIG6(segments 1–2). a, Inhibition of the L834R mutant kinase in solution by peptides 336–412 or 336–412(Y358A) (containing both segment 1 and 2). The 30-residue peptide (containing segment 1 only) is used as a control. The insert shows an expanded view at low peptide concentrations. **b**, Inhibition of the wild-type kinase in solution by peptides 336–412 or 336–412(Y358A). Titration of peptide 336–412 beyond 30 μM leads to unreliable results owing to precipitation of the protein and peptide (see Methods).

C terminus of MIG6(segment 1) is located within a channel leading into the kinase active site (Fig. 1b), used by peptidic inhibitors of protein kinases that interact directly with the active sites[15,16]. The region of MIG6 that is C-terminal to segment 1 (segment 2, Fig. 1a) contains a region of strong homology to ACK1 (also known as TNK2) (refs 3, 5 and 11). Because MIG6 and ACK1 are both sensitive to the activation state of EGFR[3,5,11,12], there may be specific interactions between segment 2 and the activation loop and/or the N lobe of the kinase domain.

To test the role of segment 2, we produced a longer peptide (residues 336–412, MIG6(segments 1–2)), and analysed its effect on a variant of the EGFR kinase domain that contains a mutation (L834R) that renders it constitutively active in the absence of concentration on vesicles[2]. MIG6(segments 1–2) inhibits this mutant kinase domain with an IC_{50} value of ~200 nM (Fig. 3a). MIG6(segments 1–2) bearing a mutation within segment 1 (Y358A) inhibited L834R much less efficiently (IC_{50} ~5 μM). MIG6(segment 1) (the 30-residue peptide) did not inhibit this mutant kinase, consistent with its dimerization-independent activity. Interestingly, MIG6(segments 1–2) seems to be much less potent in inhibiting the basal activity of the wild-type kinase domain in solution, and MIG6(segments 1–2) bearing a mutation in segment 1 (Y358A) does not show any inhibition under the same conditions (Fig. 3b). These results suggest that segment 2 is responsible for the inhibition of the activated EGFR kinase domain, and that both segments 1 and 2 are important for the high potency of inhibition.

Could MIG6 function by binding primarily to the activated kinase in an asymmetric kinase domain dimer, and not to the cyclin-like activator kinase? The MIG6(segment 1) interaction would then be important for anchorage of MIG6 to EGFR, but not directly relevant for shutting down kinase activity. Such a role may be operative in auto-inhibition of ACK1, the kinase domain of which has a conserved segment-1-binding surface, with the MIG6 homologous segments present within the same protein (Supplementary Information). We expect, however, that the asymmetric EGFR dimer will dissociate, and that activated kinase molecules can subsequently serve as cyclin-like activators. This may facilitate the lateral propagation of EGFR activation, which can spread across the cell surface even when EGF is localized to a small region[17,18]. The interaction between MIG6(segment 1) and the kinase domain would block further transmission of the activating signal.

To examine this potential, we co-transfected cells with two variants of EGFR. One form (EGFR(activator)) resembles ERBB3 in that it is catalytically inactive (the catalytic base, Asp 813, is mutated to Asn) but can serve as a cyclin-like activator. To promote interaction with MIG6, we introduced the L834R mutation, which destabilizes the inactive conformation, into the EGFR(activator). To prevent EGFR(activator) from assuming the 'activated' position in the asymmetric dimer, we also introduced the I682Q mutation[2]. The

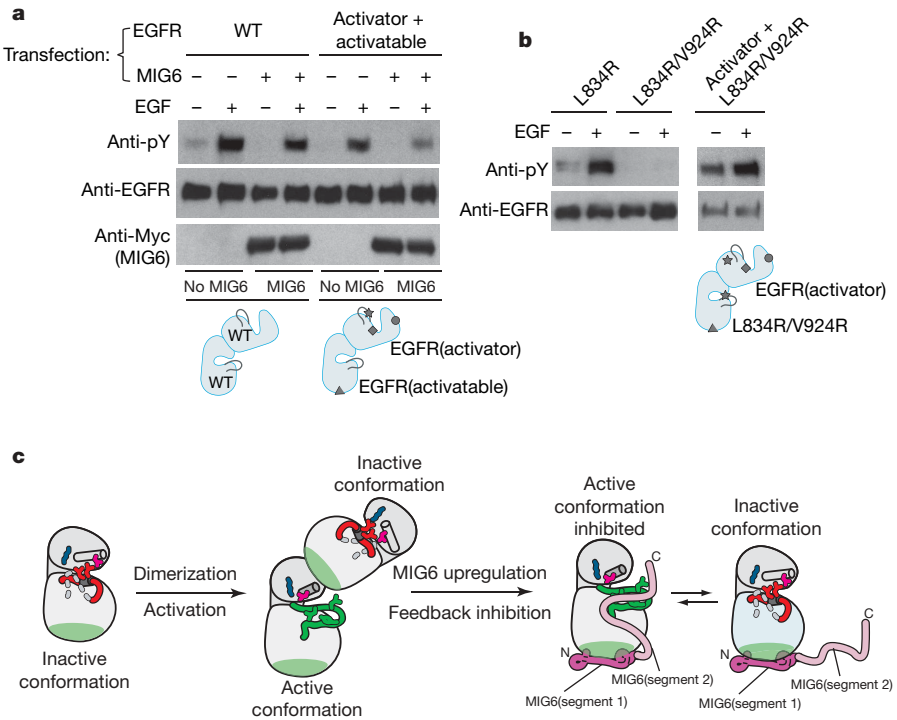


Figure 4 | A double-headed mechanism for EGFR inhibition by MIG6. a, A co-transfection experiment showing that EGFR(activator) can activate EGFR(activatable), and that MIG6 can inhibit this activation. **b**, Co-transfection experiments showing that full-length EGFR containing the L834R/V924R double mutation only shows autophosphorylation when co-transfected with EGFR(activator). Co-transfection combinations in **a** and **b** are represented by the cartoons in the respective lower panels, for clarity. The I682Q, D813N, L834R and V924R mutations are denoted in the cartoons by a circle, diamond, star and triangle, respectively. **c**, A schematic diagram showing the double-headed mechanism for EGFR inhibition by MIG6 involving both segment 1 and segment 2.

©2007 Nature Publishing Group

second EGFR variant (EGFR(activatable)) is catalytically active, but has the V924R mutation, which prevents it from serving as an activator[2]. We tested the effects of MIG6 on EGFR phosphorylation in co-transfections with these two variants. The results show that EGFR(activator) can activate EGFR(activatable) in the presence of EGF (Fig. 4a), consistent with our previous findings[2]. Co-transfection of MIG6 with EGFR(activator) and EGFR(activatable) suppresses this activation (Fig. 4a). MIG6(segment 1) does not bind to the kinase domain bearing the V924R mutation, and an intact MIG6(segment 1) is required for inhibition of EGFR in cellular assays (Fig. 2). We therefore interpret the results of the triple transfection experiment (Fig. 4a) to mean that MIG6 binds to EGFR(activator) and prevents the activation of EGFR(activatable).

Full-length EGFR bearing the activating L834R mutation is not fully phosphorylated in cells[19,20], suggesting that the formation of the asymmetric dimer is still required for robust autophosphorylation even when the kinase domain is rendered constitutively active. We confirmed this by introducing the V924R mutation, which prevents the kinase domain from serving as the cyclin-like activator, into EGFR with a constitutively active kinase domain (L834R/V924R). EGFR(L834R/V924R) fails to undergo autophosphorylation (Fig. 4b), although the kinase activity of this double mutant is comparable to that of the kinase domain bearing the single activating mutation (L834R) (Supplementary Fig. 6). EGF-stimulated autophosphorylation is restored when this double mutant is co-transfected with the kinase-dead EGFR(activator) (Fig. 4b). These results further underscore the importance of blockage of the asymmetric dimer interface by MIG6, because it can prevent both the activation of kinase domains and downstream signalling by activated kinase domains.

Our results demonstrate that MIG6 uses a double-headed mechanism for inhibiting EGFR, with the blockage of the asymmetric cyclin/CDK-like dimer being a particularly interesting aspect of the inhibition (Fig. 4c). This mechanism provides direct confirmation of the critical role of the asymmetric kinase domain dimer in the activation of EGFR family receptors. In addition, our results suggest an approach for the development of a new class of inhibitors that act by binding to the cyclin-like face of the C lobe of the kinase domains of this family. This region is not conserved in other protein kinases, and so such inhibitors may enable the development of cancer therapies with a high degree of specificity towards EGFR family members.

METHODS SUMMARY

The wild-type and mutant forms of the EGFR kinase domain were expressed and purified as described before[2]. The 60-residue MIG6 peptide was expressed in bacteria as a glutathione *S*-transferase (GST)-fusion protein, purified and treated with the TEV protease to remove the GST-moiety. The wild-type and Y358A mutant MIG6(segments 1–2) peptides were fused to a Trp ΔLE leader peptide and expressed as inclusion bodies and purified as described[21]. All other MIG6 peptides were produced using solid phase synthesis. The EGFR kinase domains (wild-type and the K799E mutant) were co-crystallized with the 60-residue, 25-residue and 40-residue MIG6 peptides and the structures were solved by molecular replacement using a structure of the EGFR kinase domain adopting the Src/CDK-like inactive conformation (PDB entry: 2GS7) as the search model. The binding affinities between the kinase domain and fluorescein-labelled MIG6 peptides were measured by monitoring the change of fluorescence anisotropy during the titration and fitting the data to a single-site binding model. Kinase assays in solution and on vesicle were performed as described before[2]. Cell-based inhibition assays were performed using Cos-7 cells co-transfected with constructs containing full-length EGFR and MIG6.

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

Received 17 July; accepted 12 October 2007.

1. Hynes, N. E. & Lane, H. A. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nature Rev. Cancer* **5**, 341–354 (2005).
2. Zhang, X. *et al.* An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. *Cell* **125**, 1137–1149 (2006).
3. Hackel, P. O., Gishizky, M. & Ullrich, A. Mig-6 is a negative regulator of the epidermal growth factor receptor signal. *Biol. Chem.* **382**, 1649–1662 (2001).
4. Xu, D., Makkinje, A. & Kyriakis, J. M. Gene 33 is an endogenous inhibitor of epidermal growth factor (EGF) receptor signaling and mediates dexamethasone-induced suppression of EGF function. *J. Biol. Chem.* **280**, 2924–2933 (2005).
5. Fiorentino, L. *et al.* Inhibition of ErbB-2 mitogenic and transforming activity by RALT, a mitogen-induced signal transducer which binds to the ErbB-2 kinase domain. *Mol. Cell. Biol.* **20**, 7735–7750 (2000).
6. Wood, E. R. *et al.* A unique structure for epidermal growth factor receptor bound to GW572016 (Lapatinib): relationships among protein conformation, inhibitor off-rate, and receptor activity in tumor cells. *Cancer Res.* **64**, 6652–6659 (2004).
7. Jeffrey, P. D. *et al.* Mechanism of CDK activation revealed by the structure of a cyclinA–CDK2 complex. *Nature* **376**, 313–320 (1995).
8. Ballaro, C. *et al.* Targeted expression of RALT in mouse skin inhibits epidermal growth factor receptor signalling and generates a Waved-like phenotype. *EMBO Rep.* **6**, 755–761 (2005).
9. Ferby, I. *et al.* Mig6 is a negative regulator of EGF receptor-mediated skin morphogenesis and tumor formation. *Nature Med.* **12**, 568–573 (2006).
10. Zhang, Y. W. *et al.* Evidence that *MIG-6* is a tumor-suppressor gene. *Oncogene* **26**, 269–276 (2007).
11. Shen, F. *et al.* Activated Cdc42-associated kinase 1 is a component of EGF receptor signaling complex and regulates EGF receptor degradation. *Mol. Biol. Cell* **18**, 732–742 (2007).
12. Anastasi, S. *et al.* Feedback inhibition by RALT controls signal output by the ErbB network. *Oncogene* **22**, 4221–4234 (2003).
13. Anastasi, S. *et al.* The evolutionarily conserved EBR module of RALT/MIG6 mediates suppression of the EGFR catalytic activity. *Oncogene* advanced online publication doi:10.1038/sj.onc.1210590 (18 June 2007).
14. Stamos, J., Sliwkowski, M. X. & Eigenbrot, C. Structure of the epidermal growth factor receptor kinase domain alone and in complex with a 4-anilinoquinazoline inhibitor. *J. Biol. Chem.* **277**, 46265–46272 (2002).
15. Depetris, R. S. *et al.* Structural basis for inhibition of the insulin receptor by the adaptor protein Grb14. *Mol. Cell* **20**, 325–333 (2005).
16. Lei, M., Robinson, M. A. & Harrison, S. C. The active conformation of the PAK1 kinase domain. *Structure* **13**, 769–778 (2005).
17. Reynolds, A. R. *et al.* EGFR activation coupled to inhibition of tyrosine phosphatases causes lateral signal propagation. *Nature Cell Biol.* **5**, 447–453 (2003).
18. Ichinose, J., Murata, M., Yanagida, T. & Sako, Y. EGF signalling amplification induced by dynamic clustering of EGFR. *Biochem. Biophys. Res. Commun.* **324**, 1143–1149 (2004).
19. Lynch, T. J. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004).
20. Sordella, R., Bell, D. W., Haber, D. A. & Settleman, J. Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways. *Science* **305**, 1163–1167 (2004).
21. Conti, E. & Kuriyan, J. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin α. *Structure* **8**, 329–338 (2000).

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

Acknowledgements We thank X. Cao and A. Fisher for cell culture support; D. King for mass spectrometry; the staff at beamlines 8.2.1 and 12.3.1 of the Advanced Light Source for technical support; and M. Seeliger, S. Deindl, P. Pellicena, J. Gureasko, S. Jacques and other members in the Kuriyan and Cole groups for technical help and discussions. We thank T. Miller for discussions on ACK1. This work is supported in part by grants from the NCI to J.K. and from NIH to P.A.C. R.B. is supported by the Susan G. Komen Breast Cancer Foundation.

Author Contributions J.K., X.Z. and P.A.C. designed the project. X.Z. performed biochemical and cell-based assays and crystallography work. K.A.P. and R.B. synthesized peptides. N.J. performed some biochemical and cell-based assays. J.K. and X.Z. wrote the manuscript.

Author Information The atomic coordinates are deposited at the Protein Data Bank under accession codes 2RF9, 2RFD and 2RFE. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.K. (kuriyan@berkeley.edu).

©2007 Nature Publishing Group

METHODS

Peptide preparation. All the MIG6 peptides used in this study are listed in Supplementary Table 1. The 60-residue peptide was expressed as a GST-fusion in *Escherichia coli* BL21 (DE3) by using pGEX6p1 (Amersham) (BamHI/XhoI) and purified using a glutathione Sepharose column. The protein was treated with the PreScission protease to release the MIG6 peptide, which was further purified using a Hitrap SP column (Amersham). The longer peptides (336–412 and 336–412(Y358A)) were cloned as Trp ΔLE fusions and expressed as inclusion bodies as described previously[21]. To prevent cleavage of the MIG6 peptides by subsequent cyanogen bromide treatment the single methionine in these peptides (M346) was mutated to leucine. This mutation does not affect the binding to the EGFR kinase domain significantly (Supplementary Fig. 4). The fusion proteins were cleaved with cyanogen bromide and the released MIG6 peptides were purified. All other MIG6 peptides were synthesized using solid-phase peptide synthesis using the Fmoc strategy with Wang resin on a Protein Technologies PS3 synthesizer. The peptide identities were confirmed by mass spectrometry.

Structure determination. The EGFR kinase domain constructs used are identical to those used previously[2], except for the K799E mutant, which includes six fewer residues at the N terminus (spanning residues 678–998 of EGFR), and proteins were purified as described[2]. The wild-type kinase domain was first co-crystallized with the 60-residue MIG6 peptide and the structure was determined at 3.5 Å resolution. This revealed that a ~25-residue segment of the peptide is bound to the distal surface of the C lobe of the EGFR kinase domain and that the rest of the peptide is disordered (Supplementary Fig. 1a). A 25-residue peptide (residues 340–364 in MIG6) was designed on the basis of the initial structure and co-crystallized with both the wild-type and a mutant (K799E) form of the EGFR kinase domain. The K799E mutation does not affect the conformation of the kinase domain or its interaction with MIG6(segment 1) (Supplementary Fig. 5), but crystals of this mutant kinase domain in complex with the peptide diffracted X-rays to higher resolution. The structure shows that this 25-residue peptide lacks the N-terminal part of the kinase binding epitope (Supplementary Fig. 1b). This peptide was then extended to include residues 325–364 in MIG6 (the 40-residue peptide) and co-crystallized with the EGFR(K799E) kinase domain. The structure of this peptide–kinase domain complex was determined at 2.9 Å. There are four kinase domains in the asymmetric unit, all of which adopt the same conformation. Two of the four kinase domains are bound to the MIG6 peptide, and the MIG6 binding surfaces of the other two are occupied by crystal contacts. Crystallization conditions, data collection and structural refinement statistics are summarized in Supplementary Table 2.

Binding assays. Fluorescein-labelled 30-residue wild-type, M346L, M346A, F352A and Y358A MIG6 peptides were diluted to final concentrations of 5, 8, 3.1, 3.5 and 2.7 μM in a buffer containing 10 mM Tris, 50 mM NaCl and 2 mM DTT, pH 7.5. These peptides in the cuvette were then titrated with the wild-type or mutant forms of the EGFR kinase domain at 20 °C. For the competition assays, the labelled 30-mer wild-type peptide (5 μM) and kinase domain (60 μM) were mixed and titrated with unlabelled competitor peptides. The fluorescence anisotropy at each titration step was monitored. The I682Q and K799E mutant kinases used in the binding assays contain the N-terminal 6×His tag and linker fragment before the kinase domain, whereas this N-terminal fragment in the wild-type and V924R mutant kinases was removed by Tobacco Etch Virus protease treatment.

Kinase assays. The assays were conducted as described previously[2]. The substrate peptide was kept at 1 mM in all the experiments. The reported rates are the initial velocities normalized by the kinase concentrations. The wild-type kinase concentrations in the vesicle-based and solution-based assays were 3.5 and 14 μM respectively. Preliminary experiments showed that peptide 336–412 (MIG6(segments 1–2) inhibited the L834R mutant kinase much more strongly and also caused precipitation when both the kinase and the peptide were at high concentrations. We therefore reduced the concentration of L834R in the assays to 200 nM. The higher intrinsic activity of this mutant and usage of $MnCl_2$ at 10 mM instead of $MgCl_2$ allowed us to measure kinase activity at such a low kinase concentration[22].

Cell-based assays. Cos-7 cells were co-transfected using Fugene 6 (Roche) with the DNA encoding the N-terminal Flag-tagged EGFR in pcDNA3.1 constructs (as used before[2]) and the wild-type or mutants of the *MIG6* genes with a C-terminal Myc tag (also in pcDNA3.1). Cells were cultured for 36 h after transfection and serum-starved for 12 h. Cells were treated with EGF (50 ng ml^{-1}) for ~5 min at 37 °C, lysed and subjected to western blot analyses. The levels of total EGFR, EGFR autophosphorylation and MIG6 were probed using the anti-EGFR antibody SC03 (Santa Cruz), anti-phosphotyrosine antibody 4G10 (Upstate) and an anti-Myc antibody (Cell Signaling), respectively.

22. Wedegaertner, P. B. & Gill, G. N. Activation of the purified protein tyrosine kinase domain of the epidermal growth factor receptor. *J. Biol. Chem.* **264**, 11346–11353 (1989).

©2007 Nature Publishing Group

LETTERS

The centromere geometry essential for keeping mitosis error free is controlled by spindle forces

Jadranka Lončarek[1]*, Olga Kisurina-Evgenieva[1]*†, Tatiana Vinogradova[1], Polla Hergert[1], Sabrina La Terra[1,2], Tarun M. Kapoor[3] & Alexey Khodjakov[1,2,3]

Accurate segregation of chromosomes, essential for the stability of the genome, depends on 'bi-orientation'—simultaneous attachment of each individual chromosome to both poles of the mitotic spindle[1]. On bi-oriented chromosomes, kinetochores (macromolecular complexes that attach the chromosome to the spindle) reside on the opposite sides of the chromosome's centromere[2]. In contrast, sister kinetochores shift towards one side of the centromere on 'syntelic' chromosomes that erroneously attach to one spindle pole with both sister kinetochores. Syntelic attachments often arise during spindle assembly and must be corrected to prevent chromosome loss[3]. It is assumed that restoration of proper centromere architecture occurs automatically owing to elastic properties of the centromere[1,2]. Here we test this assumption by combining laser microsurgery and chemical biology assays in cultured mammalian cells. We find that kinetochores of syntelic chromosomes remain juxtaposed on detachment from spindle microtubules. These findings reveal that correction of syntelic attachments involves an extra step that has previously been overlooked: external forces must be applied to move sister kinetochores to the opposite sides of the centromere. Furthermore, we demonstrate that the shape of the centromere is important for spindle assembly, because bipolar spindles do not form in cells lacking centrosomes when multiple chromosomes with juxtaposed kinetochores are present. Thus, proper architecture of the centromere makes an important contribution to achieving high fidelity of chromosome segregation.

Kinetochores on bi-oriented chromosomes are positioned on the opposite sides of the centromere[2]. However, during mitotic spindle formation both sister kinetochores sometimes attach to the same spindle pole, becoming 'syntelic'. Under this condition, microtubule-dependent forces shift sister kinetochores to the same side of the centromere. Because syntelic attachment would lead to aneuploidy, this configuration is not stable[4,5]. Kinetochore fibres (K-fibres) on syntelic chromosomes depolymerize so that the chromosome moves to the spindle pole where at least one of the two kinetochores detaches from microtubules[6–8]. Detached kinetochores can then connect to microtubules from the opposite spindle pole to achieve proper bi-orientation. However, for this mechanism to work properly the shape of the centromere must be restored such that sister kinetochores return to opposite sides of the centromere. It is generally assumed that this occurs automatically owing to elasticity of chromatin[9,10]. However, this assumption has not been tested.

Mammalian cells treated with monastrol, a small-molecule inhibitor of the molecular motor Eg5 (kinesin-5), arrest in mitosis with monopolar spindles[11]. Up to 70% of chromosomes in these cells are syntelic[11,12]. We reasoned that if restoration of proper centromere architecture occurs automatically then kinetochores on syntelic chromosomes should spring back to the opposite sides of the centromere if microtubules are rapidly depolymerized. However, kinetochores will remain juxtaposed under these conditions if their repositioning on the centromere requires microtubule-based forces (Fig. 1a).

We treated PtK_1 (female rat kangaroo kidney epithelial) cells that constitutively express a γ-tubulin–green fluorescent protein (GFP)[13] with 100 μM monastrol for 30–60 min. Mitotic cells with monopolar spindles were visualized by differential-interference contrast (DIC) microscopy and followed for 10–15 min (1 image every 30–60 s). Then, in addition to monastrol, the cells were treated with 5 μM

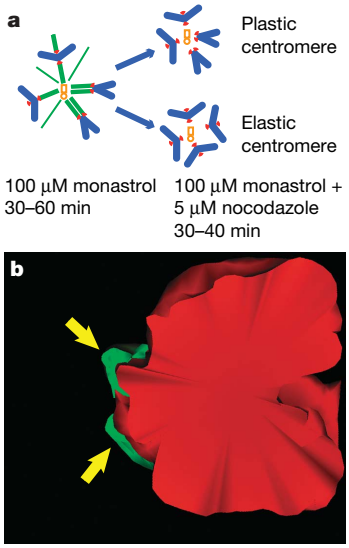


Figure 1 | Juxtaposed sister kinetochores on syntelic chromosomes do not return to opposite sides of the centromere in the absence of microtubules. a, A schematic representation of the experiment. Cells treated with monastrol form monopolar spindles with a high incidence of syntelic chromosomes; both sister kinetochores (red) on these chromosomes are positioned side-by-side and connected to the centrosomes (light brown) by microtubule bundles (green). If microtubules are depolymerized in monastrol-treated cells, kinetochores should remain juxtaposed if the chromosome's centromere is malleable (or plastic) (top diagram) or return to the opposite sides if the centromere is spring-like (that is, elastic) (bottom diagram). **b**, Surface-rendered serial-section chromosome reconstruction from a cell assayed as described in **a**. Despite complete lack of microtubules, both sister kinetochores (arrows) remain juxtaposed (electron microscopy sections are presented in Supplementary Fig. 1).

[1]Division of Molecular Medicine, Wadsworth Center, Albany, New York State Department of Health, Albany, New York 12201-0509, USA. [2]Department of Biomedical Sciences, State University of New York, Albany, New York 12222, USA. [3]Laboratory of Chemistry and Cell Biology, The Rockefeller University, New York, New York 10021, USA. †Present address: Department of Cytology, Biology Faculty, Moscow State University, Moscow 119991, Russia.
*These authors contributed equally to this work.

©2007 Nature Publishing Group

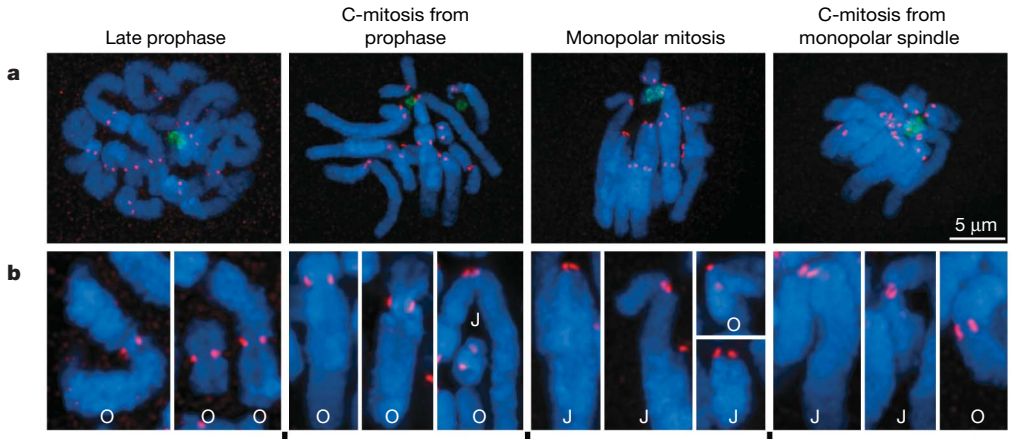
nocodazole, which completely depolymerizes microtubules in mitotic cells in ~3 min (Supplementary Fig. 1). The cells were imaged for another 30–45 min and then fixed for electron microscopy. This approach ensured that electron microscopy analyses were conducted on cells exposed to nocodazole only after they had formed monopolar spindles and accumulated syntelic chromosomes in the presence of monastrol.

Serial-section electron microscopy reconstructions revealed that sister kinetochores on many chromosomes resided on the same side of the centromere (Fig. 1b, and Supplementary Fig. 2). The frequency of chromosomes with juxtaposed kinetochores was estimated via three-dimensional fluorescence microscopy (Fig. 2). Visual inspection revealed that sister kinetochores were positioned on the same side of the centromere (within an ~90° segment) on ~50% of chromosomes in cells with monopolar spindles. This frequency did not change in cells that were additionally treated with nocodazole, suggesting that once displaced to the same side of the centromere (owing to syntelic attachment), kinetochores do not return to the opposite sides of the centromere on microtubule depolymerization. Further, the average distance between sister kinetochores measured in three-dimensional fluorescence light microscopy revealed that sister kinetochores resided closer to one another in monopolar spindles than in prophase cells and this distance did not change on microtubule depolymerization (Fig. 2). Thus, restoration of proper centromere organization during correction of syntelic attachments is not achieved through elastic recoil, but requires external forces.

Forces responsible for straightening the sister-kinetochore–centromere axis can be generated when one kinetochore on a formerly syntelic chromosome captures astral microtubules from one spindle pole while its sister connects to the other pole. The formation of astral microtubules depends on centrosomes. Therefore, we tested if correction of syntelic attachments occurs in the absence of centrosomes by ablating these organelles with a laser microbeam[14,15] ($N = 12$). Astral microtubules disappear in 3–5 min after centrosome ablation[15]. On monastrol washout in cells without microtubule asters, chromosomes remained disorganized, moving intermittently in random directions for ~1 h (Fig. 3a). Immunofluorescence analyses ($N = 20$) revealed a highly disorganized microtubule pattern (Fig. 3b). Prominent bundles of microtubules (K-fibres) emanated from the chromosomes and converged on numerous small foci (5–10 foci per cell). These foci were not γ-tubulin-positive but contained highly concentrated NuMA—a large protein responsible for spindle pole focusing[16]. Despite the lack of an organized mitotic spindle, ~1 h after centrosome ablation and monastrol washout, sister chromatids separated and exhibited short directed movements resembling anaphase motion. The extent of these movements was limited to 2–3 μm along different directions. As a result, chromatids remained in a single group and ultimately reconstituted a single nucleus.

Serial-section electron microscopy analyses of four acentrosomal cells fixed 1 h after monastrol washout revealed that sister kinetochores often remained on the same side of the centromere. K-fibres attached to these kinetochores were oriented parallel to each other (Figs 3d, and Supplementary Fig. 3). All kinetochores, whether juxtaposed or properly positioned, were associated with prominent K-fibres. Thus, in the absence of centrosomes, restoration of proper centromere organization and spindle bipolarization are impeded. In sharp contrast, monastrol-induced monopolar spindles consistently bipolarize in cells with centrosomes in ~1 h[6,12].

Kinetochores can attach to spindle microtubules either by capturing centrosome-generated astral microtubules[17] or by developing their own K-fibres[12,18,19]. In the latter case, K-fibres formed by juxtaposed sister kinetochores would be oriented parallel to one another, whereas K-fibres formed by sister kinetochores on a properly organized centromere extend in opposite directions. Thus, the shape of the centromere can be a major factor in spindle formation: proper centromere organization should promote bipolarity, whereas juxtaposed sister kinetochores should impede it. The effects of centromere architecture on spindle formation should be particularly prominent in the absence of centrosomes, when changes in the shape of centromeres induced by syntelic attachments cannot be restored.



	Late prophase	C-mitosis from prophase	Monopolar mitosis	C-mitosis from monopolar spindle
K–K distance (μm)	1.20 ± 0.22	1.07 ± 0.12	0.88 ± 0.19	0.86 ± 0.19
Range	0.81–1.83	0.75–1.41	0.43–1.23	0.47–1.43
Percentage of juxtaposed kinetochores	<10%	~10%	~50%	~50%

Figure 2 | Frequencies of chromosomes with juxtaposed sister kinetochores observed under different experimental conditions. **a**, Maximal-intensity projection spanning the entire volume of the cell; and **b**, individual chromosomes from cells shown in **a** presented at 2× additional magnification. During prophase, sister kinetochores in untreated cells are positioned on the opposite sides of the centromere and separated by ~1.2 μm. When microtubules are depolymerized during prophase–prometaphase, sister kinetochores on most chromosomes remain opposed, although occasionally they can be seen on the same side of the centromere (chromosomes marked O or J, respectively). In cells with monopolar spindles, at least 50% of kinetochores are on the same side of the centromere (juxtaposed). This change in the organization of the centromere is reflected by the decrease in the average distance between sister kinetochores (the 'C-mitosis' and 'monopolar mitosis' populations are different, with > 99.99% confidence by the two-tailed Student's *t*-test). The percentage of chromosomes and average sister kinetochore separation does not change when microtubules in cells with monopolar spindles are depolymerized with 5 μM nocodazole. Centrosomes, green (γ-tubulin–GFP); kinetochores, red (CREST); DNA, blue (Hoechst 33343).

©2007 Nature Publishing Group

To test this idea, we examined whether a functional bipolar spindle can form via acentrosomal pathways in cells with multiple juxtaposed sister kinetochores. We treated PtK_1 cells with monastrol to accumulate syntelic chromosomes, then ablated both centrosomes and depolymerized microtubules with nocodazole[14,15]. Nine of twelve cells failed to assemble a bipolar mitotic spindle when both drugs were washed out (Fig. 4a, Table 1). In contrast, if centrosomes were ablated during prophase and cells were treated with monastrol and nocodazole before nuclear envelope breakdown (when the vast majority of sister kinetochores are on opposite sides of the centromere, Fig. 2), they formed a functional spindle on drug washout (Fig. 4b, Table 1). For these two types of experiments cells were treated with the same combination of drugs. Nevertheless, the results were dramatically different, indicating that the effect was specific to the difference in the architecture of centromeres. Furthermore, centrosomal cells consistently formed a bipolar spindle after consecutive treatment with monastrol and nocodazole, indicating that restoration of properly shaped centromeres allows chromosomes to achieve bipolarization (Supplementary Fig. 4, Table 1).

Although cells with properly organized centromeres were able to form functional mitotic spindles in the absence of centrosomes, many of these cells contained syntelic chromosomes at anaphase onset (Fig. 4b, Table 1). Syntelic chromosomes were also consistently present at anaphase onset if cells were treated with nocodazole alone,

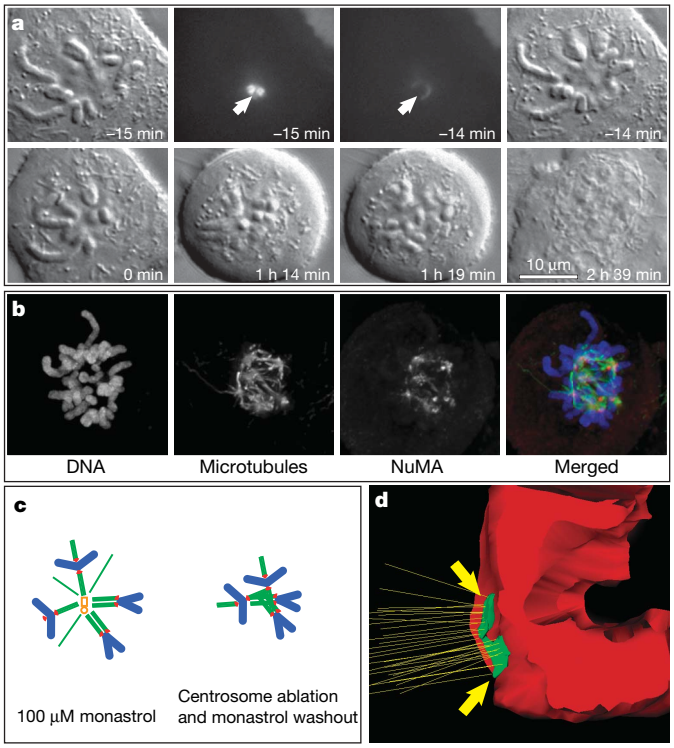


Figure 3 | Monopolar spindles do not bipolarize and sister kinetochores remain juxtaposed on monastrol washout in the absence of centrosomes. **a**, Both centrosomes in a monastrol-induced monopolar mitosis were ablated (arrows in −15 min and −14 min frames), then monastrol was washed out and the cell followed by time-lapse DIC microscopy. The cell did not organize a bipolar spindle for more than an hour, at which point cohesion between sister chromatids was lost (1 h 14 min) and the cell exited mitosis. Although individual sister chromatids attempted to move away from one another, the extent of their movement was not sufficient to form discrete chromosome groups (1 h 19 min). A single daughter cell with one nucleus was formed on completion of mitosis (2 h 39 min). **b**, Distribution of microtubules, chromosomes and NuMA in acentrosomal cells shortly before anaphase onset (45–60 min after monastrol washout). Prominent bundles of microtubules (K-fibres) were associated with all chromosomes. These K-fibres were not organized in a bipolar spindle but converged on multiple small centres that contained elevated amounts of the spindle-pole protein NuMA. All images are maximal-intensity projections through the entire cell. **c**, Schematic representation of the spindle reorganization that occurs in monastrol-induced monopolar mitoses after ablation of the centrosome and monastrol washout. **d**, Surface-rendered model of a centromere from a cell fixed 60 min after monastrol washout. Sister kinetochores remained juxtaposed and attached to prominent microtubule bundles, which terminated inside the kinetochore (original electron microscopy data are presented in Supplementary Fig. 2).

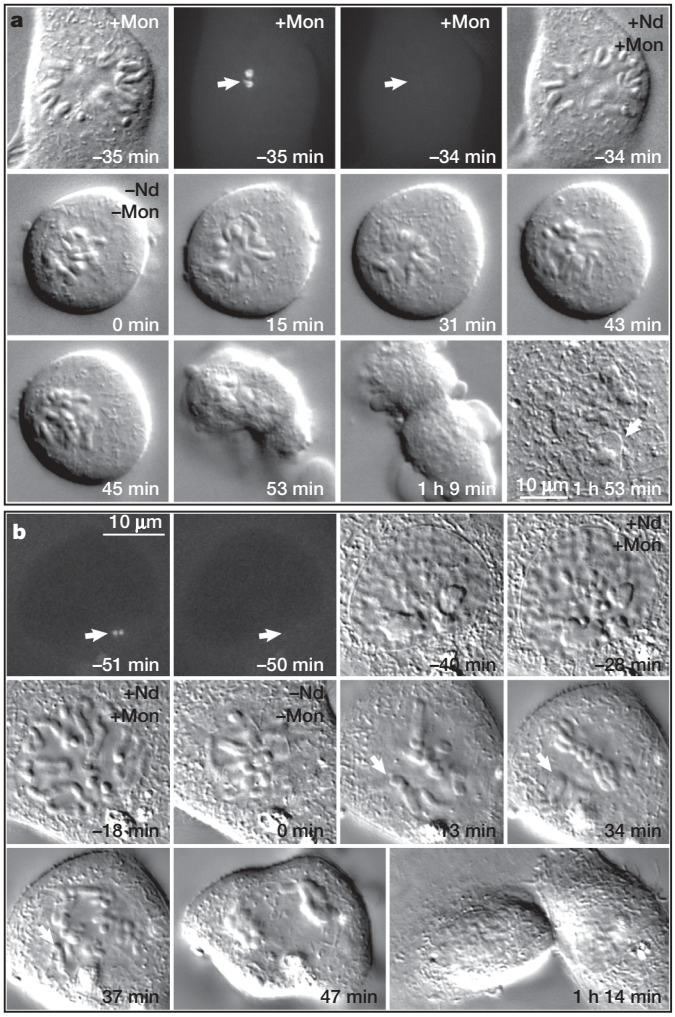


Figure 4 | Proper organization of the centromere is required for successful spindle formation in the absence of centrosomes. **a**, This PtK_1 cell was pretreated with monastrol for 30 min. Both centrosomes were ablated (arrows in −35 min and −34 min), after which point the cell was treated with 5 µM nocodazole in addition to monastrol. Thirty minutes after addition of nocodazole, both drugs were washed out. Time-lapse DIC microscopy revealed that chromosomes remained disorganized in the cytoplasm experiencing short erratic movements for approximately 45 min. Then cohesion between sister chromatids was lost (43 min and 45 min). Although the cell attempted to undergo cytokinesis, furrowing activity was disorganized (53 min–1 h 9 min) and all furrows ultimately failed so that mitosis resulted in a single daughter cell with a complexly shaped nucleus (arrow in 1 h 53 min). **b**, Mitosis in a cell where both centrosomes were ablated during late-G2–early-prophase (arrows in −51 min and −50 min). Then, during late prophase, the cell was treated with 5 µM nocodazole and 100 µM monastrol (−28 min–0 min). The cell entered mitosis 10 min after addition of the drugs (−18 min) and chromosomes become scattered in the cytoplasm. On washout of the drugs (0 min), the cell assembled a bipolar mitotic spindle (13 min–34 min) and successfully divided into two daughter cells (37 min–1 h 14 min). However, notice that anaphase was initiated in the presence of two syntelic chromosomes (arrows in 13 min and 34 min). Shown are maximal-intensity projections of three-dimensional GFP fluorescence data sets and selected frames from the DIC time-lapse recording.

©2007 Nature Publishing Group

Table 1 | Frequency of successful mitotic spindle formation in centrosomal and acentrosomal cells on different experimental conditions (see text for details).

Experimental conditions	Functional spindle (%)
Centrosome ablation before nuclear envelope breakdown	>90*
Nocodazole →C-mitosis→centrosome ablation→drug washout	90 (9/10)†
Monastrol→monopolar spindle→drug washout	>90‡
Monastrol→monopolar spindle→nocodazole→C-mitosis→drug washout	100 (10/10)
Monastrol→monopolar spindle→centrosome ablation→drug washout	0 (0/12)
Monastrol→monopolar spindle→centrosome ablation→nocodazole→C-mitosis →drug washout	25 (3/12)
Centrosome ablation (in G2)→monastrol + nocodazole (added before nuclear envelope breakdown) →C-mitosis→drug washout	90 (8/9)§

* Previously published data[14].
† Six of these cells contained one or more syntelic chromosomes at anaphase onset.
‡ Previously published data[11,12].
§ Five of these cells contained one or more syntelic chromosomes at anaphase onset. Two of these spindles were tripolar and the other six were bipolar.

then the centrosomes were ablated and nocodazole washed out (Supplementary Fig. 5, Table 1). These observations are consistent with our hypothesis that correction of syntelic attachments is impeded in the absence of astral microtubules.

Intriguingly, the presence of multiple syntelic chromosomes did not prevent mitotic exit. It is unlikely that the centromere is under tension when sister kinetochores are juxtaposed because the distance between them does not change on loss of microtubule attachments. These observations support the notion that, in contrast to lower eukaryotes[20,21], the spindle assembly checkpoint in mammals is satisfied in the absence of tension as long as all kinetochores are attached to microtubules[21–24]. Consistent with this hypothesis, immunofluorescence analyses demonstrated that the checkpoint protein Mad2 was present on multiple kinetochores in acentrosomal cells soon after monastrol washout, but gradually disappeared before the cells exited mitosis, in spite of insufficient sister kinetochore separation (Supplementary Fig. 6).

Intuitively, it seems that flexible centromeres would be disadvantageous to the cell. However, this feature can be important for promoting chromosome congression. Most mono-oriented chromosomes congress to the spindle equator within minutes, after nuclear envelope breakdown. However, sometimes mono-oriented chromosomes are seen to remain in the vicinity of the pole for up to several hours[24,25]. Electron microscopy analyses demonstrated that these persistently mono-oriented chromosomes are 'monotelic'—they are attached to the proximal spindle pole with one kinetochore, whereas the second kinetochore is positioned on the opposite side of the centromere[25,26]. In this configuration, the chances of the unattached kinetochore encountering microtubules coming from the distal pole are negligible. In contrast, mono-oriented chromosomes with juxtaposed sister kinetochores can efficiently congress to the spindle equator by sliding on K-fibres of other already-bi-oriented chromosomes[27]. During this type of congression the leading kinetochore is oriented towards the distal spindle pole, maximizing its chances to capture an astral microtubule and become attached. In this respect, syntelic chromosome attachments may serve as an efficient intermediate step towards bi-orientation.

Achieving chromosome bi-orientation depends on a complex interplay between mechanisms intrinsic to the centromere and those that act externally. The former include aurora-kinase/kinesin-13-mediated destabilization of K-fibres and release of erroneously attached chromosomes[6–8]. Activation of these mechanisms is a necessary first step in the correction of chromosome mis-orientation. However, we demonstrate here that in the absence of external forces needed to restore centromere architecture, centromere-intrinsic mechanisms are not sufficient. Our findings imply that mechanical properties and the shape of the centromere play an important part in the fidelity of chromosome segregation.

METHODS SUMMARY

Detailed layout of our laser microscopy workstation has been described elsewhere[28]. In brief, 8-ns pulses of 532-nm light from Nd:YAG laser (Diva II) were focused with the same 100× 1.4 PlanApo objective lens that was used for observations. It takes ~10–20 pulses to completely destroy the centrosome in PtK cells during mitosis. Fluorescence images presented in the manuscript are maximal-intensity projections of complete Z-series through the cell. DIC images are single Z-planes.

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

Received 6 July; accepted 3 October 2007.

1. Nicklas, R. B. How cells get the right chromosomes. *Science* **275,** 632–637 (1997).
2. Rieder, C. L. & Salmon, E. D. The vertebrate kinetochore and its roles during mitosis. *Trends Cell Biol.* **8,** 310–318 (1998).
3. Cimini, D. & Degrassi, F. Aneuploidy: a matter of bad connections. *Trends Cell Biol.* **8,** 442–451 (2005).
4. Ault, J. G. & Nicklas, R. B. Tension, microtubule rearrangements, and the proper distribution of chromosomes in mitosis. *Chromosoma* **98,** 33–39 (1989).
5. Nicklas, R. B. & Ward, S. C. Elements of error correction in mitosis: microtubule capture, release, and tension. *J. Cell Biol.* **126,** 1241–1253 (1994).
6. Lampson, M. A., Renduchitala, K., Khodjakov, A. & Kapoor, T. M. Correcting improper chromosome–spindle attachments during cell division. *Nature Cell Biol.* **6,** 232–237 (2004).
7. Kline-Smith, S. L., Khodjakov, A., Hergert, P. & Walczak, C. E. Depletion of centromeric MCAK leads to chromosome congression and segregation defects due to improper kinetochore attachments. *Mol. Biol. Cell* **15,** 1146–1159 (2004).
8. Andrews, P. D. *et al.* Aurora B regulates MCAK at the mitotic centromere. *Dev. Cell* **6,** 253–268 (2004).
9. Sullivan, K. F. & Shelby, R. D. Using time-lapse confocal microscopy for analysis of centromere dynamics in human cells. *Methods Cell Biol.* **58,** 183–202 (1999).
10. Poirier, M. G. & Marko, J. F. Micromechanical studies of mitotic chromosomes. *Curr. Top. Dev. Biol.* **55,** 75–141 (2003).
11. Kapoor, T. M., Mayer, T. U., Coughlin, M. L. & Mitchison, T. J. Probing spindle assembly mechanisms with monastrol, a small molecule inhibitor of the mitotic kinesin, Eg5. *J. Cell Biol.* **150,** 975–988 (2000).
12. Khodjakov, A., Copenagle, L., Gordon, M. B., Compton, D. A. & Kapoor, T. M. Minus-end capture of preformed kinetochore fibers contributes to spindle morphogenesis. *J. Cell Biol.* **160,** 671–683 (2003).
13. Khodjakov, A. & Rieder, C. L. The sudden recruitment of γ-tubulin to the centrosome at the onset of mitosis and its dynamic exchange throughout the cell cycle, do not require microtubules. *J. Cell Biol.* **146,** 585–596 (1999).
14. Khodjakov, A., Cole, R. W., Oakley, B. R. & Rieder, C. L. Centrosome-independent mitotic spindle formation in vertebrates. *Curr. Biol.* **10,** 59–67 (2000).
15. Khodjakov, A. & Rieder, C. L. Centrosomes enhance the fidelity of cytokinesis in vertebrates and are required for cell cycle progression. *J. Cell Biol.* **153,** 237–242 (2001).
16. Compton, D. A. Focusing on spindle poles. *J. Cell Sci.* **111,** 1477–1481 (1998).
17. Rieder, C. L. & Alexander, S. P. Kinetochores are transported poleward along a single astral microtubule during chromosome attachment to the spindle in newt lung cells. *J. Cell Biol.* **110,** 81–95 (1990).
18. Maiato, H., Rieder, C. L. & Khodjakov, A. Kinetochore-driven formation of kinetochore fibers contributes to spindle assembly during animal mitosis. *J. Cell Biol.* **167,** 831–840 (2004).
19. Tulu, U. S., Fagerstrom, C., Ferenz, N. P. & Wadsworth, P. Molecular requirements for kinetochore-associated microtubule formation in mammalian cells. *Curr. Biol.* **16,** 536–541 (2006).
20. Indjeian, V. B., Stern, B. M. & Murray, A. W. The centromeric protein Sgo1 is required to sense lack of tension on mitotic chromosomes. *Science* **307,** 130–133 (2005).
21. Li, X. & Nicklas, R. B. Mitotic forces control a cell-cycle checkpoint. *Nature* **373,** 630–632 (1995).
22. Waters, J. C., Chen, R.-H., Murray, A. W. & Salmon, E. D. Localization of Mad2 to kinetochores depends on microtubule attachment, not tension. *J. Cell Biol.* **141,** 1181–1191 (1998).

©2007 Nature Publishing Group

23. Cimini, D. *et al.* Merotelic kinetochore orientation is a major mechanism of aneuploidy in mitotic mammalian tissue cells. *J. Cell Biol.* **153,** 517–528 (2001).
24. Rieder, C. L., Cole, R. W., Khodjakov, A. & Sluder, G. The checkpoint delaying anaphase in response to chromosome monoorientation is mediated by an inhibitory signal produced by unattached kinetochores. *J. Cell Biol.* **130,** 941–948 (1995).
25. Rieder, C. L., Schultz, A., Cole, R. & Sluder, G. Anaphase onset in vertebrate somatic cells is controlled by a checkpoint that monitors sister kinetochore attachment to the spindle. *J. Cell Biol.* **127,** 1301–1310 (1994).
26. Cassimeris, L., Rieder, C. L. & Salmon, E. D. Microtubule assembly and kinetochore directional instability in vertebrate monopolar spindles: implications for the mechanism of chromosome congression. *J. Cell Sci.* **107,** 285–297 (1994).
27. Kapoor, T. M. *et al.* Chromosomes can congress to the metaphase plate before biorientation. *Science* **311,** 388–391 (2006).
28. Magidson, V., Loncarek, J., Hergert, P., Rieder, C. L. & Khodjakov, A. in *Laser Manipulations of Cells and Tissues* (eds Berns, M. W. & Greulich, K. O.) 237–266 (Elsevier, 2007).

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

Acknowledgements We thank B. McEwen, C. Rieder and V. Magidson for fruitful discussions and critical reading of the manuscript. This work was supported by grants from the National Institutes of Health grants (to A.K. and T.M.K.). Assembly of our laser microsurgery system was supported in part by a Nikon/MBL fellowship (A.K.). We acknowledge the use of Wadsworth Centre's electron microscopy core facility.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.K. (khodj@wadsworth.org) or T.M.K. (kapoor@rockefeller.edu).

©2007 Nature Publishing Group

METHODS

Live-cell microscopy and laser ablations. PtKG-23 (a clone of PtK_1 line, constitutively expressing γ-tubulin/GFP[13]) and PtKαT (a clone of PtK_2 line, constitutively expressing α-tubulin/GFP[12]) cells were maintained in Ham's F12 medium (Invitrogen) supplemented with 10% fetal calf serum at 37 °C, 5% CO_2. For laser microsurgery, cells grown on 24 mm × 24 mm coverslips were mounted in Rose chambers. Temperature of the chambers on the microscope stage was maintained between 35 and 37°C by a custom-built electronically controlled stage-heater

Monastrol, synthesized as described earlier[29], was kept as 100-mM stock in DMSO and used at 100 μM. nocodazole (Sigma) was kept as 10 mM stock in DMSO and used at 5 μM. For drug washout, the top coverslip of the Rose chamber was removed and cells (on the bottom coverslip) were washed with 5–7 changes of full growth medium. To unambiguously identify the cell of interest, its position was marked using a diamond objective-scribe (Zeiss).

Laser ablation of centrosomes was performed as described elsewhere[14,30]. Note that, after laser ablation and drug-washout, time-lapse sequences were recorded not on the laser-microsurgery microscope but on a different similarly equipped Nikon Eclipse TE2000 or TE300 microscope. This was necessary to make the laser-microsurgery workstation available for other experiments during rather lengthy periods of time-lapse recordings. Time-lapse image sequences were recorded at 30-s or 60-s intervals.

Immunofluorescence. For microtubule and NuMA staining, cells were fixed with 1% glutaraldehyde in PEM buffer (100 mM PIPES, 3 mM $MgCl_2$, 5 mM EGTA, pH 6.9), permeabilized with 1% Triton X-100, and reduced with 1 mg ml^{-1} $NaBH_4$. All reagents were purchased from Sigma. For CREST and Mad2 staining, cells were fixed with 3.8% paraformaldehyde in PBS, and then permeabilized and reduced, exactly as for microtubule/NuMA staining. The following primary antibodies were used: mouse anti-α-tubulin (cloneDM1α, Sigma-Aldrich); rabbit anti-NuMA (gift of D. Compton); human serum CREST SH (gift of B. Brinkley); and rabbit anti-Mad2 (BAbCO).

Fluorescence images were recorded with a 60× 1.4 numerical aperture PlanApo lens as *Z*-series at 200-nm steps (detailed in ref. 30). They were deconvolved using SoftWorx deconvolution software (Applied Precision) and presented as maximal intensity projections. Inter-kinetochore distances were measured in three-dimensional deconvolved data sets using SoftWorx.

Electron microscopy. Serial-section electron microscopy analysis was conducted, and surface-rendered three-dimensional models constructed, as described in ref. 31.

29. Mayer, T. U. *et al.* Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* **286**, 971–974 (1999).
30. La Terra, S. *et al.* The de novo centriole assembly pathway in HeLa cells: cell cycle progression and centriole assembly/maturation. *J. Cell Biol.* **168**, 713–720 (2005).
31. Khodjakov, A., Cole, R. W., McEwen, B. F., Buttle, K. F. & Rieder, C. L. Chromosome fragments possessing only one kinetochore can congress to the spindle equator. *J. Cell Biol.* **136**, 229–240 (1997).

©2007 Nature Publishing Group

LETTERS

How kinesin waits between steps

Teppei Mori[1], Ronald D. Vale[2] & Michio Tomishige[1,2]

Kinesin-1 (conventional kinesin) is a dimeric motor protein that carries cellular cargoes along microtubules[1,2] by hydrolysing ATP[3] and moving processively in 8-nm steps[4]. The mechanism of processive motility involves the hand-over-hand motion of the two motor domains ('heads')[5–7], a process driven by a conformational change in the neck-linker domain of kinesin[8–12]. However, the 'waiting conformation' of kinesin between steps remains controversial[13–16]—some models propose that kinesin adopts a one-head-bound intermediate[17–21], whereas others suggest that both the kinesin heads are bound to adjacent tubulin subunits[7,22,23]. Addressing this question has proved challenging, in part because of a lack of tools to measure structural states of the kinesin dimer as it moves along a microtubule. Here we develop two different single-molecule fluorescence resonance energy transfer (smFRET) sensors to detect whether kinesin is bound to its microtubule track by one or two heads. Our FRET results indicate that, while moving in the presence of saturating ATP, kinesin spends most of its time bound to the microtubule with both heads. However, when nucleotide binding becomes rate-limiting at low ATP concentrations, kinesin waits for ATP in a one-head-bound state and makes brief transitions to a two-head-bound intermediate as it walks along the microtubule. On the basis of these results, we suggest a model for how transitions in the ATPase cycle position the two kinesin heads and drive their hand-over-hand motion.

The first FRET sensor for distinguishing one-head-bound from two-head-bound states is a kinesin heterodimer in which one polypeptide chain contains a single cysteine residue in the plus-end-oriented tip of the catalytic core (residue 215), and the other chain contains a single cysteine residue in the minus-end-oriented base of the core (residue 43) (sensor termed 215–43; Fig. 1a). The second sensor is a kinesin homodimer in which a cysteine residue was introduced in both chains at the beginning of the neck linker (residue 324) (sensor termed 324–324).

To test our FRET sensors, we first examined the FRET efficiency in a kinesin dimer bound statically to a microtubule with the non-hydrolysable nucleotide analogue AMP-PNP, a nucleotide state in which both kinesin heads are bound to the microtubule[10,17,21,22]. Maleimide-modified Cy3 (donor dye) and Cy5 (acceptor dye) were reacted with the two cysteine residues in these FRET constructs, and single kinesin molecules that contained both Cy3 and Cy5 were selected for smFRET observations with total-internal-reflection fluorescence microscopy[12]. SmFRET efficiencies for the 215–43 sensor from individual microtubule-bound heterodimers showed a bimodal distribution of low (about 10%) and high (about 90%) FRET efficiencies (Fig. 1b), as expected if the two kinesin heads are bound to adjacent tubulin subunits 8 nm apart along a microtubule protofilament (Fig. 1a and Supplementary Figs 1 and 2). Our previous FRET measurements with the 215–43 sensor (Fig. 2e in ref. 12) indicated that the high-FRET peak (about 90%) corresponds to the 43 dye on the leading head and the 215 dye on the trailing head and the low-FRET peak (about 10%) corresponds to the opposite dye

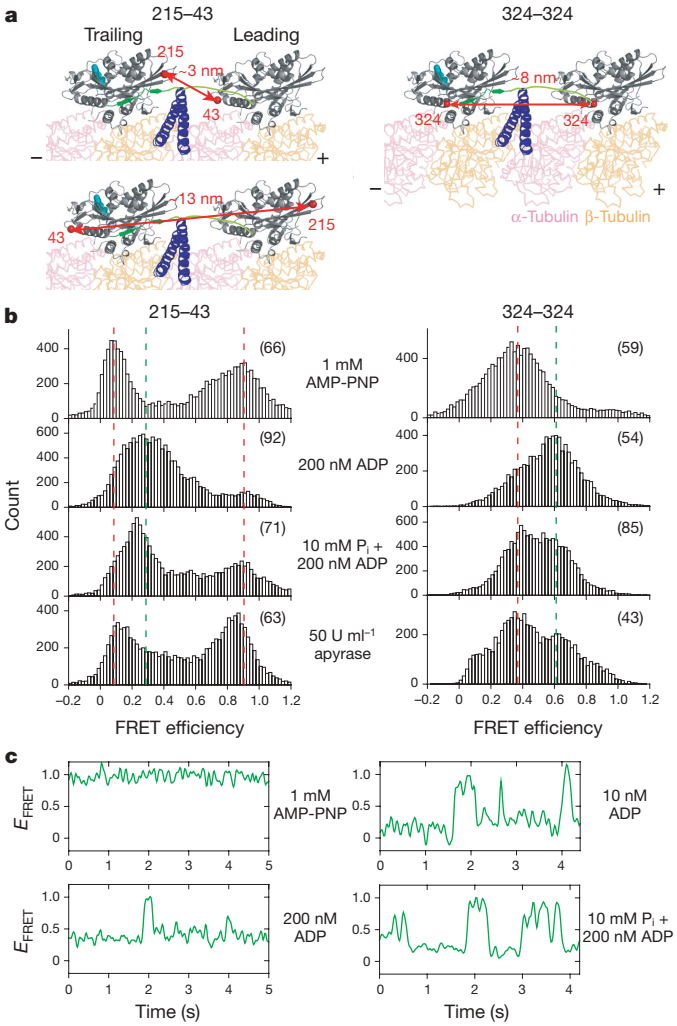


Figure 1 | SmFRET observations of head–head configuration of kinesin under various nucleotide conditions. a, Diagrams of the two-head-bound intermediate state of the kinesin dimer on the microtubule. Positions of cysteine residues for dye labelling are shown in red. The neck linker, neck coiled-coil and bound nucleotide are shown in green, blue and cyan respectively. **b**, Histograms of FRET efficiencies (from each frame of images) of dye-labelled kinesin bound to the axonemes with 1 mM AMP-PNP (ATP-like state), 200 nM ADP, 200 nM ADP/10 mM P_i or 50 U ml^{-1} apyrase (nucleotide-free state). The numbers of molecules analysed are shown in parentheses. Dotted lines illustrate peaks characteristic of putative two-head-bound (red) and one-head-bound (green) states. The correlation of FRET efficiencies with distance measurements is discussed in Supplementary Fig. 1. **c**, Examples of a FRET efficiency (E_{FRET}) trace of individual axoneme-bound 215–43 heterodimer kinesin. ADP and ADP/P_i traces occasionally showed abrupt and large FRET changes (see Supplementary Fig. 4 for additional traces).

[1]Department of Applied Physics, The University of Tokyo, Tokyo 113-8656, Japan. [2]Howard Hughes Medical Institute and Department of Cellular and Molecular Pharmacology, University of California, San Francisco, California 94158, USA.

©2007 Nature Publishing Group

positioning. The unimodal smFRET distribution centred at about 35% for the 324–324 sensor (Fig. 1b) also is consistent with a two-head-bound state. Binding of the two kinesin heads along a single protofilament (as opposed to sideways across protofilaments) is supported by additional FRET experiments with a 149–324 sensor (Supplementary Fig. 2).

We next examined smFRET under different nucleotide conditions. With ADP occupying the active site, kinesin is in a weak microtubule-binding state[3,18,21]; however, at low ADP concentrations (200 nM) single kinesin dimers remained bound to microtubules for more than 10 s. The FRET histograms at 200 nM ADP clearly differed from those described above for AMP-PNP (Fig. 1b), becoming unimodal and centred at about 30% for the 215–43 sensor and shifting from 35% (AMP-PNP) to 60% for the 324–324 sensor. The result from the 324–324 sensor suggests that the two kinesin heads come closer together. As verified in the experiments described below, these FRET distributions with ADP reflect a one-head-bound state. We next added excess inorganic phosphate (10 mM P_i) to 200 nM ADP, which probably results in partial occupancy of an ADP·P_i state in the tethered head[23]. Under this condition, FRET peaks characteristic of a two-head-bound state appeared (Fig. 1b). Similar results were obtained after the addition of 1 mM AlF_4^- to 1 mM ADP (Supplementary Fig. 3). These results suggest that the addition of P_i or AlF_4^- to ADP-kinesin shifts the equilibrium from a one-head-bound state to a two-head-bound state, as described previously[23]. When both kinesin heads are nucleotide-free (50 U ml^{-1} apyrase; see Methods), the FRET distributions for the 215–43 and 324–324 sensors were generally similar to that observed with AMP-PNP, but with broader distributions (Fig. 1b), suggesting that nucleotide-free kinesin primarily adopts a two-head-bound state (see also ref. 22) with partial occupancy of a one-head-bound state.

Interestingly, whereas the FRET signal of the 215–43 sensor in the presence of 1 mM AMP-PNP was fairly constant, a subset of molecules (about 25%) with ADP or ADP/P_i (Fig. 1c and Supplementary Fig. 4), or under nucleotide-free conditions (not shown) underwent abrupt FRET transitions (from about 30% to either 90% or 10% FRET), which most probably represent unbinding and rebinding events of one of the kinesin heads with the microtubule.

The results from our ADP FRET experiments suggest that kinesin can adopt a one-head-bound state on the microtubule. To obtain additional evidence for this hypothesis, we engineered a kinesin heterodimer in which only one head could bind to microtubules under all nucleotide conditions (mutation Y274A/R278A/K281A in loop 12, termed 'L12-triple')[24,25]. We made two different versions of this heterodimer. SmFRET measurements with a heterodimer containing a 215-labelled dye in the wild-type (WT) polypeptide and a 43-labelled dye in the L12-triple polypeptide, namely 215(WT)–43(L12), produced a unimodal distribution centred at about 30% FRET efficiency with 200 nM ADP (Fig. 2a). Another heterodimer in which the dye positions were reversed in the wild-type and L12-triple chains, namely 215(L12)–43(WT), also showed similar FRET efficiency, indicating that the distances between 43-labelled and 215-labelled dyes are similar in these two configurations. A similar result was obtained in the nucleotide-free state (Fig. 2a), in contrast with the bimodal FRET distribution in wild-type kinesin dimer after nucleotide depletion (Fig. 1b). These results further support the interpretation that a 30% FRET signal of the 215–43 sensor in the presence of ADP is diagnostic of a one-head-bound state.

The FRET histogram for the WT/L12-triple heterodimer changed markedly after the addition of AMP-PNP. For the 215(WT)–43(L12) heterodimer, the histogram became bimodal with the primary peak now emerging at 80% FRET efficiency; for the 215(L12)–43(WT)

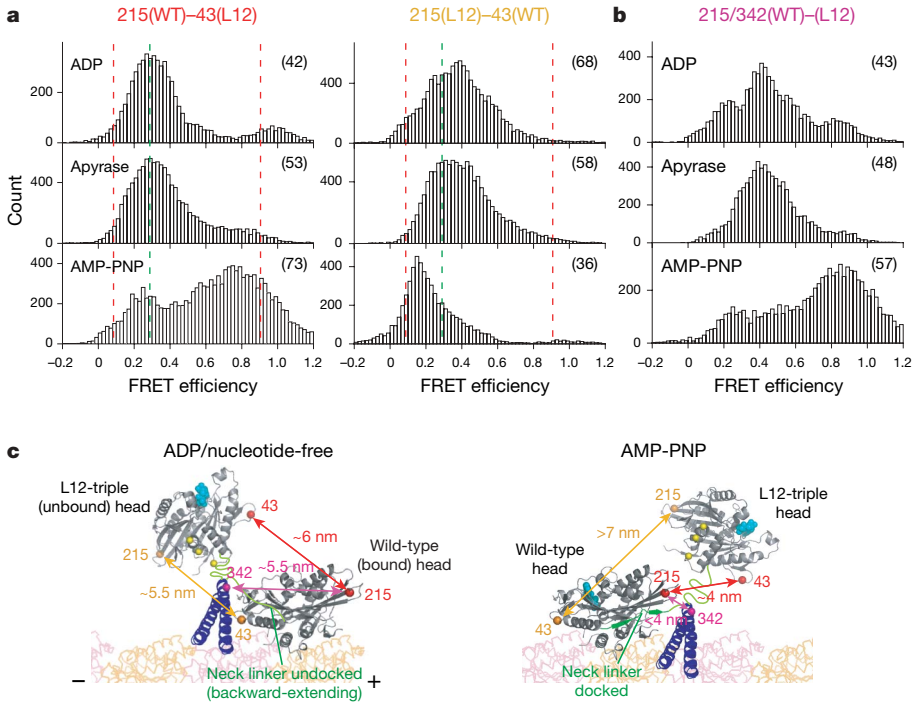


Figure 2 | SmFRET observations of mutant heterodimeric kinesin that persistently takes one-head-bound state. **a**, Histograms of FRET efficiencies of axoneme-bound WT/L12-triple heterodimer kinesin (the L12-triple mutant (yellow spheres in **c**) cannot bind microtubules[24]). The 215 and 43 cysteine residues were introduced into wild-type and L12-triple heads, respectively (left) or into L12-triple and wild-type heads, respectively (right). Dotted lines are adopted from Fig. 1b. **b**, FRET histograms with 215/342 dyes on the wild-type chain (L12-triple chain is cysteine-light)[12] to probe neck-linker conformations in the bound head. The FRET efficiency peaks in the ADP/apyrase and AMP-PNP states are similar to those observed for the wild-type kinesin dimer with AMP-PNP (backward-extending and forward-extending (docked) neck linkers, respectively; Supplementary Fig. 8). **c**, Proposed model for the nucleotide-dependent configurations of the WT/L12-triple heterodimer, based on the FRET results.

©2007 Nature Publishing Group

heterodimer, the major FRET peak shifted in the opposite direction towards lower FRET efficiencies (Fig. 2a). These large nucleotide-dependent FRET changes suggest a movement of the L12-triple (unbound) head towards the plus-end-oriented tip (residue 215) of the wild-type (bound) head on binding of AMP-PNP (Fig. 2c). This translation of the unbound head from a rear position to a forward position is most probably driven by a nucleotide-dependent docking of the neck linker[8,9] in the bound head (Fig. 2b), the implications of which will be discussed later.

We next examined the smFRET signals for dimeric kinesin moving processively along a microtubule. At saturating ATP concentration (1 mM), we could measure only an average FRET efficiency of moving kinesin molecules, but could not accurately detect conformational transitions that might occur during kinesin stepping, because the stepping rate of kinesin (about 50 s^{-1}) is comparable to the camera acquisition rate (100 frames s^{-1}). The FRET efficiency for the 215–43 sensor showed a broad distribution centred at about 50% (Fig. 3a, b), a value consistent with an average of the bimodal 10%, 90% FRET distribution observed for static two-head-bound kinesin with AMP-PNP (and differing from the 30% value of one-head-bound kinesin, Fig. 1b). The 324–324 sensor yielded a unimodal FRET efficiency centred at about 30% with 1 mM ATP (Fig. 3a, b), which is also more similar to the distribution observed with the same FRET sensor with AMP-PNP than to that observed with ADP (Fig. 1b). Thus, the results from these two FRET sensors suggest that kinesin spends most of the time bound with two heads to the microtubule when moving at saturating ATP concentration (also see Supplementary Fig. 5), which is consistent with previously published fluorescence polarization measurements[22].

We next studied head–head smFRET at subsaturating ATP concentration (2 μM). When compared with the 1 mM ATP histograms, the main FRET efficiency peak for the 215–43 sensor shifted to about 30%, whereas the peak for the 324–324 sensor shifted to about 60% (Fig. 3a, b). These FRET distributions are more similar to those observed with 200 nM ADP than to those observed with AMP-PNP (Fig. 1b). These results suggest that kinesin waits primarily as a one-head-bound intermediate when ATP binding becomes the rate-limiting step in the ATPase cycle.

With the longer dwell times at low ATP concentration, one might expect to observe FRET transitions that reveal how kinesin steps along the microtubule. At 2 μM ATP, moving 215–43 dye-labelled kinesin molecules ($n = 33$) spent most time in a roughly 30% FRET state (probably a one-head-bound state) with brief spikes (dwell time 34 ± 2 ms (mean ± s.e.m.)) towards higher (about 80%) FRET values (Fig. 3c, d, and Supplementary Figs 6 and 7). These higher FRET states could represent a transient two-head-bound intermediate state (similar to the 90% FRET state observed with AMP-PNP in which the 215-labelled head is in the trailing position; Fig. 1a, b). In this case, transitions from 30% to a lower FRET state (similar to the 10% FRET state with AMP-PNP) should also occur. Although such transitions were occasionally evident (Fig. 3c, arrowheads), the magnitude of such a FRET change is difficult to distinguish from the noise. However, the dwell times at the 30% FRET state (that is, the period between adjacent high-FRET spikes) provided additional support for the idea that transitions to both lower-FRET and higher-FRET states exist. This dwell-time histogram was best fitted by a convolution of two exponentials[7] (Fig. 3d, dotted line), suggesting that two rate-limiting ATP binding events occur between the two high-FRET spikes. Moreover, the mean dwell time derived from this double-exponential fit (140 ± 8 ms (mean ± s.e.m.)) is comparable to that predicted (160 ms) from the kinesin step size (8.3 nm)[4] and velocity at 2 μM ATP (51 ± 22 nm s^{-1}, $n = 33$). In addition, the total number of high FRET spikes from several single-molecule traces (272 spikes, $n = 33$) divided by the distance travelled by these molecules (4.57 μm) yielded an average travel distance of about 17 nm per high-FRET spike, which is close to double the kinesin step size (Fig. 3e). Collectively, these results suggest that a kinesin step at low ATP concentrations involves a short-lived, two-head-bound state (for

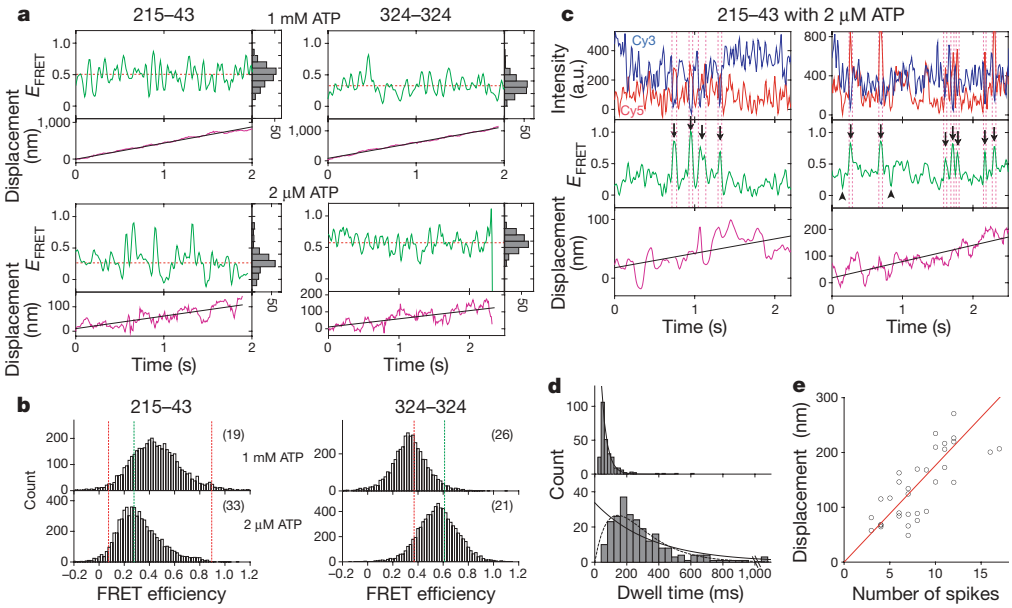


Figure 3 | Head–head configuration changes of kinesin while moving along microtubules. **a**, Examples of time traces of FRET efficiency, E_{FRET} (green; red dotted lines show the mean value; histogram at the right), and fluorophore centroid displacements (purple; black lines are a linear fit) of 215–43 or 324–324 dye-labelled kinesin moving along axonemes in the presence of 1 mM or 2 μM ATP. **b**, Histograms of FRET efficiencies (each frame) under the two different ATP conditions. Dotted lines are adopted from Fig. 1b. **c**, Typical traces of fluorescence intensities of donor (Cy3, blue) and acceptor (Cy5, red) fluorophores, FRET efficiency and axial displacement, for 215–43 heterodimer kinesin at 2 μM ATP, showing transitions between one-head-bound and two-head-bound states. FRET changes greater than 0.3 are marked by vertical magenta dotted lines, with transitions towards high and low FRET states indicated by arrows and arrowheads, respectively. **d**, Distributions of dwell times in the high-FRET state (top) and between the high-FRET spikes (arrows in **c**) (bottom). Solid and dotted lines show single-exponential and double-exponential fits[7]. **e**, Relationship between number of transitions towards the high-FRET state and the displacement per observation time; each point represents a different single molecule. The solid red line shows a linear fit (17.6 nm per transition).

©2007 Nature Publishing Group

example, a high-FRET spike), which then undergoes a transition to a longer-lived, one-head-bound state (30% FRET state).

Whether kinesin waits as a one-head-bound or a two-head-bound intermediate between steps has been debated[13–16]. Our smFRET experiments suggest that both modes of movement are possible. At high ATP concentration (when the detachment of the trailing head—triggered by ATP hydrolysis/phosphate release—is likely to be rate-limiting[20]), kinesin moves hand-over-hand, making rapid transitions from one two-head-bound state to the next (Fig. 4a). However, when ATP binding to the lead, nucleotide-free, head becomes rate-limiting at low ATP concentrations (Fig. 4b), the trailing head releases its P_i and detaches from the microtubule, producing a long-lived one-head-bound configuration with the neck linkers in two different conformational states as described previously[12] (Supplementary Fig. 8). This latter pathway is a variation of the original model proposed by Rice *et al.*[8,9], with ATP binding to the leading head taking place after the transition to a one-head-bound state. Alonso *et al.*[21] recently proposed that kinesin waits between steps with the detached, ADP-bound head in front of the bound, nucleotide-free head. However, our FRET experiments with the WT/L12-triple heterodimer suggest that the detached head is located rearward on average, shifting to a forward position on binding of ATP to the bound head (Fig. 2). The rear positioning of the detached head also might explain an apparent inconsistency in the one-head waiting-state model with high spatial resolution tracking experiments[13–16], which showed an ATP-driven roughly 17-nm step of a kinesin head[7] (a distance consistent with a detached head being positioned near the rear tubulin-binding site; Supplementary Fig. 9).

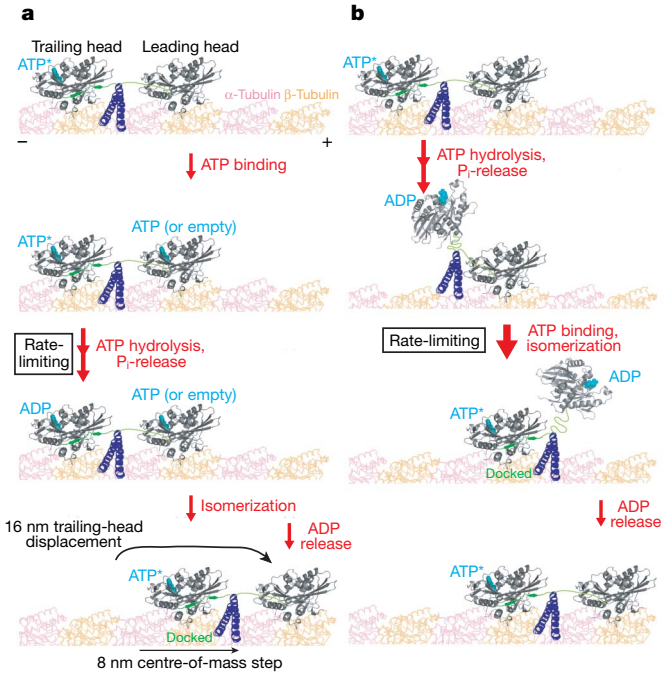


Figure 4 | A model for kinesin motility. a, At high ATP concentrations, hydrolysis and phosphate release leading to trailing-head dissociation is rate-limiting (larger arrows represent slower steps). Once released, the trailing head is rapidly translated forwards by ATP isomerization-induced neck-linker docking in the partner head (indicated by ATP*), resulting in a rapid 8-nm step and a new two-head-bound intermediate. **b**, At low ATP concentrations, phosphate release occurs before the leading head binds an ATP molecule, producing a one-head-bound state in which the detached head is positioned behind the bound partner head. Although this head may interact transiently with the rear tubulin-binding site, we postulate that the conformation of the neck linker in this head is not compatible with ADP release and strong microtubule binding. Once ATP has bound to the forward head, the detached head translates forwards and, after a diffusional search, binds to a new tubulin subunit. ADP release from this head is now permitted by the backward-extending conformation of the neck linker in this two-head-bound state[12,22,23,28].

Our experiments also shed light on how the ATPase cycles in the two kinesin heads are coordinated during processive motion. The gating model of Alonso *et al.*[21] proposes that the detached head in the 'waiting state' is parked in front of the bound head but is in a conformation that prevents its binding to tubulin. However, transient interactions of the 'detached' head with the microtubule are seen in our low-ADP FRET data (Fig. 1c) and in ATPase kinetic measurements made by Hackney[23]. Thus, an additional mechanism must keep the detached head from progressing through its ATPase cycle until its partner head binds ATP. We suggest that the detached head, positioned behind the nucleotide-free, microtubule-bound head, will not release its bound ADP when it interacts with the rear tubulin-binding site. However, ADP release can occur after the detached head has been translated to a forward tubulin-binding site by ATP-triggered neck-linker docking in the bound partner head. This model is also supported by results from Guydosh & Block[26], who showed that nucleotide dissociation occurs only when a head is in the 'forward' position. We suggest that this position dependence in the ADP dissociation rate is controlled by the conformation of the neck linker, with a forward-pointing conformation of this mechanical element inhibiting ADP release in the trailing head (Fig. 4b). Further studies will be required for a better understanding of how structural states of the neck linker affect transitions in the ATPase cycle.

METHODS SUMMARY

DNA cloning and protein purification. Cysteine residues[12] and/or mutations[24] were introduced into a 'cysteine-light' human ubiquitous kinesin-1 dimer 490 amino-acid-residues long. Heterodimers were prepared by using a coexpression vector (one kinesin containing a carboxy-terminal streptavidin tag (Strep-tag) and the second with a His_6 tag) and purified by two-step affinity chromatography of nickel-nitrilotriacetic acid and Strep-Tactin[12]. Dialysed kinesin was reacted for 4 h at 4 °C with Cy3-maleimide and Cy5-maleimide at a motor head/Cy3 dye/Cy5 dye molar ratio of 1:10:10. Unreacted dyes were quenched with 1 mM dithiothreitol and then removed through microtubule affinity purification[12]. The 215–43 and 324–324 dual dye-labelled kinesins showed single-molecule processive movement with normal maximal velocity[27] (mean velocities at 1 mM ATP concentration were $410 \pm 60\ \mathrm{nm\,s^{-1}}$ (mean ± s.d.; $n = 19$) for 215–43, and $430 \pm 80\ \mathrm{nm\,s^{-1}}$ ($n = 26$) for 324–324), indicating that the substitutions, mutations and dye-labelling did not significantly alter motor activity.

SmFRET. SmFRET measurements were performed with a custom-built prism-type laser-illuminated total-internal-reflection fluorescence microscope[12]. Dye-labelled kinesins were imaged either moving along sea-urchin axonemes in the presence of ATP and an ATP-regenerating system[12] or statically bound in the presence of AMP-PNP, ADP ($5\ \mathrm{U\,ml^{-1}}$ hexokinase converted contaminating ATP (about 10 nM) to ADP), P_i (10 mM KH_2PO_4) or apyrase at about 22 °C. High concentrations ($50\ \mathrm{U\,ml^{-1}}$) of apyrase were required to deplete residual ATP and ADP (about 10 nM) completely; with a lower concentration ($5\ \mathrm{U\,ml^{-1}}$) of apyrase we observed a histogram distribution the peak value of which was intermediate between those with $50\ \mathrm{U\,ml^{-1}}$ apyrase and 200 nM ADP (that is, both one-head-bound and two-head-bound states).

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

Received 18 April; accepted 2 October 2007.
Published online 14 November 2007.

1. Vale, R. D. The molecular motor toolbox for intracellular transport. *Cell* **112**, 467–480 (2003).
2. Hirokawa, N. & Takemura, R. Molecular motors and mechanisms of directional transport in neurons. *Nature Rev. Neurosci.* **6**, 201–214 (2005).
3. Hackney, D. D. Evidence for alternating head catalysis by kinesin during microtubule-stimulated ATP hydrolysis. *Proc. Natl Acad. Sci. USA* **91**, 6865–6869 (1994).
4. Svoboda, K., Schmidt, C. F., Schnapp, B. J. & Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **365**, 721–727 (1993).
5. Asbury, C. L., Fehr, A. N. & Block, S. M. Kinesin moves by an asymmetric hand-over-hand mechanism. *Science* **302**, 2130–2134 (2003).
6. Kaseda, K., Higuchi, H. & Hirose, K. Alternate fast and slow stepping of a heterodimeric kinesin molecule. *Nature Cell Biol.* **5**, 1079–1082 (2003).
7. Yildiz, A., Tomishige, M., Vale, R. D. & Selvin, P. R. Kinesin walks hand-over-hand. *Science* **303**, 676–678 (2004).

©2007 Nature Publishing Group

8. Vale, R. D. & Milligan, R. A. The way things move: looking under the hood of molecular motor proteins. *Science* **288**, 88–95 (2000).
9. Rice, S. *et al.* A structural change in the kinesin motor protein that drives motility. *Nature* **402**, 778–784 (1999).
10. Skiniotis, G. *et al.* Nucleotide-induced conformations in the neck region of dimeric kinesin. *EMBO J.* **22**, 1518–1528 (2003).
11. Asenjo, A. B., Weinberg, Y. & Sosa, H. Nucleotide binding and hydrolysis induces a disorder–order transition in the kinesin neck-linker region. *Nature Struct. Mol. Biol.* **13**, 648–654 (2006).
12. Tomishige, M., Stuurman, N. & Vale, R. D. Single-molecule observations of neck linker conformational changes in the kinesin motor protein. *Nature Struct. Mol. Biol.* **13**, 887–894 (2006).
13. Carter, N. J. & Cross, R. A. Kinesin's moonwalk. *Curr. Opin. Cell Biol.* **18**, 61–67 (2006).
14. Valentine, M. T. & Gilbert, S. P. To step or not to step? How biochemistry and mechanics influence processivity in Kinesin and Eg5. *Curr. Opin. Cell Biol.* **19**, 75–81 (2007).
15. Hackney, D. D. Processive motor movement. *Science* **316**, 58–59 (2007).
16. Block, S. M. Kinesin motor mechanics: binding, stepping, tracking, gating, and limping. *Biophys. J.* **92**, 2986–2995 (2007).
17. Kawaguchi, K. & Ishiwata, S. Nucleotide-dependent single- to double-headed binding of kinesin. *Science* **291**, 667–669 (2001).
18. Uemura, S. *et al.* Kinesin-microtubule binding depends on both nucleotide state and loading direction. *Proc. Natl Acad. Sci. USA* **99**, 5977–5981 (2002).
19. Carter, N. J. & Cross, R. A. Mechanics of the kinesin step. *Nature* **435**, 307–312 (2005).
20. Auerbach, S. D. & Johnson, K. A. Alternating site ATPase pathway of rat conventional kinesin. *J. Biol. Chem.* **280**, 37048–37060 (2005).
21. Alonso, M. C. *et al.* An ATP gate controls tubulin binding by the tethered head of kinesin-1. *Science* **316**, 120–123 (2007).
22. Asenjo, A. B., Krohn, N. & Sosa, H. Configuration of the two kinesin motor domains during ATP hydrolysis. *Nature Struct. Biol.* **10**, 836–842 (2003).
23. Hackney, D. D. The tethered motor domain of a kinesin-microtubule complex catalyzes reversible synthesis of bound ATP. *Proc. Natl Acad. Sci. USA* **102**, 18338–18343 (2005).
24. Woehlke, G. *et al.* Microtubule interaction site of the kinesin motor. *Cell* **90**, 207–216 (1997).
25. Kaseda, K., Higuchi, H. & Hirose, K. Coordination of kinesin's two heads studied with mutant heterodimers. *Proc. Natl Acad. Sci. USA* **99**, 16058–16063 (2002).
26. Guydosh, N. R. & Block, S. M. Backsteps induced by nucleotide analogs suggest the front head of kinesin is gated by strain. *Proc. Natl Acad. Sci. USA* **103**, 8054–8059 (2006).
27. Tomishige, M., Klopfenstein, D. R. & Vale, R. D. Conversion of Unc104/KIF1A kinesin into a processive motor after dimerization. *Science* **297**, 2263–2267 (2002).
28. Sablin, E. P. & Fletterick, R. J. Coordination between motor domains in processive kinesins. *J. Biol. Chem.* **279**, 15707–15710 (2004).

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

Acknowledgements We thank M. Nakajima for support in cloning; H. Tadakuma and N. Stuurman for microscope construction and technical support; C. Shingyoji for the gift of sea urchin sperm; K. Kikuchi for the tracking program; and A. Yildiz and A. Carter for discussions. M.T. is supported by Grant-in-Aid for Scientific Research on Priority Areas from MEXT, Japan, a Research Grant for Young Investigators from the Human Frontier Science Program, and grants from the Mitsubishi Foundation and the Asahi Glass Foundation. R.D.V. is supported by grants from the Howard Hughes Medical Institute and the US National Institutes of Health.

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

©2007 Nature Publishing Group

METHODS

FRET imaging and data collection. Donor (Cy3) and acceptor (Cy5) dyes were excited with an argon laser (514 nm; 35LAP321; Melles Griot) and a diode laser (635 nm; Radius 635; Coherent Inc.), respectively. Fluorescence images from Cy3 and Cy5 (or FRET) were separated by using a Dual-View (Optical Insights) and then projected side-by-side on an electron-multiplying charge-coupled device camera (iXon DV860 DCS-BV; Andor). For the static FRET measurements, both Cy3 and Cy5 fluorophores were illuminated directly by sequential excitation with the argon laser (0.8 mW) and the laser diode (1.5 mW) at an acquisition rate of 50 frames s^{-1}. Both lasers were used to identify molecules that showed little or no FRET (to ensure that the acceptor dye was present but not being excited by the donor dye; an example is given in Supplementary Fig. 4)[12]. For the dynamic FRET measurements, images were taken by the excitation with the argon laser (3 mW) at an acquisition rate of 100 frames s^{-1}.

Data analysis. Images were analysed with Image J (http://rsb.info.nih.gov/ij/) with custom-designed plug-in software. For the static FRET measurements, we first identified Cy3–Cy5 dual-labelled motors on axonemes by using the time-averaged images (100 frames) for donor and acceptor excitations[12]. Axonemes that were sparsely coated with fluorescence spots (less than one spot per 2-μm microtubule) were used for data analysis to reduce the chance that two distinct Cy3 and Cy5 spots happened to localize together on the axonemes within one pixel (80 nm). Average FRET efficiencies were determined for each molecule as described previously[12]. FRET efficiencies were also determined for each frame (in both static and dynamic FRET measurements) as described previously[12], except that we first applied a three-frame gaussian filter to the intensity traces derived from raw images and further applied twice the three-frame gaussian filter to the calculated FRET efficiency traces. In the presence of AMP-PNP (under which conditions kinesin dimer stably takes a two-head-bound state), the FRET efficiency was fairly constant and we did not see transition during the observation time (about 5 s) (Fig. 1c and Supplementary Fig. 4). In this case, we used the averaged FRET efficiency of individual molecules to make histograms (only for those in Supplementary Fig. 2). In contrast, with other nucleotide conditions (ADP, ADP/P_i and nucleotide-free) we often saw transitions (between one-headed and two-headed states). In these cases, FRET efficiency based on individual molecules does not properly represent individual states, so we used histograms based on individual frames to highlight each state that the kinesin dimer takes and its relative frequency (durations).

In the dynamic FRET measurements, the centroid positions of the fluorescence spots were tracked by using a gaussian fitting algorithm with images applied with ten-frame running averaging. Fluorescence spots that did not show clear unidirectional movement were not subjected to further data analysis. Transitions to a high-FRET-efficiency state ('spikes' for the 215–43 dye pair at 2 μM ATP) were identified manually as anticorrelated donor–acceptor intensity changes that accompanied FRET efficiency changes of more than 0.3. This cutoff amplitude of 0.3 was significantly greater than the noise level of FRET efficiency in our measurement with the acquisition rate of 100 frames s^{-1} (0.08 ± 0.02, $n = 20$), which was estimated as the s.d. of the FRET efficiency fluctuations in the traces for 43–215 dyes on one head[12] moving along axonemes in the presence of 2 μM ATP (Supplementary Fig. 6). The dwell time in the high-FRET state (Fig. 3d) was defined as the duration between the midpoints in the ascending and descending phases, and the dwell time in the median FRET state was defined as the duration between two adjacent transitions to the high-FRET state. The accuracy of determining the dwell time was about 10 ms (one frame).

©2007 Nature Publishing Group

naturejobs

JOBS OF THE WEEK

Tenure isn't what it used to be in the United States. Although the number of US university and faculty positions has doubled since 1975, the proportion of tenured and tenure-track positions has fallen. Temporary 'adjunct' professors — both full- and part-time — were once the minority in public and private universities. Now they make up the majority, according to an analysis of US Department of Education data conducted by the American Association of University Professors (AAUP). In 1975, adjunct professors made up 43% of the US faculty. In 2005, they accounted for nearly 70% of professorships.

This shift to a part-time workforce on short-term contracts could jeopardize the quality of US higher education and deter people from pursuing an academic career, says Gwendolyn Bradley, spokeswoman for the AAUP. Adjuncts earn a small fraction of the salaries paid to tenured or tenure-track staff, often receive no benefits and sometimes have no access to office space or administrative help. "We're reaching a tipping point where it's not going to be possible to sustain quality higher education — not because part-timers aren't qualified, but because they aren't supported," Bradley says. A decline in state funding for public universities has exacerbated the situation.

But the shift has not gone unnoticed. The American Federation of Teachers, which represents 170,000 faculty members, this year sponsored legislation in 11 states to redress the imbalance. It wants universities to have 75% of undergraduate courses taught by tenured or tenure-track employees, to pay part-timers a proportionate salary rather than a fractional salary of full-time faculty (including benefits), and to allow adjuncts access to full-time positions once they materialize.

Some universities are already taking note. The University of Michigan in Ann Arbor is upping salaries and benefits for its adjuncts, and Rutgers University in New Jersey has pledged to hire 100 tenure-track staff after a labour dispute this summer. But US universities require broader, system-wide change to keep them competitive, both in terms of educational quality and as a desirable career option.

Paul Smaglik is the former editor of *Naturejobs*

CONTACTS
Acting Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street,
London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young
(+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com

France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW:
Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:**
Tom Hancock

***Naturejobs* online production:**
Jasmine Myer
US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

Japan Head Office, Tokyo
Chiyoda Building,
2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe
Tel: +81-3-3267-8765
e-mail: a.watanabe@natureasia.com

©2007 Nature Publishing Group

NATURE|Vol 450|29 November 2007

CELL MATES

Collaborations are more integral to the life sciences than ever — and funders are trying to keep pace. **Jacqueline Ruttimann** reports.

Gaudenz Danuser is an associate professor at the Scripps Research Institute's cell-biology department in La Jolla, California. But he is not a cell biologist — Danuser was formally trained as an engineer. In the past, when he jointly applied for funding from the National Institutes of Health (NIH) to study mathematical modelling of cellular processes via the agency's standard R01 grant, Danuser would have to come to an agreement with his cell-biologist collaborator, as only one could lead the grant. But a new programme at the NIH has made his task easier and potentially more professionally rewarding. The NIH has set up a multiple principal investigator (PI) grant, which means that Danuser can get official recognition for his efforts. Previously, he says, "I was at the mercy of my friends to give me the money."

"The NIH has really taken the lead in putting administrative structures into place that will both recognize and nurture collaborations," says Mary Beckerle, former president of the American Society for Cell Biology and current member of NIH director Elias Zerhouni's advisory committee. Crediting multiple researchers is especially helpful for young tenure-track PIs looking to log experience as independent researchers on their curricula vitae, she says. Failure to document such accomplishments could "have ramifications for his or her academic survival", says Beckerle. Recent fraud cases have highlighted the need for co-authors to be clear about accountability (see *Nature* **450,** 1; 2007).

Crossing boundaries

Interdisciplinary collaborations are encouraged by numerous interdisciplinary institutes and centres, and are becoming increasingly common — there were five times more scientific papers with international co-authors in 2005 than in 1985, according to recently released figures from the Organisation for Economic Co-operation and Development. But finding the right funding agency, and the right funding category within that agency, can be a challenge.

Danuser and his collaborator Peter Sorger, a systems biologist at Harvard University, are two of the first recipients of a multi-PI grant. They are seeking to combine advanced image analysis and genetics to study proteins in the cell-division complex known as the kinetocore. Sorger welcomes the new grant.

"It allows you to fund new things you couldn't do otherwise, identify more easily who's doing what and figure out if the collaboration was a success," he says. But he is concerned that the NIH should not view multi-PI grants as a way to save money, as the total amount is the same as that of the single-PI grants: the NIH would be "paying less to fund the same science".

Still, Sorger sees the grants as a positive step. "Science has a long history of recognizing and glorifying individual achievement," he says. "We need to move to a more sophisticated way in which collaboration is recognized, but not to the detriment of the individual." Sorger likens collaboration to a strong marriage. "It doesn't just happen, you have to work at it," he says.

W. BERG

"Glue grants are an experiment into how to support science. Other investigators may feel threatened."
— Jeremy Berg

Indeed, the challenges are many. Despite the success of the NIH's interdisciplinary Specialized Programs of Research Excellence (SPORE) grants — started by the National Cancer Institute in 1992 — there have been pitfalls. SPORE grants have one director who oversees a minimum of four research projects, each of which has two co-leaders: one for basic research and one for clinical research. The grants, which target translational cancer research, require a clinical trial at the end of five years. It's been a tough target to meet. And when the $2.5-million-a-year SPORE grant is divided up among all participants, it works out to less than the average R01 grant, which is typically $220,000 a year. Even so, the grants have led to 120 phase I and phase II clinical trials and the identification of 110 biomarkers.

Similar challenges persist with the NIH's 'glue grants', so called for their effects in holding a disparate group together. These large-scale awards, initiated in 2000, bring together groups of scientists from different fields to work on a project that lasts up to 10 years and is outside the purview of a traditional R01.

Groups could have difficulty maintaining the resources that have been created once the grants have expired, says Jeremy Berg, director of the National Institute of General Medical Sciences, which started the grant. Then there's the need to find a balance between those investigators on the glue grant and the larger group of others active in the same field. "Investigators may feel threatened by the competition due to the activities of the glue grant," says Berg, who acknowledges that the NIH is actively working to explore these issues.

Berg cites glue-grant successes such as the Consortium for Functional Glycomics, through which a large collection of biologically relevant carbohydrates known as glycans has been collected or synthesized and presented on arrays. Investigators can submit requests and get help to have their samples screened against these

©2007 Nature Publishing Group

DESIGN PICS/CORBIS

arrays. The results are made public via the Internet, enabling other scientists to mine the data for general patterns relevant to cell–cell and other interactions.

Collaborations can make judging grant applications tricky. Funders of collaborative interdisciplinary research must ensure there is broad enough expertise among reviewers of proposals, points out Nicola Perrin, policy adviser at biomedical charity the Wellcome Trust. The trust's awards scheme encourages interdisciplinary research collaborations involving basic scientists, clinicians and non-biologists such as mathematicians, physicists, chemists, engineers and social scientists. Wellcome also funds interdisciplinary centres such as the Wellcome Trust Centre for Cell-Matrix Research at the University of Manchester, UK, which is currently running projects on matrix assembly, adhesion signalling, cell-fate determination and tissue regeneration.

Interdisciplinary translational research projects sometimes have a hard time making it to the clinical-trials stage in a timely manner, says Richard Seabrook, head of business development in the Wellcome Trust's technology-transfer division. To circumvent this, the trust provides advisers to help guide projects on at different stages in their lifespan. Even just articulating a project to funders — given the work's multifaceted, cutting-edge interdisciplinary bent — can be a challenge, says Seabrook. From January 2008, to help people communicate a project's method and objectives, the Wellcome Trust plans to include presentations as part of the grant-application process for translation awards.

After some disappointing early results, the Human Frontier Science Program Organization (HFSPO), an international consortium supporting multidisciplinary research, had to alter its review process for its multi-country grants. Geoff Richards, director of research grants, says that many of the first grant applications paid only lip-service to the interdisciplinary mission. "The review committees have learned to spot projects that contain thinly disguised add-ons," says Richards. The three-year grants fund basic research. Applicants must have a lab in one of the HFSPO's member

T. PIOTROWSKI

"Scientists return to developing countries, bringing this culture of collaboration with them." — Alejandro Sánchez Alvarado

C. NEGUS

Colleagues: Sandy Schmid, Dinah Loerke, Gaudenz Danuser.

countries, which include most European countries, India, Japan, South Korea, Australia, New Zealand Canada and the United States.

Initially, many HFSPO grant applicants had team members often chosen more for their geographical location than for any real contribution they could make to the project, Richards says. By including committee members from, for example, chemistry, physics and biocomputing, he says the programme has developed a review culture that spots most of the cases in which team members are poorly integrated into the proposal. To level the playing field, HFSPO grants are divided into two separate categories: one for teams of young researchers starting out and one for researchers at any stage of their career.

Taking it home

Scientists in the developing world are often eager for collaborations, but can have difficulty finding the right mechanism. Venezuelan-born neurobiologist and Howard Hughes Medical Institute investigator Alejandro Sánchez Alvarado, of the University of Utah in Salt Lake City, thanks the Millennium Science Initiative (MSI), partly funded by the World Bank, for helping him collaborate with scientists in Latin America. The initiative aims to help scientists and engineers from the developing world to do top-level research in their own countries. As Latin American scientists who trained abroad move back home to work in these specially designed research centres, he says, they are "bringing this culture of collaboration with them". Scientists working at MSI centres are encouraged to work both in the public and private sectors, inside and outside their own fields.

Cooperation is key to the success of these large efforts. A weak investigator could potentially hide behind a strong one (see *Nature* **437,** 590–591; 2005). Or burnt-out investigators could join less senior scientists solely to tap a funding resource. The multi-PI grant could help get around such pitfalls, by making it harder for one person to represent a large group.

Proper collaboration may have to come from the ground up. "I don't think you can force collaboration top-down by having grant mechanisms," says Danuser.

But the bodies that find effective ways to support interdisciplinary collaborations are likely both to gain recognition and attract expertise. "Institutions need to remain flexible and nimble as the face of science changes," says Beckerle. "They need to figure out how to fund the best science with the fewest boundaries." ■

Jacqueline Ruttimann is a freelance writer based in Chevy Chase, Maryland.

©2007 Nature Publishing Group

NATURE|Vol 450|29 November 2007

MOVERS

Richard Myers, director of Hudson-Alpha Institute for Biotechnology, Huntsville, Alabama

2002–present: Chair of genetics, Stanford University School of Medicine, California
1993–present: Director, Stanford Human Genome Center, Stanford, California
1985–93: Associate professor, University of California, San Francisco

Richard Myers is no fan of the tenure system. "You have to prove yourself individually from day one," he says. Myers also discovered early on that the quest for tenure could put undue pressure on one's career choices. During his first academic job, as an associate professor at the University of California, San Francisco, colleagues warned the young biochemist that working with others could damage his tenure bid. Nevertheless, Myers began working with his colleague David Cox.

The partnership had the opposite career effect. Myers and Cox pioneered the use of genetic signposts called radiation hybrids to build a rough blueprint of the human genome. This helped geneticists correctly arrange sequence data. Developing the technique led both Cox and Myers to jobs at Stanford.

While developing these maps and also working at the Stanford Human Genome Center, Myers and Cox nurtured another collaboration with Jim Hudson, founder of Research Genetics in Huntsville, Alabama. Hudson made oligonucleotides, clones and cDNAs for Cox and Myers, and Research Genetics distributed these tools to Human Genome Project centres free of charge. The partnership saved time and money for the genome centres, by reducing their reliance on trial and error.

Hudson sold the company in 2000 to Invitrogen, Cox left for the pharmacogenomics company Perlegen, and Myers settled into his role running Stanford's genomics centre. But soon after, Hudson sought Myers' help in building a non-profit research institution.

Initially, Myers was reluctant. "I've got a pretty good set-up and am pretty happy," he says. But Hudson eventually convinced him. Myers will leave his post as director of the Stanford Human Genome Center next year to launch the Hudson-Alpha Institute for Biotechnology, a non-profit facility based in Huntsville, Alabama. A big selling point was the planned focus on collaborations, both within the institute and with the 12 or so companies that will be housed in the same building. The institute's primary research role is applying whole-genome approaches to understanding and treating human diseases, an aim that necessitates interdisciplinary interactions.

"Fostering the spirit of collaboration is the whole point of how this thing is going to be run," says Hudson. If all goes as planned at the Hudson-Alpha institute, no one will have to go it alone — or worry about tenure. ■

Paul Smaglik

NETWORKS & SUPPORT

Authorship accountability

Christine Beveridge and Suzanne Morris have outlined an approach to determine the sequence of authors on a multi-author paper fairly (see *Nature* **448,** 508; 2007). Their criteria are based on percentage scores for each co-author, taking into account a set of items — such as figures, tables, text and ideas — that make up a manuscript. There is a percentage score for each element of the paper. They aim to develop a culture of understanding and agreement about the relative importance of different activities before the start of the work, and to settle the order of authors at manuscript-writing stage.

But there are situations in which a laboratory head or administrative boss includes his or her name without contributing anything (honorary or complimentary authorship), usually as the last author, thus misleadingly implying a leadership role. Such complimentary authors sometimes, by virtue of their position, garner awards or recognition for the work without doing any. Such an author might not hesitate to pass the buck to the co-authors if some wrong is detected. Also, a short-term visiting worker sometimes merely receives an 'acknowledgement' for genuine work in return for being trained or paid for technical work by the group leader. This also needs to be addressed when determining authorship and sharing credit and accountability.

I would suggest that each author receive a quantitative score according to relative contribution in terms of their percentage input. Credit should be earned on the basis of ideas and planning, actual experimentation, recording of observation, data analysis and manuscript preparation. All should be given equal weight, and the overall cumulative score earned by each author taken into account to decide on authorship order. The percentage of contribution over time on the set of scores needs to be periodically reassessed and calculated, making allowance for short-term visitors. A non-participating laboratory head would be included as the last author only if he or she facilitated and verified claims made in the paper. He or she would also be made accountable in the case of disputes. Such an arrangement would go a long way towards maintaining the integrity of scientific papers (see *Nature* **450,** 1; 2007). ■

Umesh Chandra Lavania is a plant geneticist and senior scientist at the Central Institute of Medicinal and Aromatic Plants, Lucknow, India.

POSTDOC JOURNAL

Double trouble

I knew that two kids would a challenge, but when both became ill at the same time, I was blindsided. It's a call every working parent dreads: "It's the school nurse. Your child is sick, please pick him up." Amazingly, this was topped by a second call about the health of my other child.

I effectively missed a week of work dealing with my toddler's uncontrollable, probably virus-induced, fever. At the same time, I was shuttling my baby off to the doctor's again and again, only to find out she had yet another ear infection and was allergic to another antibiotic. In my agitation to get something accomplished at work, I even brought her in to sit with me while I packed seeds for my winter plant nursery. This was, not surprisingly, a short-lived endeavour. At home, I was in a horrible emotional state, wishing I could help my kids more. Unable to get any work done at home or at the lab, I was sleeping even less than usual and felt that I was failing in all areas of my life. All I could do was keep moving and wait for it to pass.

Of course it did pass and the kids are now healthy. The silver lining was that their illnesses didn't cost me more time off, as they happened simultaneously. I know these are just growing pains, and I still love the challenges of being a parent. I hope that next time, I'll be more prepared and won't miss a step. ■

Moira Sheehan is a postdoc in the Department of Plant Breeding and Genetics at Cornell University.

©2007 Nature Publishing Group

NATURE|Vol 450|29 November 2007

Repeating the past

A lesson in family history.

Peter Watts

What you did to your uncle's grave was unforgivable.

Your mother blamed herself, as always. You didn't know what you were doing, she said. I could accept that when you traded the shofar I gave you for that eMotiv headset, perhaps, or even when you befriended those young toughs with the shaved heads and the filthy mouths. I would *never* have forgiven the swastika on your game pod but you are my daughter's son, not mine. Maybe it *was* only adolescent rebellion. How could you know, after all? How could any child really *know*, here in 2017? Genocide is far too monstrous a thing for history books and grainy old photographs to convey. You were not there; you could never understand.

We told ourselves you were a good boy at heart, that it was ancient history to you, abstract and unreal. Both of us doctors, familiar with the sad stereotype of the self-loathing Jew, we talked ourselves into treating you like some kind of *victim*. And then the police brought you back from the cemetery and you looked at us with those dull, indifferent eyes, and I stopped making excuses. It wasn't just your uncle's grave. You were spitting on six million others, and you *knew*, and it meant nothing.

Your mother cried for hours. Hadn't she shown you the old albums, the online archives, the family tree with so many branches hacked off mid-century? Hadn't we both tried to tell you the stories? I tried to comfort her. An impossible task, I said, explaining *Never Again* to someone whose only knowledge of murder is the score he racks up playing Zombie Hunter all day…

And that was when I knew what to do.

I waited. A week, two, long enough to let you think I'd excused and forgiven as I always have. But I knew your weak spot. Nothing happens fast enough for you. These miraculous toys of yours — electrodes that read the emotions, take orders directly from the subconscious — they bore you now. You've seen the ads for Improved Reality™: sensation planted directly into the brain! Throw away the goggles and earphones and the gloves, throw away the keys! *Feel* the breezes of fantasy worlds against your skin, *smell* the smoke of battle, *taste* the blood of your toy monsters, so easily killed! Immerse all your senses in the slaughter!

You were tired of playing with cartoons, and the new model wouldn't be out for so very long. You jumped at my third option. *You know, your mother's working on something like that. It's medical, of course, but it works the same way. She might even have some sensory samplers loaded for testing purposes.*

Maybe, if you promise not to tell, we could sneak you in …

JACEY

Retired, yes, but I never gave up my privileges. Almost two decades since I closed my practice but I still spend time in your mother's lab, lend a hand now and then. I still marvel at her passion to know how the mind works, how it keeps *breaking*. She got that from me. *I* got it from Treblinka, when I was only half your age. I, too, grew up driven to fix broken souls — but the psychiatrist's tools were such blunt things back then. Scalpels to open flesh, words and drugs to open minds. Our techniques had all the precision of a drunkard stomping on the floor, trying to move glasses on the bar with the vibrations of his boot.

These machines your mother has, though! Transcranial superconductors, deep-focus microwave emitters, Szpindel resonators! Specific pathways targeted, rewritten, erased completely! Their very names sound like incantations!

I cannot use them as she can. I know only the basics. I can't implant sights or sounds, can't create actual memories. Not declarative ones, anyway.

But *procedural* memory? That I can do. The right frontal lobe, the hippocampus, basic fear and anxiety responses. The reptile is easily awakened. And you didn't need the details. No need to remember my baby sister face-down like a pile of sticks in the mud. No need for the colour of the sky that day, as I stood frozen and fearful of some *real* monster's notice should I go to her. You didn't need the actual lesson.

The moral would do.

Afterwards you sat up, confused, then disappointed, then resentful. "That was *nothing*! It didn't even *work*!" I needed no machines to see into your head then. *Senile old fart, doesn't know half as much as he thinks.* And as one day went by, and another, I began to fear you were right.

But then came the retching sounds from behind the bathroom door. All those hours hidden away in your room, your game pod abandoned in the living room. And then your mother came to me, eyes brimming with worry: never seen you like this, she said. Jumping at shadows. Not sleeping at night. This morning she found you throwing clothes into your backpack — *they're coming, they're coming, we gotta run* — and when she asked who *they* were, you couldn't tell her.

So here we are. You huddle in the corner, your eyes black begging holes that can't stop moving, that see horrors in every shadow. Your fists bleed, nails gouging the palms. I remember, when I was your age. I cut myself to feel alive. Sometimes I still do. It never really stops.

Some day, your mother says, her machines will exorcise my demons. Doesn't she understand what a terrible mistake that would be? Doesn't history, once forgotten, repeat? Didn't even the worst president in history admit that memories belong to *everyone*?

I say nothing to you. We know each other now, so much deeper than words.

I have made you wise, grandson. I have shown you the world.

Now I will help you to live with it. ■

Peter Watts is a reformed marine biologist and failed gel-jock who is nevertheless adept at faking science, just so long as he can mix some characters and plot in among the numbers. His latest novel was nominated for several prestigious awards, winning none of them.

©2007 Nature Publishing Group