opinion

Decision time at Yucca Mountain
The United States should only proceed with its sole proposed nuclear-waste repository once geological concerns have been addressed and provided that law-makers are convinced that the assessment process has been conducted fairly.

Women worse off (cont.)
The California Institute of Technology needs to address imbalances and frustrations.

news

Japan's plans for space merger spark fears for basic research

Stem-cell list offers sixty-four lines

Elusive fossil could conceal answer to dinosaur debate

Staff survey shows women feel out in the cold at Caltech

LED pioneer seeks reward in court

Joint venture on biochips ends in disarray

It's a dog's life for Siberian foxes

Funding bonanza for astronomy and biotech in Australia
Canada unveils plans to build nanotechnology centre... 846
...as mathematicians beat retreat to Alberta 846
Green-fingered gang could hold climate key 847
Earliest malaria DNA found in Roman baby graveyard 847

news in brief 848

news feature
Out of sight, out of mind? 850
Yucca Mountain in Nevada seems like a reasonable place to bury the United State’s nuclear waste — but only if the volcanoes there remain dormant. Arguments on that point are about to erupt, reports Colin Macilwain.

Brought down to Earth 853
With its unique access to Mir cosmonauts, Moscow's Institute for Biomedical Problems was a world leader for space biology. But now it is working under greatly diminished circumstances, says Quirin Schiermeier.

correspondence
Collaboration with Japan could be more tempting 855
A pilot scheme for postdocs, blending mobility with security, ought to be expanded.

Could a website teach communication skills? 855
Curious effects created by reversal of colour 855

book reviews
The electrifying Australian
DAVID PENNY reviews Platypus: The Extraordinary Story of How a Curious Creature Baffled the World by Ann Moyal
How an egg-laying mammal challenged the world's finest zoologists.

A world without competition
PETER A. ABRAMS reviews The Unified Neutral Theory of Biodiversity and Biogeography by Stephen P. Hubbell

Fertility from a metabolic viewpoint
ROGER V. SHORT reviews On Fertile Ground: A Natural History of Human Reproduction by Peter T. Ellison

Science in culture
REX DALTON

words
Victor and victim
HOWARD P. SEGAL
The true message of Frankenstein is about morality, not mad science.

concepts
Reasoning for results
DENNIS BRAY

news and views
Achilles' heel of cancer?
BERT VOGELSTEIN AND KENNETH W. KINZLER
The p53 protein is inactivated in most human cancers. One outcome is a defect in controlling cell division. Might a virus that exploits this defect prove useful in treating cancer?

Quantum optics: Photons yield to peer pressure
PAUL KWIAT
In 1960 the invention of the laser allowed classical light to be amplified inside a cavity. New experiments show that photons in a special quantum state can also be amplified.
Cardiovascular biology: Creating unique blood vessels

PETER CARMELEIT

When tissues need more oxygen, they release molecules that encourage blood vessels to grow. The discovery of the first such molecule that is specific to one type of tissue has implications for cancer and heart failure.

Quantum engineering: Protecting the quantum world

JUAN PABLO PAZ

When quantum systems interact with the environment, classical properties emerge — a process known as decoherence. Although decoherence is unavoidable, it may still be possible to manipulate the outcome.

Immunology: Antibody alterations

ALBERTO MARTIN AND MATTHEW D. SCHARFF

In an immune response, antibody molecules are altered so they can bind to intruders more strongly. Some of the molecules that determine which alteration process predominates have now been tracked down.

Chemistry: Mystery of an interstellar ion

ANNICK SUZOR-WEINER AND IOAN F. SCHNEIDER

Hydrogen is the most abundant element in the Universe. Its chemistry governs most reactions in space, especially when it is ionized. But the abundance of the H$_3^+$ ion is not easy to explain.

100 and 50 years ago

Daedalus: A cosmic background

DAVID JONES

A weak cosmic background of neutrinos, similar to the cosmic microwave background, could exist, although its faintness will require sensitive telescopes to detect it.

brief communications

Mother's voice recognition by seal pups

ISABELLE CHARRIER, NICOLAS MATHEVON & PIERRE JOUVENTIN

Newborns need to learn their mother's call before she can take off on a fishing trip.
Palaeontology: The beaks of ostrich dinosaurs 873
MARK A. NORELL, PETER J. MAKOVICEY & PHILIP J. CURRIE

Nanostructures: Self-assembled domain patterns 875
RICHARD Plass, JULIE A. LAST, N. C. BARTELT & G. L. KELLOGG

Malaria: Cooperative silencing elements in var genes 875
KIRK W. DEITSCH, MICHAEL S. CALDERWOOD & THOMAS E. WELLEMS

correction: An unusual social display in gorillas 876

articles
Identification of an angiogenic mitogen selective for endocrine gland endothelium 877
JENNIFER LECOUTER, JOE KOWALSKI, JESSICA FOSTER, PHIL HASS, ZEMIN ZHANG, LISA DILLARD-TELM, GRETCHEN FRANTZ, LINDA RANGELL, LEO DEGUZMAN, GILBERT-ANDRE KELLER, FRANKLIN PEALE, AUSTIN GURNEY, KENNETH J. HILLAN & NAPOLEONE FERRARA

letters to nature
Characterization of extrasolar terrestrial planets from diurnal photometric variability 885
E. B. FORD, S. SEAGER & E. L. TURNER

Stimulated emission of polarization-entangled photons 887
A. LAMAS-LINARES, J. C. HOWELL & D. BOUWMEESTER

Mechanism for the destruction of H₃⁺ ions by electron impact 891
V. KOKOOLEN, CHRIS H. GREENE & B. D. ESRY

Parasitic computing 894
ALBERT-LÁSZLÓ BARABÁSI, VINCENT W. FREEH, HAWOONG JEONG & JAY B. BROCKMAN
Universal behaviour in compressive failure of brittle materials
CARL E. RENSHAW AND ERLAND M. SCHULSON

Transition of Mount Etna lavas from a mantle-plume to an island-arc magmatic source
PIERRE SCHIANO, ROBERTO CLOCCHIATTI, LUISA OTTOLINI & TIZIANA BUSA

Genetic linkage of ecological specialization and reproductive isolation in pea aphids
DAVID J. HAWTHORNE AND SARA VIA

Dynamics of travelling waves in visual perception
HUGH R. WILSON, RANDOLPH BLAKE & SANG-HUN LEE

Two-step process for photoreceptor formation in Drosophila
BERTRAND MOLLEREAU, MARIA DOMINGUEZ, REBECCA WEBEL, NANSI JO COLLEY, BENISON KEUNG, JOSE F. DE CELIS & CLAUDE DESPLAN

Virus-mediated killing of cells that lack p53 activity
KENNETH RAJ, PHYLLIS OGSTON & PETER BEARD

Rotational movement during cyclic nucleotide-gated channel opening
J. P. JOHNSON JR AND WILLIAM N. ZAGOTTA

Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation
JULIAN E. SALE, DANIELLA M. CALANDRINI, MINORU TAKATA, SHUNICHI TAKEDA & MICHAEL S. NEUBERGER

erratum: Nodulation of legumes by members of the β subclass of Proteobacteria
LIONEL MOULIN, ANTONIO MUNIVE, BERNARD DREYFUS & CATHERINE BOIVIN-MASSON

new on the market
Going separate ways
The latest developments in chromatography.
The United States should only proceed with its sole proposed nuclear-waste repository once geological concerns have been addressed and provided that law-makers are convinced that the assessment process has been conducted fairly.

This time it is for real: after years of investigation and prevarication, the US government will soon decide whether it can safely bury spent nuclear fuel and other nuclear waste at a planned repository under Yucca Mountain in Nevada. The decision is the culmination of a vast 20-year exercise to characterize the proposed site and assess its associated risks (see News Feature, page 850).

The nuclear power industry in the United States currently stores most of its spent fuel on the surface, at or near nuclear power stations. The US Department of Energy (DOE) is legally obliged to take the spent fuel and so is under intense pressure to proceed with the Yucca Mountain project — currently the United States' only option for long-term disposal.

The decision on whether to proceed with Yucca Mountain has been a long time coming. Sceptics doubt if 20 years of study at a cost of some $7 billion has really been necessary to establish the project's viability and safety. But if the outcome of the process is some measure of public acceptance for the project — not least in Nevada — then the wait will have been worthwhile. Indeed, an argument can be made that the government has not yet waited long enough. The Yucca Mountain project began its tortured existence in an era when public demands for accountability in science-based decision-making were far less onerous than they are today. In those far-off days of the 1970s and early 1980s, the DOE was a cold-war autocracy whose apparatchiks saw little need for consultation either with the general public or with technical advisers from outside its own ranks.

But times have changed since then, and an excellent and timely report (Disposition of High-Level Waste and Spent Nuclear Fuel: The Continuing Societal and Technical Challenges — see http://www.nap.edu) released earlier this year by the US National Research Council (NRC) describes what the changes mean for nuclear-waste disposal. The report notes that no nation has yet succeeded in building a permanent store for spent fuel or other highly radioactive waste. The approach recommended by the NRC's report rests on two pillars: complete openness with the public, and external scientific peer review. The Yucca Mountain process has been strong in the former, at least lately, but weak in the latter. It has been far too reliant on the expertise of the DOE's own laboratories, and not insisted enough on publishing its findings in the scientific literature or on reaching out to experts in the universities and elsewhere. For example, recent work by outside experts on lava flow suggests possible shortcomings in the models that the DOE has used to predict the integrity of the repository in the event of volcanic activity at the mountain.

In the past few years, the project has been gradually evolving in these directions. It has appointed an international panel to review its work, for example — although the panel will not report until after the United States has passed its critical decision point. President George W. Bush is expected to announce his own decision on the matter by the end of the year. It is likely that, armed with a recommendation from the DOE that the site is suitable, he will say the project should proceed. The state of Nevada is then entitled to declare its opposition to the project, and is certain to do so. Finally, under a special law, it will fall to both houses of Congress to vote the project up or down, probably in early 2002.

There is a natural tendency for the technically minded to root for projects such as Yucca Mountain to succeed. Many of those who have been involved in nuclear science and engineering feel that their work is somehow tarnished if a technical problem, such as the safe disposal of spent nuclear fuel, cannot be fixed.

But the paramount considerations must be that the solution is safe and, just as importantly, that the process by which it is derived has been fair. The project should proceed only if the DOE can establish both points.

Women worse off (cont.)

The California Institute of Technology needs to address imbalances and frustrations.

With a growing number of women students, and situated in the state that sees itself as America's most progressive, the California Institute of Technology would seem to be an ideal arena to nurture a research community free of gender warfare. But a study reported in this issue (see page 844) shows deep dissatisfaction among the one in ten of Caltech's faculty who are women. The women feel — and the facts suggest — that they have had an anomalously high chance of being spurned when leadership positions were filled; they feel discriminated against by administrators who dole out endowed chairs, which underwrite salaries; and they have now learned that they are paid less for equal work.

The intense atmosphere has led to backbiting. A senior woman faculty's reticence to fight for equal salaries was called "traitorous" by another woman professor. Women faculty who push for their rights were called "young hotheads" by senior faculty. And there is lingering bitterness among the senior and junior faculty women over battle tactics in the gender war. All this in a state where women have increasingly assumed leadership positions in the corporate and political world.

This past spring, Caltech elected the first woman 'chair' of the faculty. Last winter, the biology division saw its first woman faculty member receive an endowed chair, after some 70% of male biology faculty already had endowments. But the woman biologist's endowment came only after strong lobbying.

As a private university, the onus for improvement rests with Caltech's administration and board of trustees. Its president, David Baltimore, is lauded by a number of women faculty for his leadership and open-door policy of listening to all faculty. Ensuring that women have an equal voice in leadership and committees, bestowing endowments in a way that is demonstrably fair, and giving women equitable salaries would be the major steps forward.
Japan's plans for space merger spark fears for basic research

David Cyranoski, Tokyo

Japan is to merge its three major space laboratories to create a single organization responsible for all space-related research and development.

The merger plan is in keeping with Prime Minister Junichiro Koizumi's vision of streamlining government-funded research (see Nature 412, 364; 2001). But it is already concerning some space scientists, who fear that manned space flight and engineering-based activities will dominate the new organization.

The plan will combine the National Space Development Agency (NASDA) with Japan's main aerospace and space-science laboratories. NASDA is increasingly involved in commercial space applications, such as building and launching communication satellites. The National Aerospace Laboratory of Japan focuses on research areas such as fluid dynamics, whereas the Institute of Space and Astronautical Science (ISAS) conducts space science.

"The merger will bring together divided resources and make it easier for researchers to move freely between projects," says Fujiki Kanji, a senior official at the education ministry, which oversees the three laboratories. Sharing major facilities will make the new body more efficient, adds deputy education minister Takashi Aoyama.

Currently, each laboratory has a different funding basis and administrative standing, which will complicate the merger. Aoyama will chair a committee of laboratory managers and outside experts that will plan the details of the merger before the start of the new fiscal year, next April.

Space commercialization is likely to be a major emphasis for the new organization, Aoyama says. But some researchers worry that basic research could be swallowed up by NASA programmes such as rocket development and participation in the International Space Station, which account for four-fifths of the combined ¥194 billion ($US1.6 billion) budget of the three laboratories.

"The barriers between our basic research and their more applied research will be broken down, and if money is needed for the larger projects, they might take it away from ours," says one ISAS researcher.

Others take a more sanguine view, saying that the change could reinvigorate a field that is facing a crisis of confidence after the failed launch of ISAS’ M-V rocket last February, and of two of NASDA’s H-II rockets.

http://www.isas.ac.jp/e/index.html
http://www.nasda.go.jp/index_e.html
http://www.nal.go.jp/Welcome-e.html

Stem-cell list offers sixty-four lines

Jonathan Knight, San Francisco

The US government has published a list of the 64 human embryonic stem-cell lines that can be used by publicly funded researchers under the Bush administration's new rules.

Publication of the list, which the National Institutes of Health (NIH) posted on its website on Monday, ended weeks of head-scratching among researchers over the large number of stem-cell lines that President George W. Bush claimed existed in his 9 August speech. The list names 10 companies and research labs worldwide that have derived cell lines.

But questions remain about the nature and quality of many of the cell lines, and not all the labs are ready to give the cells out.

The highest concentration of stem-cell lines is in Sweden — 19 at Göteborg University and 5 at the Karolinska Institute in Stockholm. Another 10 cell lines are at two centres in India, with 4 more in Israel.

But CyThera, a San Diego-based company developing stem-cell therapies for degenerative diseases, says it does not know when it will be ready to distribute its nine lines. "We first need to characterize them fully," says its chief operating officer Lutz Giebel.

Audrey Chapman, a science-policy official at the American Association for the Advancement of Science, says the list only begins to address the concerns outlined in an association statement produced on 17 August (see Nature 412, 753–754; 2001).

"The question really is how much more the NIH knows but is not ready to share, or whether they have shared all they know," she says.

www.nih.gov
**Elusive fossil could conceal answer to dinosaur debate**

**Rex Dalton, San Diego**

A Chinese dinosaur fossil of immense potential interest to palaeontologists is circulating between private dealers in Europe, highlighting scientists’ concern that the underground trade in such artefacts is inhibiting their study.

The fossil of a psittacosaurid is special because it has outer coverings, known as integuments, protruding along the tail, something not previously observed in the dinosaur group to which it belongs. If it were to be established that these protrusions carried hair or feathers, that could mean that most dinosaurs had these features.

Mark Norell, an expert on Asian dinosaurs at the American Museum of Natural History in New York, says the story is a telling example of how private dealings can hide important scientific discoveries. “This specimen could redefine how we look at dinosaurs,” he says. “But we can’t look at it with Chinese palaeontologists.

A museum researcher says that the fossil was examined there in some detail, but he has not been allowed to discuss it. The fossil was being kept in a shop called Stone Age in Trieste, Italy.

David Cyranoski, Tokyo

A prominent Japanese materials scientist is suing his former employer for $2 billion (US$16.5 million) over the level of compensation he received for inventing the technology behind blue light-emitting diodes (LEDs).

Shuji Nakamura, who shared the 2014 Nobel Prize in physics for inventing blue lasers, is claiming ownership, in part, of the patent, and wants the technology to be licensed to outside manufacturers. “Nichia will not license it to others,” he says. Nakamura is basing his case on Japan’s ambiguous patent law, under which researchers should get “fair compensation” for their patented research. Researchers do not usually have contracts with detailed reward provisions. “We need this kind of reward, otherwise inventors are just left out in the cold,” Nakamura says.

**Staff survey shows women feel out in the cold at Caltech**

**Rex Dalton, San Diego**

Women staff at the California Institute of Technology are highly dissatisfied with academic conditions there, according to an internal study that has also found they are paid less than men at the prestigious university.

The investigation finds that women are unhappy with their career advancement, compared with 27% of men. Almost two-thirds of the women were dissatisfied or had reservations about their positions, their lack of involvement in faculty leadership, and the method by which they are promoted and receive tenure.

“The study shows a lot of dissatisfaction among women,” says astronomer Annelia Sargent, chair of the faculty committee that conducted the two-year study. The committee is now preparing recommendations for Caltech administrators about the issues it raises.

Among the 283 academic staff at Caltech during the period studied, only 27 were women. But over one-third of undergraduates and one-quarter of graduate students there are women.

The study found that 54% of women were dissatisfied or had reservations about their career advancement, compared with 27% of men. Almost two-thirds of the women were dissatisfied or had reservations about faculty committee participation, while less than a quarter of men had reservations and none were dissatisfied.

David Baltimore, the institute’s president, says “Caltech is trying very hard to bring women into parity with men in the academic world. We are very conscious and sensitized to the issues, and will try to address whatever imbalances there are.”

Sargent’s committee is likely to recommend changes in how women are promoted and given tenure, and in the openness of academic affairs.”When a process is not transparent, there is always room for a perception of injustice,” says Sargent.

The study also found that women were paid less than men, although no figures are disclosed in its report. “Clearly, if there are inequities in salaries we will need to take a serious look at that,” says Steve Koonin, provost of the institute.

**LED pioneer seeks reward in court**

David Cyranoski, Tokyo

A prominent Japanese materials scientist is suing his former employer for $2 billion (US$16.5 million) over the level of compensation he received for inventing the technology behind blue light-emitting diodes (LEDs).

The claim — filed by Shuji Nakamura, now at the University of California, Santa Barbara, against Nichia Corporation of Shikoku — has sent shock waves through Japanese industrial research.

LEDs are semiconductor-based devices that convert electrical current into light. Nakamura is claiming ownership, in part, of the patent, filed in 1991, covering a processing method for making the light-producing layer of gallium nitride-based semiconductors. Nichia, where Nakamura worked from 1979 to 1999, used this technology to produce blue and green LEDs as well as blue lasers.

Nakamura is basing his case on Japan’s ambiguous patent law, under which researchers should get “fair compensation” for their patented research. Researchers do not usually have contracts with detailed reward provisions. “We need this kind of case to force employers to make clear reward conditions for inventors,” says Nakamura.

Nakamura is claiming ownership, in part or in full, of the patent, and wants the technology to be licensed to outside manufacturers. “Nichia will not license it to others,” he says.
Joint venture on biochips ends in disarray

Quirin Schiermeier, Moscow

A successful scientific collaboration between Russia and the United States for developing microchips to process genetic information has collapsed in acrimony.

The collaboration, which involved the Engelhardt Institute of Molecular Biology (EIMB) in Moscow, the Argonne National Laboratory in Illinois and two US-based commercial partners, had provided substantial funding for the Moscow laboratory in exchange for access to its expertise in DNA chip technology.

But in recent months, financial support for the Russian laboratory has dried up as EIMB and Argonne argue over the contractual arrangements between the parties. As a result of the dispute, Andrei Mirzabekov has resigned as director of Argonne’s Biochip Technology Center. In turn, Argonne has suspended the transfer to the Russian laboratory of some of the royalties it receives from the collaboration’s industrial partners.

The collaboration, which began in 1995, has generated almost 40 scientific publications and 30 patents and patent applications. It also triggered a partnership with Packard Motorola. The companies have used research from the collaboration to develop commercial microchips based on tiny gels applied on a glass surface which can perform thousands of biological reactions within seconds, and associated instruments to process and analyse results from the chip. Researchers use the technology to detect genetic mutations and for drug design.

Motorola and Packard have contributed almost US$20 million to the project over five years, making it one of the most successful biotechnology agreements between the US Department of Energy, which runs Argonne, and industry. About 15% of this money went to the EIMB, which also, Mirzabekov says, received funding through a licensing agreement with Argonne.

Argonne officials agree that the arrangement has been a roaring success. “This has been a true model relationship,” says one. But according to Bill Ragland, a research manager at Argonne, it is now time to reduce the size of the Moscow research operation supported by the collaboration.

Moving on: director Andrei Mirzabekov (inset) is looking for new partners for Moscow’s Engelhardt Institute of Molecular Biology.

Mirzabekov claims that some US officials have turned sour on the arrangement because of security concerns about the possible use of biochip technology in detecting biological warfare agents. Argonne officials deny that this has any connection with their decision.

According to Argonne, royalty payments of about $500,000 for the year 2000 have been stopped because of contractual disagreements over intellectual property rights on microchip technology. The technology was initially described by a team at the Moscow lab in the 1990s. Argonne rejects Mirzabekov’s claim that rights for the technology, which was subsequently developed by the joint venture, should now revert back to the EIMB.

Argonne officials say they are keen to transfer the outstanding money, but first want Mirzabekov to sign a contract assuring them that he will continue to abide by earlier financial agreements. Mirzabekov is unhappy with the contract, saying he already has one, and that Argonne is effectively seeking to renegotiate it. “It makes us very worried about the future, thinking that Argonne could depart from other contracts as well,” he says.

Mirzabekov says he has hired lawyers in the United States, but cannot afford to fight a lengthy legal battle with the US lab. He hopes instead that the Russian science ministry will settle the case directly with the US Department of Energy.

Mirzabekov is now searching for new partners in the United States or Western Europe. He hopes to use biochips to detect drug-resistant strains of tuberculosis — a growing health problem in Russia.

It’s a dog’s life for Siberian foxes

Bryon MacWilliams, Moscow

A unique project in Siberia that has helped researchers understand how animals become domesticated could soon expire from lack of funds.

Researchers at the Institute of Cytology and Genetics in Novosibirsk have compressed into decades a process that might otherwise require thousands of years. Through intense selective breeding, they have turned the silver fox (Vulpes vulpes) into an amiable, tail-wagging house pet.

The work, started in 1959 by the late geneticist Dmitry Belyaev, has demonstrated the close links between behavioural and developmental genetics.

The silver fox — known as the red fox in the United States — is closely related to the dog, but had never been domesticated. The 45,000 foxes so far involved in the Siberian project were chosen by researchers for their propensity to being tamed.

The selection process resulted in the emergence of traits in physiology, morphology and behaviour that are most noticeable in the changes in coat colour and the presence of floppy ears and curled tails.

But the institute lacks the $10,000 a year it needs to keep the project going. “This is the only kind of experiment of its kind in the world,” says Ludmila Trut, head of the research group at the institute. “It would be a tragedy if it were to wither.”

The research “is really important because evolutionary change in behaviour is very hard to study,” says Deborah Goodwin, deputy director of the Anthrozoology Institute at the University of Southampton in Britain.

The population of breeding foxes has been reduced from 700 to 100 since 1996 and the staff, too, is declining. The project is now overseen by seven researchers of retirement age.

http://www.bionet.nsc.ru/ICIG

Tail of woe: lack of money leaves the project to domesticate foxes facing an uncertain future.
Canada unveils plans to build nanotechnology centre...

David Spurgeon, Montreal

Canada is set to invest Can$120 million (US$78 million) over five years in nanotechnology, creating a new institute based in the western province of Alberta.

On 17 August, Prime Minister Jean Chretien and Ralph Klein, premier of Alberta, unveiled plans for the Institute for Nanotechnology, which will be attached to the University of Alberta in Edmonton. Employing about 200 people, the institute will be funded jointly by the federal and provincial governments.

Federal support for the institute will come from Canada's National Research Council (NRC). It will be the first NRC institute to be governed jointly with the host university. The centre will collaborate closely with the University of Alberta, the NRC says, making joint appointments and sharing research facilities.

Peter Hackett, the NRC's vice-president for research, says he is "overjoyed" by the announcement. "Canada has decided to make a major investment in an emerging field. We did this before as a nation in 1984 in biotechnology — with the Biotechnology Research Institute in Montreal — and that's turned out to be very successful."

Hackett says that the new institute will be "interdisciplinary from day one". The institute will offer "a different culture and different way of looking at engineering, chemistry, physics and biology", he says. "Everyone will share this perspective of the nano-world."

Michael Brett, a researcher in nanostructure and thin-film engineering at the University of Alberta, agrees. The cooperative arrangement between the university and the NRC will create "a much richer institute", he says, because of the mix of scientists and graduate students and the availability of university facilities.

The main areas of investigation at the institute will relate to biomaterials, biomedical devices, computing and quantum computing.

Erica Klarreich

North American mathematicians will soon be able to immerse themselves in their thoughts at a purpose-built hideaway high in the Canadian Rockies.

When it opens in 2003, the Banff International Research Station will allow invitees to "live, eat and breathe mathematics", says Nassif Ghousseoub, director of the Pacific Institute for the Mathematical Sciences (PIMS) in Vancouver.

Inspired by existing retreats for mathematicians in Oberwolfach, Germany, and Luminy, France, the research station will invite up to 40 mathematicians at a time to participate in intense, five-day workshops.

"We've seen from Oberwolfach and Luminy that this kind of setting is a fantastically successful way to advance a mathematics research agenda," says Philippe Tondeur, director of the mathematics division at the US National Science Foundation (NSF).

The station will cost about $1.5 million per year to run, and will be supported in roughly equal parts by the NSF, the Natural Sciences and Engineering Research Council of Canada, the Alberta Science and Research Authority and PIMS. The NSF and PIMS have already announced their support, and the other partners are expected to confirm their participation on 24 September.

As many as 1,600 mathematicians could pass through Banff each year. In addition to the five-day workshops, it will accommodate groups of 10–15 mathematicians to work for up to four weeks on specific research topics.

... as mathematicians beat retreat to Alberta

Funding bonanza for astronomy and biotech in Australia

Peter Pockley, Sydney

In a bid to answer perennial charges that it has short-changed science, the Australian government is to provide A$151 million (US$80 million) over five years for 15 major facility projects.

Astronomers won the largest single grant, gaining A$23.5 million for studies using the international Gemini telescopes in Chile and Hawaii. But biotechnology was the most favoured field overall, with with total funds of A$47 million for four projects, reflecting the government's emphasis on research with likely commercial applications.

Intense competition for the grants, which were announced on 21 August, meant that only about one in six of the 86 applicants was successful.

Nonetheless, Australian astronomers can now double their observing time on the 8-metre Gemini telescopes. They will also develop instruments to help Australia's bid to host the proposed international Square Kilometre Array radiotelescope.

The second-largest single grant, of A$18 million, will help to establish a neuroscience facility based primarily in Melbourne for studying and treating schizophrenia and Alzheimer's and Parkinson's diseases. Stem-cell research got a smaller grant for a new cell-engineering facility based primarily in Melbourne for studying and treating Parkinson's diseases. Stem-cell research got a smaller grant for a new cell-engineering centre at Monash University in Melbourne.

Research at synchrotron light sources will continue with a A$14.8-million grant for Australian beamlines at facilities in Japan and the United States. Other projects to win backing include facilities for proteomics and genomics, a photonics foundry, a livestock genetic database, a marine station for oil and gas exploration in tropical waters, and a centre for wine research.

Bright outlook: Australian astronomers can look forward to more time on Gemini.
Green-fingered gang could hold climate key

Legions of devoted British gardeners may soon be enlisted into climate-change research, as part of an investigation of how weather patterns make their gardens grow.

The UK Climate Impacts Programme (UKCIP), a government-funded organization at Oxford University, is considering asking gardeners to report on the trees, shrubs and flowers in their backyards. The data could help researchers to identify long-term trends in plant development and growth, and then to match them with climate patterns.

The idea reflects rising interest in Britain in phenology — the study of seasonal plant and animal activity. Phenologists use observational data such as the first sightings of migratory birds in spring and first leaf-fall in autumn to identify changes in the biosphere.

For example, the average dates of leaf unfolding and colouring have shown that the growing season in Europe is extending (see Nature 397, 659; 1999).

UKCIP officer Megan Gawith says that most research on climate change focuses on the possible effects on commercial agriculture and horticulture. "There has been no research on the domestic garden or heritage garden sector," she says. Together with organizations such as the National Trust and the Royal Horticultural Society, the UKCIP aims to investigate how climate change affects typical features of British gardens.

Data from domestic gardens may be "very valuable", says Frank Chmielewski, a crop scientist at the Humboldt University of Berlin and coordinator of the International Phenological Gardens, a pan-European network of 62 gardens containing genetically identical trees and shrubs. But he warns that it will take about 10 years to generate data to investigate climate change, and that many observations will be needed, as gardens are rarely typical of the surrounding environment.

"It would be better to advise gardeners to make phenological observations over a larger area," Chmielewski says, "not only in their own garden but also in the gardens of their neighbours." Observers should ignore flowerers, he says, because they are more influenced by local conditions than are trees and shrubs.

Britain has around 27 million self-confessed gardeners and a national obsession with the weather, making it ideal for such a study. Yet Mark Schwartz, a phenomenology researcher at the University of Wisconsin, Milwaukee, cautions that participants will need to be vigilant. "For the data to be useful, the observed events have to be defined quite sharply," he warns.

Gawith says the new scheme will probably be associated with an already established one, such as the UK Phenology Network, which already has 3,000 volunteer observers.

Earliest malaria DNA found in Roman baby graveyard

Malaria experts are helping archaeologists to work out what caused the deaths of 47 babies whose bodies have been unearthed together in a fifth-century cemetery outside Rome.

The experts think the tiny skeletons — including the bones of 22 miscarried fetuses — may have been victims of Plasmodium falciparum, an extremely virulent form of the organism that causes malaria. The Roman site could mark the northern edge of P. falciparum’s penetration into Europe.

The archaeologists who excavated the site, led by David Soren of the University of Arizona in Tuscon, are intrigued by its unusual configuration. The cemetery, at Lugnano, Umbria, was created in a ruined villa. The bones were found at different depths, with more at higher levels.

"This is not the normal burial pattern for a Roman cemetery," says Soren. There were also signs of witchcraft at the burial site, including puppy skeletons, perhaps meant to ward off demons thought to cause disease. Study of the soil between the remains indicates that all the babies were buried within a few weeks.

Medics suggest a malaria epidemic as the cause of the deaths. The honeycomb pattern of many of the bones suggests anaemia, a symptom of malaria, and severe malaria implied in the era’s literature.

"Malaria is the most logical explanation for the deaths," says Mario Coluzzi, a parasitologist at the University of Rome. DNA studies, due to be published next month in Ancient Biomolecules, show the presence of P. falciparum DNA in the bones of the oldest infant — the earliest malaria DNA ever identified.

"This is by no means proof that the cause of the infants’ death was malaria," says Robert Sallares of the University of Manchester Institute of Science and Technology, U.K., who led the DNA study, "but it allows it to remain a contender."
news in brief

Committee calls for end to secrecy in animal research

London The public should have greater access to information about the use of animals in research, a committee advising the British government says.

A report from the Animal Procedures Committee, an independent group that includes practising researchers and animal-welfare representatives, says that a change in the law is needed to address public concerns about secrecy surrounding the licensing and conduct of animal procedures.

The group called for more publication of negative results to prevent repetition and wider dissemination of details of infringements. The committee also says that a clear and concise summary of procedures should be included on licence applications and that these should be made public. The government says that it broadly accepts the recommendations and will act on them.

No charge for Japan’s cloned mice

Tokyo The mouse clones that formed the backbone of the international effort to make sense of the mouse genome (see Nature 409, 685–690; 2001) are to be made available to researchers free of charge.

Dnaform, a spin-off from Japan’s Institute of Physical and Chemical Research (RIKEN), based in Tsukuba, near Tokyo, began supplying the 21,000 clones on request on 27 August. Each clone can be used to study an individual piece of complementary DNA (cDNA) and the gene associated with it.

The announcement was made by Yoshihide Hayashizaki, chief researcher on the mouse cDNA project, at last week’s Mouse Molecular Genetics Meeting in Heidelberg, Germany. Previously, only researchers collaborating with RIKEN had been allowed to use the clones.

Brazil plans to make generic AIDS drugs

London The Brazilian government has announced plans to start producing a generic version of a drug patented by Swiss pharmaceutical company Roche.

The government claims that the country will save US$35 million a year by producing its own version of nevirapine, one of a cocktail of drugs used to treat AIDS. It says that negotiations with Roche over possible delivery of the drug at a reduced price had failed to come to a satisfactory conclusion.

Meanwhile, the South African Treatment Access Campaign (TAC) last week took the country’s health minister, Manto Tshabalala-Msimang, and nine provincial health ministers to court over their AIDS policies. The campaign demands that nevirapine, which is used to prevent mother-to-child transmission of HIV, be made freely available to pregnant women infected with the virus.

Canada chews on transgenic food advice

Montreal Canada needs a centralized information system to advise people about the content of food containing genetically modified ingredients, a report from the Canadian Biotechnology Advisory Committee has concluded.

The committee, an independent body set up to advise the federal government on policy issues, also recommended that a chief safety officer be appointed to oversee the safety of genetically modified food and that federal government make a clear distinction between its roles in promoting and regulating biotechnology.

But the committee’s decision to call for voluntary rather than compulsory labelling has angered food-safety groups, such as the non-profit Council of Canadians. Several such organizations boycotted public consultations relating to the report, claiming that the committee was biased in favour of biotechnology interests.

Solar scheme sails on despite missile crash

Washington A private astronomy group has announced plans to press ahead with a scheme to put a solar sail into orbit, despite the failure of a test launch last month.

The Planetary Society, based in Pasadena, California, says that the first of two solar-sail launches will take place early next year. A suborbital launch on 20 July to test the deployment of two 15-metre triangular sails ended when the sail failed to separate from the converted Soviet intercontinental ballistic missile that was carrying it. The rocket and craft crashed on the Kamchatka Peninsula in eastern Russia.

“We would find it a diversion to build another suborbital test craft,” said the society’s executive director, Louis Friedman, adding that insurance money from the failed test would help to pay for two new rockets and satellites.

French receive taste of crop vandalism

Paris French activists attacked an experimental trial of genetically modified (GM) crops last week, the first significant act of vandalism against GM crops in France.

Around 150 people uprooted insect- and herbicide-resistant plants belonging to US firm Monsanto at a 1,000-square-metre plot in Beaupréau, southern France. The action followed a threat in July by the Confédération Paysanne, the farmers’ union led by José Bové, to destroy all open field trials of GM crops. Bové argues that outdoor trials could allow GM crops to spread beyond the trial sites, but says he does not object to tests in greenhouses.

French research minister Roger-Gérard Schwartzengraben described the destruction as “unacceptable”.

Johns Hopkins faces fresh probe over study safety

Washington Johns Hopkins University in Baltimore, Maryland, is set to come under further scrutiny after the US Office for Human Research Protections announced plans to investigate a study that was overseen by the university’s researchers.

The research into reducing exposure to lead in the home was carried out by the Kennedy Krieger Institute, a paediatric research centre based in Baltimore, but was supervised by scientists from Johns Hopkins. Two lawsuits brought by families involved in the study are currently under consideration in Maryland courts. They centre on allegations that participants in the experiments were not informed of the health risks of the housing in which they were living.

The announcement of the investigation comes a month after the office suspended all federally funded human studies at the university for five days, after the death in June of a healthy volunteer in an asthma study.

Lab life: the British public could soon be better informed about experiments on animals.
Out of sight, out of mind?

Yucca Mountain in Nevada seems like a reasonable place to bury the United State’s nuclear waste — but only if the volcanoes there remain dormant. Arguments on that point are about to erupt, reports Colin Macilwain.

A rock and a hard place: Yucca Mountain, proposed site of the US nuclear waste repository.

At the summit of Yucca Mountain, a gentle summer breeze breaks the scorching heat of the Nevada Desert. “People have two reactions when they come up here,” says William Boyle, a geologist and senior technical adviser to the US Department of Energy’s Yucca Mountain Project, which plans to store nuclear waste at the site. “Some of them say: ‘Well, this is pretty well as near to the end of the world as you can get.’ The others say: ‘Are those volcanoes over there?’”

Across the Solitario Canyon, three rusty volcanic plugs protrude from the sandy-coloured rocks that dominate the surrounding landscape. One of them, about 20 kilometres away at Crater Flat, indicates relatively recent volcanic activity in geological terms: it is only about 75,000 years old.

The volcanoes have not been discussed that much during the 20-year characterization to establish whether Yucca Mountain is a suitable nuclear-waste repository. But they will be — and not merely by environmentalists opposed to the project.

The Nuclear Regulatory Commission (NRC) is the government-funded watchdog that must license the proposed facility, and it thinks that the Department of Energy (DOE), which is responsible for getting the repository built, is seriously underestimating the risk presented by the volcanoes. The argument will come to a head this autumn, just as President George W. Bush is expected to issue his recommendation on whether the project should proceed.

For now, the DOE does not need a firm design proposal for the repository. Rather, it is publishing several documents, including a science and engineering report summarizing all of the studies undertaken during the site-assessment process, that will inform Bush’s decision on whether or not to proceed. If he decides to do so, the project team has another two years to finalize its design and submit a licence application to the NRC.

For the NRC to approve the facility, the DOE must meet the standards published earlier this year by the Environmental Protection Agency (EPA). These demand that for 10,000 years after the repository opens, no members of the public should be exposed to more than 15 millirem (0.15 millisieverts) of radiation per year.

Desert manoeuvres

To prepare for its moment of truth, the DOE has mounted one of the largest scientific risk-assessment exercises in history, at a cost of about $7 billion. Over two decades, its characterization of the repository plan has roped in experts including geologists, microbiologists, seismologists, volcanologists, statisticians, mathematicians, sociologists, politicians, theologists and members of the public. “So far as I know,” says Robert Levich, a geologist and project spokesman, “this has been studied more thoroughly than any other geological site in the world.”

The future of US nuclear power hinges on the process. If there is no repository and high-level radioactive waste has to be stored locally on the surface, as if it is now, it is unlikely that the United States will build any more nuclear power stations.

Given such high stakes, attempts to find a home for the nuclear waste got off to a rocky start. In the early 1980s, a process was under way to select a suitable repository from a number of reluctant contenders, including salt beds in Texas or Louisiana, basalt in Washington state and granite in the northeastern United States. But political interventions excluded the other options, and in 1987 Congress decreed that only the Nevada site would be considered.

Many experts say that Yucca Mountain was emerging as the most technically suitable site in any case. “Yucca Mountain was the best site at the time, because of the lack of water,” says Michael Voegele, a geological engineer who has worked on repository designs for more than 20 years. He is now chief scientific officer for Bechtel SAIC, the Las Vegas-based private contractor that helps to manage the site-characterization project for the DOE.

The DOE’s plan is for Yucca Mountain to accommodate high-level radioactive waste about 300 metres below the desert floor, and yet 300 metres above the water table. It will be housed inside stable, consolidated volcanic ash — or tuff — which has lain moreor
The hole picture: an 8-kilometre tunnel has been bored through Yucca Mountain to allow for tests.

Tubular cells: a range of canisters holding radioactive waste will line the floors of dozens of tunnels bored through the mountain.

Tunnel vision

Concerns about the water and doubts about the wisdom of burying the waste irrevocably have caused the DOE to abandon its original plan for the repository, which would have involved filling tunnels in the rock with nuclear waste and permanently sealing them up. Instead, every disused fuel rod and shaft of glass into which waste has been sealed will be housed in multilayered metal canisters, weighing 20–70 tonnes when full, which will sit on the tunnel’s floor (see diagrams, above).

The effectiveness of the canisters is being probed by members of the Nuclear Waste Technical Review Board (NWTRB), an official watchdog set up by Congress to keep a close eye on the project. Under President Bill Clinton, the board gradually filled up with members who were sceptical of the project’s early self-assurance. “The NWTRB has had a major impact on the project,” says Kevin Crowley, staff director for radioactive waste management at the National Academy of Sciences. “They’ve pushed for fuller consideration of uncertainties.”

At a meeting in Las Vegas in June, for example, NWTRB members questioned the integrity of the DOE’s models of how the canisters will behave. Each of the $1-million cans features a protective layer made of a heat-resistant nickel alloy called alloy 22. Like aluminium, this protects itself from corrosion by accumulating a thin, passive oxide layer on its surface. By extrapolating corrosion rates recorded in laboratory tests over two or three months, DOE scientists believe they know how the canisters will fare over 10,000 years. But Paul Craig, a retired engineering professor formerly at the University of California, Davis, who serves on the NWTRB, is not convinced. “To extrapolate four or five orders of magnitude without some kind of mechanism is just not credible,” he says.

Discussions between the NWTRB and the DOE have also focused on the extent to which heat emitted by the waste will keep the storage tunnels dry. Last year, the NWTRB asked the project to consider the comparative merits of a radically different approach — a low-temperature repository, in which the fuel canisters would not be crammed so tightly in the tunnels and would be air-cooled during the first few decades of storage, reducing the rate at which they would corrode. Uncertainty about the percolation of water through the tuff has also led project planners to consider a $4-billion titanium umbrella, or drip shield, to protect the canisters from direct exposure to dripping water.

DOE officials remain confident that these uncertainties will not throw the Yucca Mountain Project off course. Even taking the most pessimistic views about the ingress of water and the integrity of the canisters, the mean annual dose that could leak from the repository during the 10,000 years still falls way below the upper limit of 15 millirems. Residents of the community whose water supply would be most affected — Lathrop Wells, a truckers’ stopover mainly comprising two filling stations and a brothel — would be exposed to only one-millionth of this dose.

But the risks posed by volcanic activity may be a different matter. So far, these have not drawn much attention from the NWTRB, although it will examine them at a special meeting in Las Vegas on 12 September. Although the prospect of the repository being breached by lava is remote, the consequences could be catastrophic.

The DOE contends, based on studies in the 1980s by Bruce Crowe at the Los Alamos National Laboratory, that Yucca Mountain will be shielded for several million years. Inside the experimental tunnel that the DOE has bored some 8 kilometres into the mountain, the walls are dry and pockmarked with thousands of metal plaques that mark holes from which geological samples have been extracted. But the appearance of dryness is deceptive. It is the tunnel ventilation that keeps the surface of the walls dry; the rock behind contains about 10% water by volume. This slowly seeps down from the sporadic rain — 15 centimetres a year, on average — that lands on the desert above.

…
National Laboratory in New Mexico and others, that the chance of such a breach occurring is just over one in a hundred million per year—a sufficiently small risk that some scientists would be happy to discount it. But the NRC puts the risk at one in ten million per year, which researchers at the NRC say cannot be ignored.

The disagreement stems from the methods used to estimate the risk of an eruption occurring at different areas around Yucca Mountain. The original DOE assessment assumed that all areas had an equal probability of experiencing a volcanic eruption, says John Trapp, an NRC geologist. Trapp and his colleagues at the NRC’s Center for Nuclear Waste Regulatory Analyses (CNWRA) at the Southwest Research Institute in San Antonio, Texas, used a more complex model that takes into account the fault characteristics of the rock in different areas (C. B. Connor et al. J. Geophys. Res. 105, 417–432; 2000).

The NRC researchers also believe that the DOE is severely underestimating the fallout that would result from volcanic activity. At the heart of this issue is the way in which an eruption would damage the repository.

Dyke dynamics

Eruptions are caused by a vertical wall of lava, called a dyke, that rises towards the surface until it breaks through at the weakest point. At present, the DOE models assume that a dyke would cut cleanly through the repository, expelling the contents of a few waste canisters into the atmosphere. The models anticipate that the radioactive fallout from such an eruption would scatter onto the surface and eventually arrive in the water supply. They predict an average dose at La Jolla Wells of up to a millirem per year—far higher than the projected dose from any other cause, but still an order of magnitude less than the limit permitted by the EPA rule.

But according to Trapp and other volcanologists, a dyke rising through the repository would move at only about a metre per second—and so would be liable to flow through many of the tunnels, heating their contents and perhaps releasing much larger quantities of radioactive waste. “They are equating a dyke to an eruption,” says Trapp. “But a dyke goes real slow.”

Trapp’s assessment appears to backed up by two unpublished papers by groups led by Andrew Woods at the University of Cambridge, UK, who studies lava flow. In a paper currently under review at the Journal of Geophysical Research, Woods and Onno Bokhove, a mathematician at Bristol University, model the behaviour of a dyke rising through the repository. They conclude that the shock waves created by the dyke would be powerful enough to break open the end of some repository tunnels and allow most of the contents to be expelled to the surface.

Although the paper finds that most of the canisters would not be breached directly by the shock wave, a second paper by Woods and his collaborators at CNWRA, which has been submitted for publication in Eos, finds that after the initial shock wave has passed, most canisters would fail as a result of being heated— and weakened—by the lava flow.

The combined implication is that far more waste would be expelled into the atmosphere by an eruption than the DOE has allowed for. NRC scientists have not yet published an estimate of how many canisters of waste may be released in this way, but Brittain Hill, a volcanologist at CNWRA and co-author of the paper submitted to Eos, notes that there will be about 150 canisters in each parallel tunnel at the depository, and that a dyke would probably cut into about 10 tunnels. “The DOE models have a dose of 1 millirem from 10 waste canisters,” he says. So for this scenario, the potential fallout could clearly exceed the EPA’s limit.

Grounds for appeal

Little headway was made when NRC and DOE volcanologists met in Las Vegas hotel on 22 June to discuss the implications of the unpublished work. Greg Valentine, a geologist at the Los Alamos National Laboratory who works on the DOE project, says that the NRC’s points “are credible, to the extent that we can’t just dismiss it. It is something that needs to be looked at.” But Trapp and others argue that the DOE scientists have not fully taken the issues on board. “The meeting was a tremendous disappointment,” he says. Hill claims that the DOE is “giving short shrift to what they admit is the only source of risk.”

The DOE is having to get to grips with the issue just as Bush prepares his decision on whether the repository should proceed. Senior Republican politicians support the project, and Bush is widely expected to approve it. The Nevada state government then has the right to veto the plan, which it is certain to exercise. But if such a veto is issued, the final say on whether the project can proceed will rest with the US Congress.

Both houses would need to vote for the project by a simple majority, and recent voting history suggests that they are prepared to do so. But since the Democrats wrested back control of the Senate in May, Senator Harry Reid (Nevada, Democrat) has threatened to use his new position as Senate Majority Whip to try to block the project.

In such circumstances, project advocates can ill afford the arrival of a credible scientific challenge to their assumptions about the site’s safety. And with just such a challenge looming, some observers are surprised that the DOE has not done more to revisit its models of volcanic activity. As the National Academy of Science’s Crowley puts it: “When your regulator says you’ve got a problem—you’ve got a problem.”

Colin Macilwain is Nature’s news editor.

Yucca Mountain Project ➤ http://www.ymp.gov
Nuclear Regulatory Commission High-Level Waste Program ➤ http://www.nrc.gov/NMSS/DWM/hlw.htm

A yucca plant in the Nevada Desert, from which the mountain and the project get their names.
Brought down to Earth

With its unique access to Mir cosmonauts, Moscow’s Institute for Biomedical Problems was a world leader for space biology. But now it is working under greatly diminished circumstances, says Quirin Schiermeier.

When the flaming remnants of the Mir space station plunged into the South Pacific on 23 March it marked not only the end of a successful mission, but of an era of human space flight. This expensive and risky pursuit, no longer the ambition of single nations, is now a global partnership.

For the next decade, human activities in space will be dominated by the International Space Station (ISS). And Anatoli Grigoriev, director of the Institute for Biomedical Problems (IBMP) in Moscow, has mixed feelings about this state of affairs. Without the ISS, his institute would have no future. Yet the IBMP’s involvement in the project will be very different from the scientific autonomy it enjoyed on Mir. “We used to be hosts, but now we are only guests,” Grigoriev laments.

Grigoriev’s career has spanned the rise and fall of the Russian space programme. He joined the IBMP as a medical graduate in 1966, just five years after Yuri Gagarin’s pioneering flight had electrified the nation. Working on the regulation of blood salt levels under microgravity, he rose through the ranks until he was appointed to his present position in 1988.

The IBMP, with a complement of some 4,000 scientists, engineers and technicians, was then at its zenith. Although the cracks were beginning to show, the Soviet Union was still a genuine superpower. The space programme, for which the IBMP provided full medical support, was the nation’s pride and joy, and the first module of Mir had been launched just two years before. “That time was a unique and fulfilling experience,” says Valeri Bogomolov, one of the institute’s deputy directors.

Through experiments on animals sent into orbit, the IBMP had built up its expertise on the biological effects of space flight. But it was the institute’s work on Mir cosmonauts, investigating the physiological effects of long-duration space flight, that dominated its agenda under Grigoriev’s leadership. IBMP scientists studied how the body adapts to microgravity, and developed countermeasures, such as exercise regimes and gravity-simulation suits. Many other aspects of space research — including water conservation, microbiology and radiation biology — were also tackled.

Decline and fall

But this work was set against a background of steady decline. As research budgets dwindled after the break up of the Soviet Union, Grigoriev’s institute had farther to fall than most. Today the IBMP, housed in a series of run-down brick buildings on the outskirts of Moscow, has lost 90% of the staff it had in 1988.

The IBMP’s fortunes were mirrored, more publicly, by those of Mir. The station was originally intended to operate until 1991. But when it became clear that Russia could not afford to launch a successor, the space programme’s efforts became focused on keeping Mir alive. More modules were added, the last in 1996. In the meantime, Mir cosmonauts set endurance records: Valeri Polyakov, now a deputy-director at the IBMP, stayed in orbit for an unmatched 437 days and 18 hours in 1994–95, conducting medical observations on fellow crew members.

But by 1997 the rot had set in. Mir suffered fire, computer failures and a collision with a supply vehicle that depressurized one module. In the Western media, Mir was lampooned as an orbiting rust bucket. Its commander, Vasili Tsibliev, whose health declined as disaster followed disaster, was presented as a pathetic figure, in contrast to the heroic portrayal of the NASA astronaut Michael Foale, who was on Mir at the same time. For staff at the IBMP, it was a dispiriting experience. “It was a very difficult situation,” says Bogomolov.

The station’s demise has deepened the depression at the institute. The IBMP is now preparing for research on the ISS and, in principle, Russia is an equal partner. But in practice, financial constraints have made
the former superpower a space wallflower. Russia’s Zvezda service module, docked onto the ISS in July 2000, could be financed only by selling 4,000 hours of Russia’s total research time aboard the first phase of the ISS — more than half of its total allocation.

Many of the IBMP’s staff fear that this deal, struck in 1998, will be the start of a larger sell-off. Russia, although invited, is not participating in the International Life Science Working Group that is managing the selection of biology and medical research proposals for the ISS. Russia says it cannot afford to join this club, whose members have pooled their experimentation time and facilities. Staying away gives Russia the option of selling more of its allocated research time. But it also means that Russia will be scientifically isolated on the ISS.

Isolation station

That is unfortunate, because the IBMP’s separation from the international scientific community has been a long-standing problem. The institute does have some researchers who are recognized as world leaders in their fields — neurophysiologist Inesa Kozlovskaya, for example, has published more than 250 papers in English-language journals. But she is an exception. Founded as a military institute, much of the IBMP’s research was published only in Russian, and often restricted to reports delivered to senior officials. Tellingly, when Nature handed Grigoriev the latest issue of the journal, it was the first copy he had ever opened.

Since the end of the Cold War, international collaborations have opened up. In 1991, NASA, the US National Space Biomedical Research Institute in Houston and the IBMP agreed to exchange physiological data. Many Western space agencies sent astronauts to Mir, and ground-based collaborations have also been established (see ‘Cosmonauts behaving badly’, right).

Most scientists who have collaborated with the IBMP are positive about the experience. But they warn that good personal contacts are needed to ensure that things run smoothly. “It is essential that the modes of operation of a joint project, its scientific goals, and the methods and statistical tools to be applied are defined in detail,” says Rupert Gerzer, director of the Institute of Aerospace Medicine in Cologne, part of DLR, the German aerospace research agency. “Otherwise you can experience unpleasant surprises.”

Ary Goldberger of Harvard Medical School has carried out research on heartbeat dynamics using IBMP data. “The Russian data were invaluable,” he says. “I immortalizing their unique data in an open archive should be a major priority.” The Russian data are not always well collated, says Goldberger. But he adds that he has also experienced problems in obtaining data from NASA.

Grigoriev works hard to dispel reservations about working with the IBMP, and points to the institute’s unique experience in providing medical support for long-duration space missions. “Much is hard for us to do these days,” he says. “But without Mir there would be no ISS.”

“Russia’s long-term background in the area is very valuable,” says Victor Schneider, of NASA’s life-sciences division. “But the Russian programme, limited in scale, has only whetted our appetite.”

Grigoriev agrees that additional research is necessary, in particular on bones, red blood cells, metabolism, and the cardiovascular and immunosystems. How much of a contribution the IBMP’s scientists make to this work may depend on the health of Russia’s fragile economy. But Grigoriev believes that western scientists and space agencies have more to gain from working with the IBMP than just access to its data from Mir. “There are differences in scientific traditions, customs and ambitions,” he says. “We have not yet learned the best from each other.”

Until now, cultural differences in the approach to human space-flight have tended to cause tensions, with NASA’s buttoned-down approach clashing with the Russian penchant for improvisation. But as long as the IBMP remains part of the ISS, Russia’s cosmonauts will have the chance to show that they still have something to contribute.

© 2001 Macmillan Magazines Ltd

http://www.ibmp.rssi.ru/webpages/engl/welcom.html
Collaboration with Japan could be more tempting

Sir — Your News story “Japan aims to forge stronger European links” (Nature 411, 875; 2001) clearly identifies why it is difficult to recruit foreign scientists to work in Japan. The Japan Society for the Promotion of Science (JSPS) has been running a pilot ‘2+2 programme’, which allows a British researcher to spend two years in Japan with the assurance of two years’ subsequent funding in the United Kingdom. To promote greater collaboration between research in the European Union (EU) and Japan, this should now be established in other European countries. I would not have applied for this scheme without the return element, yet have already felt the benefits.

Researchers tempted by a period of research in Japan, or wishing to establish a long-term collaboration, face other problems in addition to those mentioned in your News story. Permanent university staff have heavy teaching loads which, combined with family commitments, allow only the fortunate to take long-term sabbaticals. Postdoctoral researchers are far more mobile (often not from choice). Schemes such as the 2+2 programme can turn this mobility into a positive factor, helping them establish long-term collaborations early in their careers, while providing a degree of security.

Unfortunately, much may depend upon the Japanese economy. Provision under the JSPS scheme for Japanese language lessons has recently been reduced, and the tax-free salary status is under threat. Several other improvements could make a significant difference to the quality and number of applicants, for example one paid-for trip home during the two-year period, and greater opportunity for overseas research collaboration. Changes are also required in the host countries — I am unable to contribute to my university pension while in Japan, for example.

Other schemes for EU-Japanese collaboration should be open to all, unlike Britain’s otherwise excellent Biotechnology and Biological Sciences Research Council (BBSRC) Japan Partnering Awards (http://www.bbsrc.ac.uk/international/bbsrc/japan.html), which restricts applicants to current BBSRC grant holders or employees.

Although you report that in the past three years only two people have applied for JSPS funding from the EU, the JSPS does have other types of grant allotted for specific countries or institutions within Europe, and it expects to fund 440 fellowships for the financial year 2001 (see http://www.jsps.go.jp/e-home.htm).

Angus Davison
Department of Ecology and Evolutionary Biology, Biological Institute, Tohoku University, Aramaki-Aza-Aoba, Aoba-ku, Sendai 980 8578, Japan

Could a website teach communication skills?

Sir — The Opinion article “Learning to speak and write” (Nature 411, 1; 2001) points out that scientists need to be taught how to communicate the results of their research effectively. In Correspondence, D. A. Watson and his colleagues and M. Attrill (Nature 411, 992; 2001) refer to courses where this is being done.

Young scientists who do not have English as their first language are keenly aware of a further obstacle: the need to acquire these skills in English. Without them, many are media-shy and feel handicapped in effectively communicating the full significance and scope of their work, even if they have good science-communication skills in their native languages.

The Internet is changing the pace and practice of disseminating scientific information (Nature 410, 1023–1025; 2001), but the basic skills of scientific writing are still indispensable. To induce more young scientists to improve their scientific writing skills, the success story of the team running the web-based Journal of Young Investigators as discussed in your News Feature (Nature 411, 13–14; 2001) must be repeated by others, launching similar affordable Internet journals in other languages. This will go a long way in building the younger generation’s confidence and skills.

In the absence of any formal training in scientific writing, it takes weeks or more for scientists to prepare their first few manuscripts in English. Publication will often be further delayed when journal editors ask for manuscripts to be rewritten. In an era when speed of communication is vital, the global scientific community needs a freely accessible website that will help people learn science-communication skills through home study, and assist them in evaluating the level of proficiency they are achieving.

Virendra K. Bhasin
A. R. L. Teixeira’s Chagas Disease Multidisciplinary Research Laboratory, Faculty of Medicine, University of Brasilia, Brasília-DF, Brazil

Curious effects created by reversal of colour

Sir — The inadvertent reversal of Masaccio’s Trinity (see above left for correct orientation) in my essay “Maintaining M. Masaccio” (Nature 412, 382; 2001) made nonsense of my point about the relationship between the asymmetrical composition and the off-centre approach by the spectator, but the error may yet be turned to good account.

As Leonardo da Vinci recognized, reversing a composition can make it look very different and even ‘wrong’: “When you are painting, you ought to have by you a flat mirror in which you should often look at your work … [I]t will seem to be by the hand of another master and thereby you will better judge its faults.”

This property of a reversed image has been widely recognized, with respect to the arrangement of forms, but it occurs to me, looking at the flipped Trinity, that colour might also be involved. Masaccio’s carefully orchestrated asymmetries of red and blue look distinctly uncomfortable in reverse. Is this just a question of familiarity, or is there a ‘handedness’ in colour distribution as well as in composition?

Martin Kemp
Department of the History of Art, University of Oxford, 59 George Street, Oxford OX1 2BE, UK
The electrifying Australian

How an egg-laying mammal challenged the world’s finest zoologists.

Platypus: The Extraordinary Story of How a Curious Creature Baffled the World
by Ann Moyal

David Penny

If any animal has mystery and charisma, it is the platypus. This book tells its story, and it’s a winner. In Platypus, Ann Moyal uses this extraordinary animal to record the continuing 200-year puzzle of understanding the Australian fauna. Her book has something for everyone — the excitement of a detective story, the history of biological ideas, the frustration at the fact that morphologists appear not to want to understand the evolution of structure in relation to function.

The arrival in England in 1799 of a platypus skin, preserved in rum, highlighted the distinctive nature of the new (to Europeans) Australian fauna. For the next 100 years, every leading zoologist debated the nature of Australian fauna. For the next 100 years, every leading zoologist debated the nature of Australia. Analyses of the platypus by Jean-Baptiste Lamarck, Georges Cuvier, Etienne Geoffroy St-Hilaire, Richard Owen and Thomas Huxley all had their opinions, occasionally uncharacteristically humble. It was 25 years before Johann Meckel discovered the presence of mammary glands. “If these glands produce milk, let’s see the butter,” retorted one disbeliever.

But at least the platypus was now recognized as a mammal. And therefore, went the reasoning, it must give birth to live young. There was no ‘sense’ in laying eggs if the young were to suckle — even if the aborigines said that the platypus laid eggs. Thousands of platypuses were shot to settle the question. This is the dark side of zoological history, the ‘search and destroy’ attitude that turned zoological research into the killing fields. One Scottish naturalist, William Caldwell, returned home from Australia with the remains of more than 1,300 echidna (the other egg-laying mammal). While I was an undergraduate, it was said, “the difference between a botanist and a zoologist is that botanists actually like animals”. To the classical zoologist, the only good animal was a “dead and dissected” one. But finding that the platypus laid eggs didn’t settle the still unresolved mystery of what selective forces led an egg-laying mammal to develop mammary glands and for its young to start suckling.

Finding out the function of the platypus’s bill took even longer. Everard Home, from the Royal College of Surgeons in London, had suggested by 1802 that it was an exploratory organ that replaced sight and sound in the underwater world where the platypus foraged. Yet it was nearly 200 years later that a German-Australian research group reported that the bill detected electric fields; a freshwater shrimp could be detected ten centimetres away. The platypus electric had been found, and European scientists had at last discovered something that the aborigines did not already know.

As Moyal’s book shows, ‘big science’ is not just a phenomenon of the past 60 years — it was already flourishing at the end of the eighteenth century. Some of the earliest discoveries of the unusual nature of the Australian fauna date from Captain Cook’s first voyage to the South Seas in 1768–71. The scientific purpose of this expedition was to observe the transit of Venus in Tahiti. The French also mounted several major expeditions to the Southern Hemisphere, one with two vessels and 23 scientists. Even if England and France were at war, in the Antipodes the scientific expeditions could meet and toast the internationalism of science.

My only quibble with Moyal’s version of the scientific thought of this period is that she accepts William Whewell’s (1830s) simplistic division of the geological theories of the time into ‘catastrophist’ and ‘uniformitarian’ — uniformitarian referring to the idea that change occurs at a uniform rate throughout time. Charles Lyell objected (in his third volume of Principles of Geology, 1830–33) to Whewell’s description of his theory as uniformitarian. Much earlier, James Hutton had argued against uniform rates of geological change and, even later, Charles Darwin frequently insisted, in respect of biological evolution, that rates could vary — it was the basic mechanisms of change that were the same. Hutton, Lyell and Darwin denied both catastrophism and uniformitarianism. They each combined the concepts of a long timescale and the continuity of testable mechanisms through time (although with variable rates) to give the viewpoint known as ‘actualism’. In one of those twists of fate, the loser (Whewell, a strong opponent of both Lyell and Darwin) laid the framework for writing this history. But it really is the time to give up the idea that they were uniformitarians; orthogenesis fits much better into that category.

But such quibbles aside, everybody will enjoy this book, and the story continues to unfold. We still do not know which features of the platypus are truly ancient and which are more recent specializations that allowed an early mammal to become both an aquatic and a burrowing animal. Read it and enjoy.

David Penny is at the Institute of Molecular Biosciences, Massey University, Box 11-222, Science Towers D5-01, Palmerston North, New Zealand.
A world without competition

The Unified Neutral Theory of Biodiversity and Biogeography
by Stephen P. Hubbell
Princeton University Press 2001. 448 pp. $29.95, £19.95 (pbk); $75, £52 (hbk)

Peter A. Abrams

During the past 50 years, physicists have sought a ‘unified theory’ from which all major categories of forces could be derived. Any physicist sufficiently intrigued by the title of Stephen Hubbell’s book to wade through its pages will, however, come away disappointed. This will be partly a result of the constraints on achieving any generality and unification in the field of ecology, and partly because of the limitations of Hubbell’s theory. Nevertheless, Hubbell’s book contains many interesting ideas and intriguing new findings, and few practising ecologists will fail to get something out of it.

The ‘unification’ of the title refers to the fact that this theory seeks to predict both the number of species found in a community and the distribution of the relative abundances of those species. Hitherto, these two aspects of biological communities have been explained independently by most (but not all) theoretical biologists. The number of species and their distribution of abundances are clearly dependent on some of the same processes. An extinction reduces the number of species as well as removing one of the species previously in the category of least abundance. Nevertheless, the ‘best’ theory for species-abundance distributions might not effectively account for species number, or vice versa.

In any event, vast expanses of ecological endeavour are totally independent of these on an assumption that the vast majority of ecologists believe to be almost universally false. He assumes that every individual in every species in a biological community is identical, and that the total abundance of all species is fixed. In his favoured version of the theory, each individual of each species has a low probability of mutating into a new species. All changes in distribution and abundance occur because of purely random variation in births, deaths, migration and speciation. This is what makes the theory ‘neutral’.

Hubbell has long championed the idea that the many tree species in the tropical forests where he works coexist largely because they are competitively equivalent. Random changes in abundance occur slowly enough for speciation events to have replaced the rare, random extinctions. The Unified Neutral Theory of Biodiversity and Biogeography generalizes this idea and derives many consequences from the assumption that all individuals are equal. Unfortunately, decades of experiments studying hundreds of species pairs have identified no conclusive cases of competitive equivalence, so one would expect such an assumption to lead to nonsensical consequences.

The surprising thing about Hubbell’s theory is that it is pretty accurate in its predictions of many attributes of ecological communities — particularly the distribution of abundances of tree species. On the other hand, this ability is perhaps not as surprising as it initially appears; the parameters determining the shape of this distribution — speciation rate and number of individuals in the set of connected communities — are largely immeasurable. Most of the fits are achieved by adjusting these parameters (or their product) to minimize the difference between theory and observation.

There are two main resolutions to the apparent contradiction of small-scale experiments and Hubbell’s global theory: first, non-neutral theories would be capable of fitting the data at least equally well; second, the assumption of equality is close to correct. But Hubbell largely ignores the first possibility. And his arguments for the second are not presented until the final 15 pages of the book, and I at least found them unconvincing. Graham Bell, another biologist who has recently analysed ‘neutral’ models, also favours the second alternative. However, work by Steinar Engen and collaborators indicates that alternative models are likely to be equally capable of fitting observed species-abundance relationships.

Biodiversity: Hubbell’s predictions are surprisingly accurate for tree species.

questions of species number and statistical distributions of abundance. A fisheries manager will find nothing here to help predict the sustainable harvest of cod, nor will an entomologist derive anything about the characteristics of an effective natural enemy of a devastating pest. The unification proposed by Hubbell has limited scope.

Attempts to construct unified theories have produced laws that, at least initially, appear to be universally applicable. Thus, our hypothetical physicist reader will be surprised to find that Hubbell’s theory is based
One of the likely consequences of Hubbell's and Bell's arguments is that there will probably be studies to resolve this issue in the near future. Theory must build on an understanding of simple cases and limiting assumptions, and Hubbell's deductions from 'neutral' assumptions have provided a rigorous basis for future work. At a time when books are often bland reviews of papers published many years before, this book has a high proportion of previously unpublished results and stimulating ideas. It is unlikely to be ignored.

Peter A. Abrams is in the Department of Zoology, University of Toronto, 25 Harbord Street, Toronto, Ontario M5S 3G5, Canada.

---

**Fertility from a metabolic viewpoint**

*On Fertile Ground: A Natural History of Human Reproduction* by Peter T. Ellison Harvard University Press, 2001. 358 pp. £16.95, $27.95

**Roger V. Short**

Peter Ellison is Professor of Anthropology and Dean of the Graduate School of Arts and Sciences at Harvard. He is best known for his exploitation of methods for measuring reproductive hormones in human saliva, thereby making it possible to carry out non-invasive studies of the reproductive endocrinology of people living in remote parts of the world, such as the Ituri rainforest in Central Africa or the mountains of Nepal. This book therefore describes some of his work on the fertility of traditional human societies.

Unfortunately, *On Fertile Ground: A Natural History of Human Reproduction* does not live up to its grandiose title; it would have been better called "Females, fatness and fertility", as this is its central theme. The opening chapters present a superficial (and sometimes incorrect) account of human reproduction, with strained analogies: "the area at the base of the brain called the hypothalamus and an associated bit of glandular tissue called the pituitary gland that hangs off the base of the brain like a tiny holiday ornament." It is excessively and unnecessarily North American in its outlook and citation of references. For example, pages are devoted to the bizarre and discredited views of Margie Profet, who postulated that menstruation was designed to cleanse a woman's body. Despite this, Ellison shows in the 1980s that sucking frequency and duration are unquestionably related to the duration of lactational amenorrhoea. In its place, he proposes a "strong alternative hypothesis, in which "relative metabolic load" is said to be responsible for the lactational inhibition of ovulation. Although we are given no clues as to how the body might sense this "relative metabolic load", or how this information is then used to inhibit ovulation, we are assured that his hypothesis fits the available data better than the nursing frequency hypothesis, and is sounder theoretically! How come that high-yielding dairy cows in negative energy balance, which are milked only twice a day, show little inhibition of ovulation postpartum, whereas low-yielding beef cattle, frequently suckling their calves, have a far longer period of lactational anovulation?

There is a good rebuttal of Rose Frisch's hypothesis that the percentage of body fat is the trigger for the onset of female puberty. But Ellison then goes on to propose an alternative "pelvic size hypothesis", suggesting (without any evidence) that there are mechanical determinants of the timing of puberty. Curiously, there is no mention of the timing of the onset of spermatogenesis in boys, or how it is that boys grow to be considerably taller than girls.

Indeed, there is no mention of the fact that "sperm counts above 15,000 or 20,000 per millilitre of semen are considered 'normal'". And I was amazed to learn that "the thermal oscillations produced in the mucin strands by the woman's body heat also resonate with the beating of the sperm cells' tail to aid its progress," although no reference is given to the source of this incredible information.

Sadly, this book is not the place I would recommend anybody to go for a comprehensive and factually correct account of the "natural history of human reproduction".

Roger V. Short is in the Department of Obstetrics and Gynaecology, University of Melbourne, 132 Grattan Street, Melbourne 3053, Australia.
Science in culture

Through a looking glass
Did major artists use optical devices to plot points to help them paint?

Rex Dalton

Last year, as artist David Hockney was working on his soon-to-be-published book on lens use in creating paintings, a common fascination with optics brought him together with Arizona physicist Charles Falco. The resulting collaboration may leave a lasting imprint on the worlds of art and science.

Hockney holds, controversially, that before the advent of photography in the mid-nineteenth century, major European artists used optical devices for over 500 years to help capture exactness of expression or perspective in their pictures.

Hockney's best known works include huge collages constructed of hundreds of small photographs, giving him a special eye for creating pictures at some distance with a lens. In 1999, he started to develop his theory of lenses as artists' aids after attending a London exhibition of the works of the early nineteenth century French painter Jean-Auguste Dominique Ingres, where he was struck by their extraordinarily confident detail. Like a detective with a refined draughtsman's eye, Hockney then began examining paintings to build a wealth of evidence from the Renaissance to the nineteenth century. Artists he contacted used a lens include Caravaggio, Velázquez, van Eyck, Holbein, Leonardo and Lotto.

Hockney's views are controversial on several counts. Ideologically, they offend a traditional notion that the creative process is independent of any form of copying. And on the practical front, there is no documentation of optical devices being used. But Hockney argues that few artists — if any — give away their individual tricks.

With optical science and mathematical formulas, Falco, a condensed-matter physicist at the University of Arizona, Tucson, has helped to affirm many of Hockney's theories. After an article in the New Yorker (31 January, 2000) described Hockney's theories, Falco contacted the artist, offering his assistance. In a lengthy series of correspondence — much of which is included in Hockney's book — Falco buttressed Hockney's artistic finds with scientific facts. In particular, Falco showed Hockney that a curved mirror is itself a lens, something not widely understood by art historians. This allowed Hockney and Falco to use mirrors to recreate centuries-old conditions to mimic how paintings may have been created.

For instance, if the geometry of a painted scene and the size of the canvas are known, the focal length of the lens can be calculated. And if the focal length and the depth of field are known, the lens diameter can be determined. Then a lens can be created to test its use for a painting.

In one of the more vivid examples, Falco helped to dissect an AD 1543 painting by Lorenzo Lotto entitled Husband and Wife. In a fax to Hockney, reproduced in the book, Falco calls it “a smoking gun” providing “extremely strong scientific evidence” for Hockney's theory. With assumptions based on measurements of the subjects in the painting, Falco calculated fairly precise properties of a lens for the painting. The focal length of the lens was determined, then the lens's curvature. Such lenses were available more than 100 years before Lotto's painting. Knowing the theoretical properties of the lens, Falco reconstructed it, and used it to recreate distortions in a pattern in a tapestry in the painting, showing how Lotto may have sketched the piece.

Falco, who is an art collector and aficionado of industrial design, particularly classic motorcycles, says he is enjoying the opportunity of a lifetime. Working with Hockney has opened up new research vistas for him in applied optical sciences. The feeling is mutual. “He really gave me a boost,” says Hockney of Falco. “He allowed us to confirm things.” For a BBC documentary to be shown in October, Hockney went to Florence, Italy, to examine the work of Filippo Brunelleschi, universally credited with first creating perspective in about AD 1420. “The birth of perspective came from an optical projection,” says Hockney, shored up by a re-enactment in Florence. “Art historians have never said that, but the evidence is very strong; the science can’t be denied.”

Some prominent art historians are impressed by Hockney's theories and writing, although not all are convinced by every example. “It is absolutely fascinating what Hockney is doing,” says Martin Kemp of Oxford University. “He is making us look at things afresh.” Falco's contribution, he adds, “is most valuable. It is not proof in the strictest sense, but it is supportive and shows no contradictions.” But there are doubts among art historians such as Walter Liedtke, a curator of European paintings at the Metropolitan Museum of Art in New York City. “I think Hockney goes too far with his ideas,” says Liedtke.

It is important to remember that “each artist has a personal style,” notes Liedtke. “We have a brain; the camera or lens does not. Experience and imagination will always intervene.” Hockney, of course, agrees on the importance of style. But he feels he is unmasking a secret artistic technique.

Rex Dalton is Nature's West Coast correspondent based in San Diego.
Victor and victim

The true message of Frankenstein is about morality, not mad science.

Howard P. Segal

T

he increasing use of the word "Frankenfood" by critics of genetically altered food is merely the latest instalment in the unending saga of anti-scientific sentiments allegedly originating with Mary Shelley’s Frankenstein (1818). It bespeaks the novel’s continuing influence that so many immediately recognize the derivation of the term. Yet ‘Frankenfood’ also echoes persistent distortions of the novel. Those who use Frankenstein to bash not just biotechnology but science overall have apparently never read the book, or have never read it carefully.

In truth, Frankenstein is hardly a Luddite tract — its message is not a call to destroy laboratories or experiments in the manner of the legendary English machine-breakers of the same period. Nor is its message akin to the famous 10-year moratorium on all scientific research proposed by the Bishop of Ripon in 1927.

Instead, Frankenstein insists that scientists must take moral considerations into account before, during and after research creating and destroying life in the laboratory. Victor builds an eight-foot-tall being simply because larger body parts are easier to work with. He never ponders his creature’s appearance until he brings it to life.

Indeed, it is critical to the novel’s message that the creature is nameless. Contrary again to so many movie and stage versions, the name ‘Frankenstein’ belongs not to the creature but only to Victor and his family. Abandoned by Victor at ‘birth’, the unnamed being later compares its miserable plight to God’s loving creation of Adam.

Yet it is too easy to characterize Victor as the quintessential mad scientist of most popular treatments. Rather, Shelley portrays him as extraordinarily self-centred. Victor himself notes that he is “not recording the vision of an amadman”. Instead, from the age of 15 until his death, he repeatedly blames “fate” for all his misfortunes. As Shelley understood, a truly mad scientist might escape moral responsibility for his actions.

Popular culture has also misrepresented the creature as an uncaring monster who kills innocent people without remorse. By now, ‘Frankenstein’ has nearly become a generic term for monster. In the novel, however, the creature feels guilty about its deeds and ironically, apart from appearance, is far more appealing than Victor. In fact, it is self-educated, sensitive and articulate, unlike the ignorant, grunting beasts of most popular versions. The creature is thereby capable of becoming Victor’s missing moral compass.

If Victor blames “fate”, the creature blames itself, planning to build a bonfire and commit suicide as penance for its actions. H ad Shelley believed that scientists should not explore the “cause of generation and life”, she would surely have portrayed a genuinely moral monster.

Frankenstein insists that scientists must take moral considerations into account and assume responsibility for the outcomes of their experiments.

© 2001 Macmillan Magazines Ltd

Howard P. Segal is in the History Department, University of Maine, Orono, Maine 04469-5774, USA.

NATURE | VOL 412 | 30 AUGUST 2001 | www.nature.com
Reasoning for results

Dennis Bray

Let’s start on safe ground. We all agree, surely, that theory — the formulation of hypotheses — is important in biology. Techniques are essential, as is the careful collection of quantitative data. But without ideas to give them shape and meaning, those endless successions of base sequences, expression profiles, electrical recordings and confocal images are as featureless as a plate of tofu. All really big discoveries are the result of thought, in biology as in any other discipline. Allostery, genes, DNA structure, chemi-osmosis, immunological memory, ion channels were all once just a twinkle in someone’s eye. And the work of most contemporary research laboratories still takes place within a framework of hypothesis, although practitioners may not always recognize this fact. As Charles Darwin once remarked: “How odd it is that anyone should not see that all observation should be for or against some view if it is to be of any service.”

But assuming that biological theory exists, does it therefore follow that theoretical biology is a distinct and legitimate subject? My guess is that there is less agreement on this proposal; there may even be virulent opposition to it. One of the most common arguments against it is that living systems are so prodigal, so unpredictable and, above all, so historical that any attempt at a grand theoretical framework is doomed to failure. The shape of a leaf in a forest is the product of so many chance events, not only during its growth but also in the aimless blundering of its evolution, that it is inconceivable that it could be predicted from first principles. An all-encompassing theory of biology is no more possible than a predictive theory of other large, inchoate systems, such as the weather, the stock market or London’s Heathrow Airport.

Moreover, the argument goes, even in cases in which laws and patterns have been observed in living systems — as in zebra stripes, Hodgkin-Huxley equations and protein coiled-coils — these are not truly biological laws. They belong to mathematics, chemistry or physics, and reflect what living systems have in common with the non-living world, not their own unique biological characteristics.

One can argue with these views, but they are rational and worthy of debate. What is harder to understand is the irrational opposition to theory that often surfaces. Many experimentalists seem to regard theoreticians as carpetbaggers. “We spend months in the lab getting data,” they say, “and then along comes some character who doesn’t know how to hold a pipette and explains our results to us.” Apparently, you’re not licensed to theorize unless you put the time in and get the data. This is unfortunate because people are good at different things, and some really enjoy reading papers, juggling possibilities and formulating ideas, even if they can’t work a pipette. It is true that certain ‘biological theorists’ in the past have indeed been carpetbaggers, rushing into print with ill-conceived and carelessly prepared ideas, and this gave the subject a bad name. But this is a sociological problem, not a scientific one. The cure is to embrace theoretical work and let it become part of the mainstream of biological research. The quality and accuracy of predictions will then inexorably rise.

Returning to the chase, the term ‘theoretical biology’ seems to me just silly—a frozen joke, an oxymoron. Whoever coined the phrase (I’d love to know who it was) was probably being deliberately provocative by drawing parallels with theoretical physics. But although the name is misbegotten, the thing it has come to represent is healthy and growing like a weed. Surfacing briefly in a seminar the other day, I realized that I had no idea whether the traces on the screen were actual electrical recordings made in a physiological experiment or the output of a computer program. It was impossible to tell.

Computer models of action potentials, synaptic integration, heart contraction and even the movements of ions and molecules in cells are now so accurate that they can often be used as experimental objects in lieu of the thing they represent. Biologists can now design and test small genetic circuits in cells. It seems inescapable that, at least at the level of molecules and cells, biology is moving from an era of data-collection to one of hypothesis-driven research. Progress in this new field will be driven by informed and increasingly quantitative theories — whatever name we choose to give it.

Dennis Bray is in the Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK

Some people enjoy reading papers, juggling possibilities and formulating ideas, even if they can’t work a pipette.

Theoretical biology

Theory needs to be embraced and to become part of the mainstream of biological research. The quality and accuracy of predictions will then inexorably rise.

Further Reading

Mayr, E. This is Biology (Harvard Univ. Press, Cambridge, MA, 1997).

Virtual cell: http://www.nrcam.uchc.edu
Achilles’ heel of cancer?

Bert Vogelstein and Kenneth W. Kinzler

The p53 protein is inactivated in most human cancers. One outcome is a defect in controlling cell division. Might a virus that exploits this defect prove useful in treating cancer?

A decade or so ago, it was discovered that the p53 protein is subtly mutated, and thereby inactivated, in almost all types of human cancer. These findings raised the exciting possibility that drugs that restore p53’s function would be broadly applicable cancer treatments. That hope has not yet been realized, but there have been some imaginative approaches to the problem—the latest of which is described by Raj and colleagues on page 914 of this issue.

For ten years after its discovery in 1979, the p53 gene was thought to be an oncogene. Under normal circumstances, the protein products of oncogenes stimulate appropriate cell division or interfere with cell death. However, certain mutations in these genes result in their products being switched on all the time, or gaining a new function. This generally leads to inappropriate stimulation of cell growth, resulting in tumours. All commonly used drugs inhibit the ability of their target proteins to bind to other molecules or to catalyse reactions, and so mutated oncoproteins are, in theory, perfect drug targets. Theory has been put into practice with the development of STI-571, a drug that inhibits the aberrantly activated enzyme found in chronic myeloid leukaemia cells.

Yet most genetic alterations in common human cancers affect tumour-suppressor genes, rather than oncogenes. The normal function of a tumour-suppressor gene is to keep cell numbers down by stopping cells from multiplying or by promoting cell death. When mutated, tumour-suppressor proteins are switched off, and again the result is inappropriate net cell growth. In some cancers, an entire tumour-suppressor gene may be deleted. So tumour suppressors are poor targets for conventional drugs—one cannot inhibit an activity that is not there. Unfortunately, the same work that identified p53 as a potential target for anticancer drugs also showed that it behaves genetically like a tumour-suppressor gene, not an oncogene. One of the main functions of p53 is in the control of cell division. In the presence of damaged DNA or other stresses, p53 prevents cells from progressing from a resting phase of the cell-division cycle (G1 phase) to the DNA-replication phase, or from a second resting phase (G2 phase) to nuclear division (mitosis). In this sense, p53 is a true “guardian of the genome”, and its loss can have severe consequences under certain conditions.

There are several approaches to restoring normal p53 function (Box 1). The most direct involves gene therapy to transfer a wild-type p53 gene to cancer cells that express mutant p53. Although this has shown promise in animals and humans, it is currently impossible to deliver any gene specifically to cancer cells, and externally derived p53 is likely to be as harmful to normal cells as to cancer cells. So gene therapy with p53 will probably be highly toxic unless more specific modes of delivery can be developed.

Other ambitious strategies aim to convert mutant p53 protein into its wild-type form. Normally, one end of the p53 protein (the carboxy terminus) folds back to inhibit its central, DNA-binding region. Peptides have been developed that interfere with this process and in so doing can tickle certain mutant p53 proteins to adopt an active conformation. Small compounds that interact directly with the DNA-binding domain and partially activate mutant forms have also been identified, and sophisticated RNA-based enzymes have been developed for similar purposes. Although more potent substances must be developed, this approach is exciting because the compounds should be specific to cancer cells—no normal cell contains mutant p53.

Another angle of attack is provided by the fact that mutant p53 proteins accumulate in cancer cells, whereas normal cells have little p53 unless they are stressed (for example by DNA damage). So, mutant p53 can be decorated with proteins that are toxic only when bound to p53, providing a treatment that would theoretically be harmless to normal cells. It might also be possible to stimulate

**Box 1 Using p53 to kill cancer cells**

The p53 protein is a tumour suppressor—it keeps cell numbers down by stopping cells from multiplying or by promoting cell death. Loss of p53 occurs in most human cancers, so it would be useful to be able to restore its function. Several innovative strategies have been suggested:

- Introduce normal p3 genes into a cancer cell with mutant p53.
- Introduce a small compound that converts mutant p53 proteins from an abnormal to a normal shape.
- Add a protein that attaches itself to mutant p53 and kills cells.
- Stimulate the host’s immune response to mutant p53 peptides.
- Introduce drugs that disrupt the interaction between the MDM2 or E6 proteins and p53.
- MD2 and E6 negatively regulate p53; they are present at abnormally high levels in some cancer cells, so ‘quench’ any normal p53.
- The strategy described by Raj et al. introduces the adenovirus-associated virus, which mimics damaged DNA. Cells with mutant p53 cannot activate the usual p53-dependent ‘checkpoint’ that is induced by DNA damage, and eventually die.
- Infect cells with viruses that can replicate only in cells without normal p53; the viruses kill these cells.

---

**Figure 1** The strategy used by Raj et al. to kill cells that lack p53. This is one of many approaches being developed to restore p53 function to cancer cells (Box 1).
patients to produce antibodies that react with cells containing mutant p53 (ref. 14).

Inside cells, p53 is bound to other proteins that negatively regulate it. For example, the MDM2 protein binds to p53 inside the nucleus, enabling p53 to be exported from the nucleus into the hands of a protein-degrading machinery known as the proteosome. In some types of cancer (liposarcomas, for example), p53 is normal but there are too many copies of the MDM2 gene, so there is more MDM2 protein around to inactivate p53. The E6 protein from human papilloma viruses has a similar function, leading to cervical cancer. Drugs that interfere with the binding of MDM2 or E6 to p53 might free p53 from these protein shackles, suppressing sarcomas or cervical cancers.

The strategy outlined by Raj et al. is likewise directed not towards p53 itself, but (in this case) towards the consequences of p53 action (Fig. 1). The authors infected cultured cells with adenovirus (AAV) and found that this activates the p53 checkpoint, presumably because the unusual genome of AAV, which contains single-stranded DNA and DNA hairpins, mimics damaged DNA. Indeed, a DNA molecule corresponding to the AAV hairpin, but which contains no genes, had the same effect. Cells containing normal p53 paused in the G2 phase to rid themselves of the AAV genome, and then began dividing again. But cells without functional p53 could not maintain their arrest in the G2 phase, and instead began a catastrophic nuclear division that led to cell death. The AAV strategy also proved successful in shrinking experimentally induced cancers in mice.

This approach might be more specific than standard DNA-damaging drugs, which interact with a plethora of cellular constituents and have diverse toxic effects. It would not matter if an AAV entered normal cells, because these cells have functional p53. One problem, however, is that the virus is inactivated and non-infectious; because it cannot spread from cell to cell, it must be delivered specifically to almost every cancer cell.

There is a viral-based strategy that potentially solves part of this delivery problem. The ONXY-015 adenovirus was originally developed as an anticancer agent that can replicate only in p53-deficient cells, killing them directly. Although the replicative abilities of this virus are more complex than once thought, the idea of a live virus that can infect p53-deficient cancer cells but cannot multiply in normal cells is attractive. Infection of only a few cancer cells could theoretically spread to the whole cancer. But there are drawbacks. It is not trivial to infect even a small fraction of tumour cells by systemic administration (for example, by injecting adenoviruses intra-nasally), and the host’s immune system may fight back. Immune reaction to the activated AAV might not be such a problem.

No cancer patient should expect any of these innovative approaches to cure them in the near future. But continued research into the structure of p53, the proteins to which it binds and the pathways it controls should eventually lead to the maturation of one or more of these strategies — each of which has the potential to benefit millions.

Bert Vogelstein and Kenneth W. Kinzler are at the Howard Hughes Medical Institute and Johns Hopkins Oncology Center, Baltimore, Maryland 21231, USA.


Quantum optics

Photons yield to peer pressure

Paul Kwiat

In 1960 the invention of the laser allowed classical light to be amplified inside a cavity. New experiments show that photons in a special quantum state can also be amplified.

Quantum entanglement between two particles means that measuring the behaviour of one instantly determines the behaviour of the other, even when they are physically far apart. Erwin Schrödinger once described this peculiar connection as “the characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought”. Now, on page 887 of this issue, Lamas-Linares, Howell and Bouwmeester report some of the first basic experiments on amplifying entangled photons through a process analogous to that occurring in lasers. Because entanglement is a key ingredient in many quantum information procedures, such as quantum computing, teleportation and cryptography, a method for reliably producing entangled states of several photons could have significant ramifications, especially for quantum communications and optical techniques.

Entanglement describes a system with several components in which the individual parts carry no information but nevertheless share quantum correlations with each other that are stronger than those allowed by classical physics. For example, photons can be polarized — the polarization describes the oscillation direction of the electric field associated with a light wave. Polarization filters, such as Polaroid sunglasses, will let through photons polarized in one plane but block those polarized at right angles, and so can be used to measure photon polarization. If two photons have entangled polarizations, each photon individually would appear completely unpolarized (with no particular oscillation direction) and yet measuring the polarization of one completely determines the polarization of the other. It is as if you flipped two coins, each of which was equally likely to come up heads or tails, and yet they always gave the same results — that is, both heads or both tails.

Although normal coins do not behave like this, it has been known for some time how to produce pairs of photons that display such bizarre quantum mechanical correlations. Ultraviolet photons sent into a nonlinear optical crystal will sometimes split into two infrared daughter photons (a process known as ‘spontaneous downconversion’), each travelling in its own beam, which are polarization entangled. Unfortunately, the likelihood of each photon splitting in this way is less than 1 in 10 billion. So even though typical experiments use pulses that have several billion ultraviolet photons in them, often not even one will lead to a downconversion pair. The chance of two of them splitting, which can lead to an entangled quartet, is much smaller still: previous experiments have reported four-photon counting rates of only 1 to 2 events per minute6–8 (out of 5 billion pulses per minute).

Lamas-Linares et al. have significantly increased this rate by using the entangled output of one crystal as a seed for ‘growing’ more entanglement in another crystal (actually the same one, used a second time). Without the seed the second crystal would produce an entangled beam of about the same brightness as the first crystal. By adding the seed photons, one might naively expect production rates simply equal to the sum of the rates for each process individually (that is, a doubling of the...
output of each crystal alone). Not so, thanks to the magic of quantum interference (Fig. 1a) and stimulated emission.

Because the two downconversion processes can interfere, the production of photon pairs can be enhanced by a factor of four (Fig. 1b). The enhancement of downconversion was first demonstrated in 1994 (ref. 7), but the new work extends this to pairs that are polarization entangled. Even more striking is the enhancement effect on four-photon processes (Fig. 1c). The simple interference analogy would again lead one to predict a maximum enhancement factor of four. But the enhancement for four-photon entanglement is higher still — a factor of sixteen — owing to the increased efficacy of stimulated emission as the number of photons increases. Loosely speaking, the photons experience more peer pressure when there are more photons to pressure them into conforming. Lamas-Linares et al.’s results clearly show the predicted enhancement, evidence that they have observed the stimulated emission of entangled radiation.

A word of caution: the extra enhancement by a factor of four of the quartet-photon state relative to the twin-photon state should not be misinterpreted to imply that creation of the former is more likely than creation of the latter. Lamas-Linares et al. report two-photon counting rates of over 10,000 per second, but four-photon counting rates of only 1 per second. Still, this is a 50-fold increase over the results of initial experiments done only a few years ago4–6. Equally important, the quality of the four-photon interference has seen a similar improvement — the interference contrast has been raised from about 85% to 97%.

The experiment of Lamas-Linares et al. is only a first step toward producing a ‘laser’ of entangled photons — strictly speaking, it is the analogue of a laser cavity in which the light is cycled just twice, instead of tens or hundreds of times. It will be interesting to see how far the technology can be pushed. Unfortunately, the entanglement is most visible only if a definite number of photons can be identified in each of the two beams. This becomes more and more difficult as the number increases (because of photon loss and detector inefficiency), and could limit the usefulness of the approach. Nevertheless, just as no one knew in 1960 all the possible applications of the optical laser, it can safely be said that the potential of an entangled-photon laser for quantum information applications is also largely unknown.

The authors’ suggest that this work may lead to some new form of quantum cryptography, with higher transmission rates than are presently available using entangled photon twins8–10. I think this is unlikely, both because any loss will reduce the entanglement, and because the generation rate of these multiple-pair pulses is so much lower than the generation rate of single-pair polarization-entangled photons. Entanglement might also be used to improve the signal-to-

Figure 1 In quantum mechanics, one plus one can equal zero, or four. a. An optical interferometer consists of two mirrors and two beamsplitters, which reflect half the light and transmit the rest. If we look at one output of the interferometer, with either path blocked we detect one-quarter of the light sent into the interferometer. If we call the contribution from each path i, we might expect that with both paths open the detector would see 2i. But in fact the intensity on the detector can range from 0 (if the contributions from the two paths interfere destructively) to 4i (if the contributions from the two paths interfere constructively). b. In a similar process, infrared downconversion pairs may be produced during the first passage of the ultraviolet (UV) photons through the crystal, or during their return passage. Because these indistinguishable processes can interfere, the final output intensity of pairs can vary from 0 to four times the pair intensity from a single passage through the crystal. c. Lamas-Linares et al.1 show that further enhancement is possible with the four-photon process — going through the crystal twice increases the rate of four-entangled photon production 16-fold.

100 YEARS AGO
Mr. W. W. Davis has a paper in “Studies from the Yale Psychological Laboratory”… on some relationships between temperament and effects of exercise. His tests and observations are scarcely sufficient to establish very definite relations, but the conclusions at which he arrives are not without interest. The observations suggest that nervous persons, in training for the development of strength, require light practice, and phlegmatic persons require vigorous practice. The phlegmatic type of temperament is apparently characterised by the presence of much reserve energy of muscle and nerve cell. The nervous type has less reserve energy but a greater ability to use the energy at hand. It is not difficult to apply these principles to practical physical training. They make necessary on the part of the trainer a knowledge, secured either by means of observation or experiment, of the temperament of each man under his charge. From Nature 29 August 1901.

50 YEARS AGO
The ground-state of the hydrogen atom is a hyperfine doublet the splitting of which, determined by the method of atomic beams, is 1,420.405 Mc./sec. Transitions occur between the upper (F = 1) and lower (F = 0) components by magnetic dipole radiation or absorption. The possibility of detecting this transition in the spectrum of galactic radiation, first suggested by H. C. van de Hulst, has remained one of the challenging problems of radio-astronomy. In interstellar regions not too near hot stars, hydrogen atoms are relatively abundant, there being, according to the usual estimate, about one atom per cm. Most of these atoms should be in the ground-state. The detectability of the hyperfine transition hinges on the question whether the temperature which characterizes the distribution of population over the hyperfine doublet — which for want of a better name we shall call the hydrogen ‘spin temperature’ — is lower than, equal to, or greater than the temperature which characterizes the background radiation field in this part of the galactic radio spectrum. If the spin temperature is lower than the temperature of the radiation field, the hyperfine line ought to appear in absorption; if it is higher, one would expect a ‘bright’ line; while if the temperatures are the same no line could be detected... We can now report success in observing this line. From Nature 1 September 1951.
news and views

Cancer, blindness in diabetics and arthritis

Contributes to numerous diseases, including response to specific signals. This process is by stimulated emission.

Peter Carmeliet

Creating unique blood vessels

When tissues need more oxygen, they release molecules that encourage blood vessels to grow. The discovery of the first such molecule that is specific to one type of tissue has implications for cancer and heart failure.

On a farm, systems of hosepipes are used to deliver vital water to every field. Similarly, in our bodies, blood vessels deliver essential oxygen and nutrients to each organ. New blood vessels form when the endothelial cells that line the interior of the vessels, and the smooth-muscle cells that form a jacket around the outside, grow in response to specific signals. This process is known as angiogenesis and, when deranged, contributes to numerous diseases, including cancer, blindness in diabetics and arthritis. One might assume that it would be irrelevant whether the blood vessels in different organs are alike or not, just as it would not matter what type of hosepipes irrigate the farm, as long as all fields receive the water they require. Yet, beyond their basic need for oxygen and nutrients, different tissues make different demands on their blood vessels. This suggests that, as well as the general signals (such as vascular endothelial growth factor, VEGF) that control angiogenesis in all organs, there should also be tissue-specific angiogenic molecules. On page 877 of this issue, LeCouter and colleagues describe their discovery of the first such molecule. Their results imply that other tissue-specific angiogenic signals exist. If so, this would create many new — and probably safer — opportunities for stimulating or inhibiting angiogenesis in diseased tissues without affecting healthy organs (Box 1).

Because different tissues have distinct needs, the blood vessels that supply them also differ (Box 1). Take, for instance, the vessels in endocrine glands — organs such as ovaries and testes that produce steroid hormones involved in growth, metabolism, stress, reproduction and sexual development. These hormones must be able to reach the bloodstream, so vessels in endocrine glands are leaky and their endothelial cells contain fenestrations (tiny gaps through which fluid and small solutes can pass). In contrast, endothelial cells in the brain are tightly linked to each other and are engulfed by numerous periendothelial cells, which constitute a barrier that protects nerve cells from potentially toxic molecules from the blood.

VEGF is involved in the growth of fenestrated blood vessels. But, given that it is also expressed near non-fenestrated vessels, it cannot be the only factor involved. Moreover, when endothelial cells from endocrine glands are cultured in a dish, they lose their fenestrations. This plasticity suggests that endocrine glands produce signals that are essential to maintain the specificity and growth of their endothelial cells.

LeCouter et al. have now identified such a molecule, which they call endocrine-gland-derived VEGF (EG-VEGF). This molecule functionally resembles and complements VEGF in its ability to induce the formation of endothelial fenestrations. However, EG-VEGF and VEGF are structurally dissimilar and probably work through different receptors. Moreover, VEGF affects endothelial cells non-selectively, and even acts on non-endothelial cells such as motor neurons. EG-VEGF, on the other hand, affects only endothelial cells in endocrine glands. It is likely to be the first member of a large class of tissue-specific vascular growth factors.

Box 1 Blood-vessel growth factors

The properties of general and specific molecules that promote the growth of blood vessels (angiogenesis). Vascular endothelial growth factor (VEGF) is an example of a general molecule, produced by most tissues. Endocrine-gland-derived VEGF (EG-VEGF) is the first tissue-specific angiogenic molecule to be identified; its discovery suggests that there may be others. Tissue-specific angiogenic molecules are needed because blood vessels differ. Bottom, blood vessels in endocrine glands need to be leaky, and have pores known as fenestrations through which hormones, produced by the endocrine glands, can enter the bloodstream. VEGF and EG-VEGF cooperate to induce the growth of fenestrated blood vessels. Top, endothelial cells in the brain are very different: they do not have fenestrations, and are ensheathed in an impermeable layer of periendothelial cells (pericytes), which prevent potentially harmful molecules from leaking out of the blood. Different vessels vary in the number of pericytes per endothelial cell.
quantum world

Juan Pablo Paz

When quantum systems interact with the environment, classical properties emerge — a process known as decoherence. Although decoherence is unavoidable, it may still be possible to manipulate the outcome.

The world around us looks classical even though the fundamental laws of physics are based on quantum mechanics. At the atomic level, electrons and protons are blurred entities that cannot be described as point-like particles following trajectories. But macroscopic objects have well defined properties: they are either here or there, and not everywhere. So how does the classical world arise from the quantum?

The consensus today is that classical behaviour is an emergent property of quantum systems, induced by their interaction with the environment. This interaction, at the heart of life for complex macroscopic objects, is responsible for the process of decoherence, which is the biggest obstacle to building a viable quantum computer. An attractive idea to control decoherence is to manipulate the environment — so-called environment engineering. As Carvalho et al. describe in Physical Review Letters, a cold trapped ion could be manipulated in this way by applying a number of judiciously chosen laser fields to create an environment in which the stable state of the ion can be chosen (almost) at will.

Decoherence makes most of the states of a quantum system unstable, so that only a small subset of all possible states, the ‘pointer

news and views

These results provide new insight into how organs develop a blood supply — a question with significance for tissue regeneration and tumour growth. For tumours to grow or tissues to regrow (for example, during the healing of burned skin), they need oxygen, which at first reaches cells by simple diffusion from existing blood vessels. But when the gap between the rapidly multiplying cells and the blood vessels exceeds the limit of diffusion, distant cells become deprived of oxygen (hypoxic). These cells increase their expression of VEGF and other angiogenic molecules by activating hypoxia-inducible transcription factors, so attracting new blood vessels, which restore oxygenation. When such hypoxic upregulation of VEGF is impaired, angiogenesis in tumours is reduced and, surprisingly, motor neurons degenerate.

LeCouter et al. show that the expression of EG-VEGF is also upregulated by hypoxia, providing a molecular explanation as to how growing endocrine glands acquire not only more numerous but also more specialized blood vessels. Moreover, because arteries are produced from different precursors to veins, and smooth muscle cells are generated from distinct precursors in different organs, their results also raise the possibility that tissue-specific angiogenic molecules control these processes, too.

Do tissue-specific blood-vessel regulators contribute to the excessive angiogenesis that occurs in cancers and inflammatory disorders? We don't yet know, but we should perhaps hope that they do. LeCouter et al. show that the forced overexpression of extrinsic EG-VEGF in rat ovaries leads to the production of cysts, with excessive numbers of blood vessels nearby. If the abnormal expression of intrinsic EG-VEGF likewise contributes to angiogenesis in ovarian disorders, then it might be possible to develop drugs that selectively block this process. Inhibition of VEGF does block deregulated ovarian angiogenesis but, because VEGF is non-selective, such drugs might increase the risk of motor-neuron degeneration and heart failure because of poor blood supply. Blocking EG-VEGF would be expected to have fewer side effects.

On the flip side of the treatment coin, LeCouter et al.'s results might also have implications for inducing the growth of new blood vessels in blood-deprived (ischaemic) tissues. A good blood supply is crucial; for example, a lack of blood supply to the heart can cause heart failure. VEGF and other general angiogenic molecules have been tested for their ability to stimulate angiogenesis in ischaemic heart tissue, and have proved successful in animals.

But the downside of the long-term use of such general molecules is that they might also stimulate angiogenesis elsewhere — in hidden tumours, for example. Moreover, although these molecules work in animals, initial results from clinical trials reveal only modest long-term benefit, so it remains unknown whether VEGF alone can induce the necessary formation of stable blood vessels. Additional angiogenic molecules might be needed, such as placental growth factor, which amplifies the activity of VEGF in ischaemic but not normal heart tissue. However, this combination has not yet been shown to stimulate angiogenesis. Given that coronary vessels are derived from different precursors to vessels in other tissues, heart-specific angiogenic molecules might be necessary, and safer than more general molecules.

Research into angiogenesis has arrived at the critical stage of translating the findings from studies of mice to humans. LeCouter et al.'s study has raised the hope that we might be able to achieve this goal more rapidly, and more safely, than we thought.

Peter Carmeliet is at the Centre for Transgene Technology and Gene Therapy, Flanders Interuniversity Institute for Biotechnology, University of Leuven, B-3000 Leuven, Belgium. e-mail: peter.carmeliet@mei.kuleuven.ac.be

states, survive the interaction with the environment. So pointer states, dynamically selected by the environment, are the only ones in which macroscopic systems are allowed to exist. In the real world, pointer states are selected by nature. But in recent years, physicists have found valuable ways to modify the properties of the decoherence process in some systems. They do this by creating an artificial environment that is tailored to control the properties of pointer states.

Ignacio Cirac, Peter Zoller and co-workers originally devised the idea of environment engineering in 1996. Their work contained the basis for a general method that has been tested by impressive experiments in which an ion is stored and laser-cooled in a special trap. Ion traps use a particular configuration of oscillating electromagnetic fields to confine the ions to a well-defined region of space, and have been used successfully in high-precision spectroscopy, as well as to demonstrate the feasibility of quantum information processing. The trapped ion can be made to interact with an environment whose properties are designed by the experimenter.

Carvalho et al. propose a way of preparing pointer states that can persist even in the presence of the form of decoherence that is most effective in disturbing the motion of an ion in a trap: that generated by the interaction with random fields in the trap electrodes. Even with this disturbing interaction, the authors claim their method can be used to select a single pointer state for the center of mass of a trapped ion, which becomes stable under the effect of the artificial environment. This quantum state can be chosen at will by manipulating the frequency and intensity of the applied laser fields. The method will only protect a single stable state, so rather than specifying several pointer states, which may be relatively stable, the procedure prepares and protects a single specified state in the presence of natural decoherence. The proposal seems to be within the reach of current technology but would require ‘stiffer’ traps than those used in current experiments.

The method’s basic strategy is simple. Suppose that you want to prepare and protect a given quantum state, \( |\psi\rangle \). The first step of the procedure is to find an operator, \( D \), such that \( D |\psi\rangle = 0 \) (that is, \( D \psi = 0 \)). The authors then explain how to design the laser fields, which drive internal transitions of the ions into rapidly decaying states, in such a way that the artificial environment interacts with the ion precisely through the operator \( D \). The key feature of the method is that the net effect of the laser fields is to drive the irreversible decay of the motional state of the ion towards \( |\psi\rangle \). If the desired state is a superposition of energy eigenstates (with given coefficients), the method determines the necessary values of the amplitudes of the driving laser fields. There is a practical limitation to the method. To protect a state in which the ion occupies two positions at once separated by a distance of about 80 nanometres, it is necessary to apply about 15 laser beams to the ion. This is because the number of laser beams is proportional to the number of terms in the expansion of \( |\psi\rangle \). Increasing the separation into the mesoscopic domain becomes increasingly hard. This kind of quantum state (in which the ion occupies two locations at once) illustrates the most counterintuitive features of quantum mechanics, and is a laboratory cousin of the famous Schrödinger’s cat (whose fate is to be in a superposition of alive and dead states).

Thus, while the method does not seem to be useful for preparing mesoscopic or macroscopic Schrödinger cat states, it could serve to prepare and protect states of small and tender quantum kittens. The authors also describe in detail how to prepare other interesting quantum states.

A remarkable feature of the debate on the transition between quantum and classical behaviour is that, for the first time in its rather long history, experiments are probing this boundary and playing an important role. Environment engineering, as discussed by Carvalho et al., is a useful proposal that will certainly be tested soon in more detail. Moreover, there are other fertile proposals on how to fight decoherence. Many of these ideas, which were born in the context of studies on quantum information processing (such as quantum error correction), are also being examined experimentally. Most of these are ‘controlled decoherence’ experiments, which a few years ago appeared to be just thought experiments exploring the nature of the transition from quantum to classical physics. In the near future, a new generation of experiments, in which decoherence is caught in the act, will be able to probe the fuzzy border between the quantum and the classical worlds.

Juan Pablo Paz is in the Departamento de Física, Universidad de Buenos Aires, Pabellón 1, Ciudad Universitaria, 1428 Buenos Aires, Argentina.

e-mail: paz@df.uba.ar


Immunology

Antibody alterations

Alberto Martin and Matthew D. Scharff

In an immune response, antibody molecules are altered so they can bind to intruders more strongly. Some of the molecules that determine which alteration process predominates have now been tracked down.

One of the ways in which the immune system fights off intruders is to produce antibodies, which bind to and neutralize foreign molecules (antigens). The immune system — specifically B cells — must be able to generate enough different antibodies to recognize every possible antigen, so extraordinary antibody diversity is generated before exposure to foreign antigens. But these primary antibodies almost always have low affinity for their targets, and cannot neutralize pathogens or toxins. So, after exposure to an antigen, the variable (V) regions of the antibody genes, which encode the antigen-binding site, acquire many changes — some of which result in higher-affinity binding sites. In some species, this occurs by a process known as gene conversion; in others, by somatic hypermutation. It is not clear why these two different processes evolved, and why their species distribution is so sporadic. Moreover, the biochemical mechanisms underlying gene conversion and somatic hypermutation have been hard to identify. These questions formed the backdrop for Saleet al.’s work.

The authors used the chicken DT40 B-cell line, which is unusual among cultured animal cells because it can undergo high
rates of both homologous recombination and gene conversion\(^{15,20}\). Homologous recombination is the general name given to genetic exchange between two similar DNA sequences; gene conversion is a specific type of homologous recombination. Homologous recombination is also involved in repairing DNA lesions, and studies of yeast have shown that proteins known as RAD proteins are required\(^1\). In DT40 cells, inactivating the chicken RAD54 protein reduces both homologous recombination and gene conversion, and makes the cells highly sensitive to \(\gamma\)-irradiation\(^2\), which damages DNA. By contrast, the chicken XRCC2, XRCC3 and RAD51B proteins are also involved in homologous recombination, but have different functions from RAD54 — inactivating them results in only slight increases in cellular sensitivity to \(\gamma\)-irradiation, and greatly increases sensitivity to drugs that crosslink DNA molecules together\(^3\). Sale et al.\(^4\) looked at the effects of deleting RAD54, XRCC2, XRCC3 or RAD51B in the chicken B-cell line, and confirmed the need for RAD54 during gene conversion. But loss of the other three proteins had little effect on this process. The implication is that, unlike RAD54, these three proteins are not involved in the intrachromosomal homologous recombination that is required for gene conversion, but participate in other homologous-recombination pathways\(^5\). The authors suggest that these other pathways might in effect limit the level of somatic hypermutation in normal chicken B cells. They make this proposal because they found that, surprisingly, the deletion of XRCC2, XRCC3 or RAD51B led to very high rates of somatic hypermutation — much higher than the rates of gene conversion. How could this come about? The formation of double- or single-stranded DNA breaks is likely to be required for somatic hypermutation\(^6\)\(^-\)\(^9\). Such breaks could precede gene conversion, too. Sale et al. propose that many more breaks might be produced in the V regions than are revealed by the frequency of gene conversion. Some of these DNA breaks are repaired as a result of gene conversion; others are fixed in normal cells by a different homologous-recombination process (perhaps between matching genes on paired sister chromatids) involving XRCC2, XRCC3 and RAD51B. In cells that lack these proteins, the breaks persist and may trigger the mutational process.

It seems then, as suspected\(^1\), that the two mechanisms of V-region diversification can coexist in the progeny of a single B cell. As chicken B cells (and probably B cells from other species) have all of the factors required for both gene conversion and somatic hypermutation, why does one process predominate? This is part of the larger question of how particular sets of DNA-polymerase enzymes and accessory molecules gain access to the right DNA ends at the right time to carry out such distinct transactions as replication, repair and recombination in all cells\(^10\), and antibody rearrangement and diversification in B cells. As suggested by the effects of inactivating XRCC2, XRCC3 and RAD51B, these events are likely to be regulated by the abundance and activity of the relevant molecules.

For example, as mentioned above, several chicken B-cell lines that carry out gene conversion also display high rates of homologous recombination\(^6\)\(^-\)\(^9\). Perhaps gene conversion predominates over somatic hypermutation in these cells simply because the homologous-recombination enzymes are very active. Alternatively, gene conversion and somatic hypermutation might require different combinations of molecules to be recruited to the V region. Other factors are also likely to contribute, including the accessibility of the pseudogene and of the V region.

Finally, Sale et al.'s work\(^4\) highlights the advantage of working with cultured cell lines; they provide large, relatively homozygous cellular populations that can easily be genetically manipulated and subjected to detailed biochemical analysis. As the authors point out, the ability to readily inactivate genes in the DT40 cell line by gene targeting provides a powerful tool. By using it we should be able to discover the full spectrum of genes that are required for gene conversion and somatic hypermutation; to pin down the biochemical mechanisms involved; and to find out exactly how somatic hypermutation is targeted to the V regions of antibody genes.

Alberto Martin and Mathew D. Scharff are in the Department of Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461, USA.

e-mail: scharf@acm.yu.edu


**Chemistry**

**Mystery of an interstellar ion**

Annick Suzor-Weiner and Ioan F. Schneider

Hydrogen is the most abundant element in the Universe. Its chemistry governs most reactions in space, especially when it is ionized. But the abundance of the \(\text{H}_3^+\) ion is not easy to explain.

The elementary molecular ion \(\text{H}_3^+\) is a basic building block of interstellar chemistry and star formation. Yet its abundance in diffuse molecular clouds is hard to understand. Much of the debate about \(\text{H}_3^+\) has centred on its rate of destruction by interstellar electrons — a rate that varies wildly depending on what calculation or measurement you use. But on page 891 of this issue, Kokouline et al.\(^7\) report a new calculation that might reconcile these contradictory results, thanks to the inclusion of a previously overlooked molecular effect.

For such a simple polyatomic molecule, our understanding of \(\text{H}_3^+\) has a long and troubled history. It was discovered in the laboratory in 1912 by J. J. Thomson, but it wouldn’t be seen in space for many more years before it was seen in space. \(\text{H}_3^+\) is usually formed through the exothermic reaction \(\text{H} + \text{H}^+ + \text{H} \rightarrow \text{H}_3^+ + \text{H}\), so it is found wherever molecular hydrogen is present and partly ionized, such as most of the cool regions of the Universe. Being fairly reactive, \(\text{H}_3^+\) initiates most of the interstellar chemical reactions in molecular clouds. It is a generous proton donor, and by reacting with neutral molecules (\(\text{H}_3^+ + \text{X} \rightarrow \text{H}_2\text{X} + \text{H}^+\)) it triggers chains of ion–molecule reactions leading to the formation of large organic and prebiotic molecules.

So the abundance of \(\text{H}_3^+\) is a crucial parameter for models of interstellar molecule formation. Following the discovery of molecular clouds in the 1940s, astronomers searched in vain for signs of interstellar \(\text{H}_3^+\). It wasn’t until the late 1990s that T. Okada and his group successfully detected the infrared spectrum of \(\text{H}_3^+\) in dense\(^8\) and then diffuse\(^9\) interstellar clouds of molecules. This long search turned out to be the first of many hurdles in working out the abundance of interstellar \(\text{H}_3^+\).

In diffuse interstellar clouds — which contain about 100 particles per cm\(^2\) at a temperature around 100 K and are partly ionized by cosmic rays — the main process that destroys \(\text{H}_3^+\) is its dissipative recombination with low-energy electrons:

\[
\text{H}_3^+ + e^- \rightarrow \text{H}_2\text{H} + \text{H} + \text{H} + \text{H} + \text{H} + \text{H}
\]

This process has long eluded the attempts of...
energy gap, which makes the dissociative fragmentation of H$_3^+$ (formed from H$_3$) very inefficient, allowing the efficient study of other processes. Kokoouline et al. include a previously ignored decay mechanism in their new calculation, which allows the efficient study of other processes.

Kokoouline and colleagues take a decisive step towards resolving this puzzle by identifying a previously neglected decay mechanism. This mechanism also involves a coupling between electronic and nuclear motion, but requires a full three-dimensional treatment of the nuclear dynamics. The Jahn–Teller effect results from the combined vibration and rotation of the three hydrogen nuclei when distorted by the incoming electron from their equilibrium positions in the ground state of H$_3^+$. The authors show that this symmetry-breaking effect induces a cascade of energy-reducing transitions that may raise the rate of reaction (1) up to a value of $10^{3-1}$ cm$^{-3}$ s$^{-1}$, compatible both with recent astrophysical models and Langmuir-probe experiments. Moreover, the effect should increase for rotationally hot H$_3^+$ target ions, perhaps explaining the larger valuemeseasured in storage-ring experiments. The new calculation not only reproduces the tendency for interstellar H$_3^+$ to break up into three rather than two fragments, but also correctly predicts the vibrational energy of those fragments. A complete quantitative treatment is still required, but the main destruction mechanism for interstellar H$_3^+$ has now almost certainly been identified.

Annick Suzor-Weiner is at the Centre de Recherche Astrophysique de Meudon, Université Paris-Sud, Orsay, France.

Fig. 1 Bridging the gap between theory and experiment. Values for the destruction of interstellar H$_3^+$ ions by low-energy electrons (dissociative recombination) differ widely between calculations and measurements. The potential energy surfaces for the ground states of H$_3^+$ and H$_2$ molecules are separated by a large energy gap, which makes the dissociative recombination of H$_3^+$ difficult to explain by usual mechanisms. Kokoouline et al. include a previously ignored decay mechanism in their new calculation, which allows the efficient fragmentation of H$_3^+$ (formed from H$_3$ + e$^-$). (Figure reproduced from ref. 9.)

Second, using the higher destruction rate from the storage-ring data, astrophysical models cannot explain the large abundance of H$_2^+$ ions observed in diffuse interstellar clouds, which have high electron density. As a result of the rapid destruction of H$_2^+$, the measured abundance would require stable molecular clouds to be unreasonably large. At most, present astrophysical models can accept the lower rate constant of $10^{-1}$ cm$^3$ s$^{-1}$.

Third, reaction (1) has long been considered theoretically impossible for the very cold electrons in the interstellar medium. In the conventional ‘Born–Oppenheimer’ approximation of molecular physics, which treats separately the fast electronic processes and the slow dynamics of the heavy nuclei, an efficient fragmentation of the H$_3^+$ complex is expected to occur only if it has a similar electronic energy to a dissociative state of the neutral H$_3$. But the potential energy well of the H$_3^+$ ground state lies far above the repulsive potential surface of the neutral H$_3$ ground state, the only path to dissociation at very low energy (Fig. 1).

The gap between theory and experiment began to narrow when theorists realized (from studies on diatomic ions such as HeH$^+_2$) that dissociative recombination can be helped by dynamical couplings between electronic and nuclear motion, beyond the Born–Oppenheimer approximation. The reaction can then proceed in steps, through a series of H$_3^+$ bound excited states, whose potential energy surfaces mimic that of the H$_3$ ground state and span the energy gap between it and the dissociative H$_3$. A two-dimensional calculation including this mechanism led to an H$_3^+$ destruction rate close to $10^{-3}$ cm$^{-3}$ s$^{-1}$ at very low temperature. This value is acceptable for astrophysical models but is still far below the rates observed experimentally.

Kokoouline and colleagues take a decisive step towards resolving this puzzle by identifying a previously neglected decay mechanism. This mechanism also involves a coupling between electronic and nuclear motion, but requires a full three-dimensional treatment of the nuclear dynamics. The Jahn–Teller effect results from the combined vibration and rotation of the three hydrogen nuclei when distorted by the incoming electron from their equilibrium positions in the ground state of H$_3^+$. The authors show that this symmetry-breaking effect induces a cascade of energy-reducing transitions that may raise the rate of reaction (1) up to a value of $10^{-3}$ cm$^{-3}$ s$^{-1}$, compatible both with recent astrophysical models and Langmuir-probe experiments. Moreover, the effect should increase for rotationally hot H$_3^+$ target ions, perhaps explaining the larger valuemeseasured in storage-ring experiments. The new calculation not only reproduces the tendency for interstellar H$_3^+$ to break up into three rather than two fragments, but also correctly predicts the vibrational energy of those fragments. A complete quantitative treatment is still required, but the main destruction mechanism for interstellar H$_3^+$ has now almost certainly been identified.

Annick Suzor-Weiner is at the Centre de Recherche Astrophysique de Meudon, Université Paris-Sud, Orsay, France. E-mail: annick.suzor-weiener@ppm.u-psud.fr

Ivan F. Schneider is in the Department of Physics, Université de la Havre, France.

Mother’s voice recognition by seal pups

Newborns need to learn their mother’s call before she can take off on a fishing trip.

In gregarious mammals, mother and pup need to be able to recognize each other’s voices in order to be reunited in a crowd, a skill that has only been studied in domesticated species where mother and young stay together during the rearing period. In otariids, females have to leave their newborn to feed at sea, but offspring nevertheless develop long-term recognition of their mother’s voice. Here we show that pups of the subantarctic fur seal (Arctocephalus tropicalis) can acquire this ability when they are just 2–5 days old, and that the mother times her departure accordingly.

On Amsterdam Island in the Indian Ocean, fur seals are born from late November to early January. The rearing period lasts for about 10 months and consists of foraging trips at sea that last for 2–3 weeks at a time, interspersed with suckling periods ashore lasting for 2–3 weeks at a time, interspersed with periods of suckling, and lasting for an average of 10 months.

The rhamphotheca extends about 1 mm farther to the right naris. The upper and lower rhamphotheca on each jaw taper to the tips of the mouth. Although the premaxilla and dentary are present, specimens RTMP 95.110.1 from Dinosaur Provincial Park, Alberta, Canada (Fig. 1a). This specimen preserves traces of the keratinous covering (rhamphotheca) on the beak on the tips of the right premaxilla and dentary. As preserved, the soft tissue extends up to 3.5 mm from the rostral and mandibular margins, and fills a gap left by the anterior divergence of the premaxilla and dentary. The rhamphotheca on each jaw tapers posteriorly and terminates just behind the midpoint of the right naris. The upper rhamphotheca extends about 1 mm farther than the lower rhamphotheca.

The beaks of ostrich dinosaurs

Primitive ornithomimids, a ubiquitous group of specialized Cretaceous dinosaurs nested within a clade of predominantly carnivorous theropods, are known to have had teeth, whereas derived ornithomimids had an edentulous beak, which has prompted speculation about their dietary habits. Here we describe two new ornithomimid specimens in which soft-tissue structures of the beak have been preserved. These creatures probably used their beaks to strain food sediment in an aqueous environment, rather than for predation on large animals.

Specimen RTMP (Royal Tyrrell Museum of Palaeontology) 95.110.1 is an almost complete skeleton of the Campanian ornithomimid Ornithomimus edmontonicus from Dinosaur Provincial Park, Alberta, Canada (Fig. 1a). This specimen preserves traces of the keratinous covering (rhamphotheca) on the beak on the tips of the right premaxilla and dentary. As preserved, the soft tissue extends up to 3.5 mm from the rostral and mandibular margins, and fills a gap left by the anterior divergence of the premaxilla and dentary. The rhamphotheca on each jaw tapers posteriorly and terminates just behind the midpoint of the right naris. The upper rhamphotheca extends about 1 mm farther than the lower rhamphotheca.

To test the efficiency of the pups’ learning, we measured the time taken by pups to find their mother after her return from her first sea trip. Of the monitored mother–pup pairs (n = 15), 66% met up within 7 min, and the remaining pairs were united within 11 min (regression analysis, double-reciprocal model: percentage of successful meetings calculated as 1/(−0.000356 + 6.45/time required); P < 0.05).

In colonial mammals, the ability of pups to recognize their mother’s voice is crucial for their survival as, in most species, females only feed their own offspring. Mother–young recognition must therefore develop as soon as it is needed, before the first separation. In several bird species, vocal identification between parents and young generally coincides with offspring mobility. In A. tropicalis, factors other than call recognition may be involved — for instance, females may leave the shore when pups stop suckling.

Our results show that, in spite of the variable duration of ontogeny, the mother’s departure date is linked to the pup’s ability to recognize her voice. This supports the idea that recognition of a mother’s call by her pup is an important factor in allowing her to go to sea.

Isabelle Charrier*, Nicolas Mathevon*, Pierre Jouventint
*Laboratoire de Biologie Animale, Université Jean Monnet, 42023 Saint-Etienne cedex 2, France
e-mail: mathevon@univ-st-etienne.fr
†CEFE CNRS UPR 9056, 34213 Montpellier, France
‡SNAM CNRS UMR 8620, 91400 Orsay, France

posteriorly than the lower, and is also slightly longer anteriorly. The surfaces of the dentary and the premaxilla are densely pitted next to the preserved soft tissue, indicating extensive neurovascularization in the region of the rhamphotheca.

The other specimen, IGM (Institute of Geology, Mongolia) 100/1133, is a sub-adult of Gallimimus bullatus1 from Tsaagan Khushu, Nemegt Basin, Mongolia (Fig. 1b). Although slightly crushed, the skull is remarkably complete. The interior surface of the rhamphotheca is preserved on both the right premaxilla and dentary. Small, separate columnar structures extend sub-perpendicularly to the buccal surface of the rhamphotheca. These structures, which are remnants of the lingual surface of the beak, project up to 5.6 mm, as preserved, at the tip of the premaxilla, but are shorter posteriorly. They are tightly packed and evenly spaced (about 0.5 mm from centre to centre). Presumably, these columnar structures would have been at least partially covered laterally by an outer columnar structures would have been at least partially covered laterally by an outer columnar structures would have been at least partially covered laterally by an outer

In an analysis of 17 diverse modern anseriform species, some straining was observed in all species except the piscivorous merganser5, which has large, tooth-like lamellae. The Gallimimus condition closely resembles that of the shoveller Anas clypeata, an extreme filter-feeding strainer with a diet of plants, freshwater molluscs, ostracods and foraminiferans5. Other features of the ornithomimid skeleton have been used as grounds to suggest that they were not carnivorous±

Ornithomimids with numerous small gastroliths have been reported from deposits in China1. Although gastroliths are patchily distributed in Dinosauria, they are most often associated with herbivorous (Psittacosaurus mongoliensis and sauropods) or putatively herbivorous (Caudipteryx zhouri) groups. Although there is not enough evidence to indicate that these ornithomimids behaved similarly to anseriforms, it is unlikely that these delicate features were used for predation on large animals. Furthermore, ornithomimids are abundant in mesic environments (such as Nemegt8, Iren Dabasu9 and Dinosaur Park Formations10), but are rarer in more arid environments (such as Djadokhta8,11). This suggests that ornithomimids may be ecologically tied to food supplies in wetter, more mesic environments, which would be consistent with their beak morphology.

Mark A. Norell*, Peter J. Makovicky†, Philip J. Currie

*Canadian Museum of Natural History, Ottawa, Ontario K1A 0R7, Canada
†American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024-5192, USA
e-mail: norell@amnh.org
‡Field Museum, 1400 South Lake Shore Drive, Chicago, Illinois 60605-2496, USA
§Royal Tyrell Museum of Palaeontology, Drumheller, Alberta T0J 0Y0, Canada


Figure 1 Two ornithomimid specimens, showing preservation of soft-tissue structures of the beak. A, The anterior end of the rhamphotheca of Ornithomimus edmontonicus (RTMP 95.110.1) in lateral view (reversed). Red arrow indicates preserved soft tissue. Scale bar, 1 cm. B, Gallimimus bullatus (IGM 100/1133) in right lateral view. Scale bar, 2 cm. C, Detail of red box in B, showing soft-tissue lamellae along the rostral margin. Scale bar, 3 mm.

Figure 2 Artist’s rendition of a feeding ornithomimid, showing the sediment-straining method probably used by these dinosaurs.
Self-assembled domain patterns

The ordered domain patterns that form spontaneously in a wide variety of chemical and physical systems as a result of competing interatomic interactions can be used as templates for fabricating nanostructures. Here we describe a new self-assembling domain pattern on a solid surface that involves two surface structures of lead on copper. The evolution of the system agrees with theoretical predictions, enabling us to probe the interatomic force parameters that are crucial to the process.

The two structures showing this remarkable behaviour are a disordered Pb/Cu surface alloy, produced by deposition of lead vapour onto clean Cu(111), and a lead-overlayer phase, which is produced by continued deposition after the alloy structure has covered the surface. Figure 1a-g shows a sequence of low-energy electron microscope (LEEM) images in which the overlayer structure (bright) grows over the surface alloy structure (dark) during lead deposition at a temperature of 673 K. The most striking feature of this sequence is the evolution of a pattern from circular islands (average diameter, 67 nm) to stripes and then to circular holes within the lead-overlayer matrix.

This sequence of domain patterns, referred to as droplets, stripes and inverted droplets, respectively, has been predicted to be a general property of domain formation on solid surfaces. To our knowledge, these results represent the first unambiguous observation of the expected sequence of domain patterns with changing area fraction during deposition, and agree with the predicted positions of the droplet-to-stripe phase transitions for short-range attractive and long-range dipolar interatomic interactions on solid surfaces.

One reason why these domain patterns form in this system is the high mobility of the islands (Fig. 1a), which contain many thousands of atoms and can move by as much as a micrometre before being incorporated into the more tightly packed droplet structure (Fig. 1b). This mobility allows patterns to form that have hundred-nanometre periodicity and allows the underlying interatomic interactions to be determined.

As the islands move in the force fields of the other islands, the island trajectories contain information about the force fields. Droplet trajectories are consistent with the expected long-range dipolar repulsion. Because magnetic interactions are not relevant in this system and we can rule out electrostatic interactions from measurements of work-function differences, we propose that elastic interactions arising from a surface-stress difference between the alloy and overlayer structures are responsible for the stabilization of domain patterns.

The behaviour of the patterns at fixed coverage and varying temperature is also qualitatively consistent with theory. For both droplets and stripes, the feature size decreases with increasing temperature. The most likely reason for this is a reduction in the domain-boundary free energy with increasing temperature. For droplets and inverted droplets, reducing the temperature allows improvement in the long-range order. Figure 1h is a LEEM image showing the degree of long-range order that can be obtained in the inverted-droplet phase.

If self-assembled structures are to be useful as nanostructure templates, two-dimensional, self-assembled patterns need to be stable at room temperature and resistant to air exposure. This is the case for our Pb/Cu system. Features that correspond to lead-overlayer droplets can be clearly seen in atomic-force-microscopy images taken after cooling to room temperature and 2 hours exposure to air (Fig. 1j).

Richard Plass*, Julie A. Last*, N. C. Bartelt*, G. L. Kellogg*
Sandia National Laboratories, Albuquerque, New Mexico 87185-1415, USA
e-mail: gkello@sandia.gov
†Sandia National Laboratories, Livermore, California 94551-0969, USA


Malaria

Cooperative silencing elements in var genes

Each Plasmodium falciparum malaria parasite carries about 50 var genes from a diverse family that encode variable adhesion proteins on the infected red blood cells of the host, but individual parasites single out just one var gene for expression and silence all the others. Here we show that this silencing is established during the DNA-synthesis phase (S phase) of the cell cycle and that it depends on the...
The proteins (PFEMP1) encoded by the var genes enable red blood cells infected with P. falciparum to adhere within the circulatory system, thereby avoiding destruction by the spleen. Different PFEMP1 forms can attach to different host molecules, leading to patterns of sequestration that produce disease complications such as cerebral malaria or the severe malaria of pregnancy\(^\text{1}\). Because the exposed PFEMP1 molecules are subject to antibody attack during infection, P. falciparum parasites produce subpopulations with antigenically diverse PFEMP1 forms which must be continually chased by the immune response. The underlying switches in var-gene expression that result in this process of antigenic variation have yet to be understood.

Silent var promoters become transcriptionally active when removed from their chromosomal context\(^\text{2}\), suggesting that further control elements in the intact var gene are required to control or silence expression. Possible locations for such control elements are in the conserved var introns, which separate the two exons of all var genes and have sequences that are found in many heterogeneous sterile transcripts in erythrocyte-stage parasites\(^\text{3}\).

To test whether var introns function in gene control, we inserted a representative intron sequence (int) into two different plasmids containing a luciferase (luc) reporter driven either by a var promoter sequence or by a promoter sequence from the unrelated gene hrp3, which encodes histidine-rich protein 3 (pVLH and pHLH, respectively; Fig. 1a). Transfection experiments with pVLH and pVLH/int indicated that the presence of int downstream of luc strongly repressed the luciferase signal (Fig. 1b, left), irrespective of int orientation. In contrast, the presence of int in the pHLH/int plasmid caused no detectable change in the reporter activity (Fig. 1b, right). The fact that the var, but not the hrp3, promoter was affected indicates that int repression is specific and occurs by cooperative action with the upstream var 5′-flanking region.

These initial experiments revealed strongly repression, but not complete silencing, of luc expression from pVLH/int (Fig. 1b). Transition through S phase of the cell cycle is known to be required for the assembly of silent chromatin structures in other organisms\(^\text{4}\). Because parasites continually take up DNA during their growth in plasmid-loaded erythrocytes\(^\text{5}\), our experiments may have included heterogeneous populations of plasmids acquired before and after S phase, leading to incomplete silencing of the var promoter.

To test this possibility, we measured luc expression from homogenous plasmid populations before and after transition through S phase. Assays done 18 h after invasion of plasmid-loaded erythrocytes (before the onset of S phase) confirmed that luc was expressed without repression (Fig. 1c, left). After their maturation to mature schizont stages, we separated transfected parasites from plasmid-loaded erythrocytes using Percoll/sorbitol gradients. Returning these parasites to fresh cultures with plasmid-free erythrocytes eliminated further plasmid uptake after re-invasion and ensured that all plasmid DNA in the parasites had been through S phase. Complete reporter silencing from pVLH/int was then evident (Fig. 1c, right). We detected no silencing in control transfections with the original pVLH, or with a pVLH/hsp86 construct in which the int sequence was replaced with a P. falciparum hsp86 intron of similar size (Fig. 1c, right).

The regulatory elements and S-phase-dependent silencing we identify here implicate cooperative DNA-binding complexes and modifications of chromatin structure in the control of var-gene transcription. Experimentally manipulable plasmid systems that incorporate mechanisms for var control may offer a useful approach by which to identify and characterize these complexes.

**Kirk W. Deitsch, Michael S. Calderwood, Thomas E. Wellems**

Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892-0425, USA

e-mail: twd@helix.nih.gov

---

**Figure 1** The role of cooperative control elements in S-phase-dependent silencing of var expression. **a**, Maps of pVLH/int(+) and pHLH/int(+). Plasmids pVLH/int(+), pVLH/int(−), pHLH/int(+) and pHLH/int(−) were constructed from the pVLH (pVLH-1) and pHLH (pHLH-1) plasmids by inserting a var intron (int) in the (+) or (−) orientation downstream of the 5′-flanking region of the hrp2 gene. Plasmid pVLH/hsp was obtained by inserting the intron sequence from the unrelated P. falciparum hsp86 gene into pVLH at the same position. GenBank accession numbers for plasmids pVLH(+), pVLH/hsp and pHLH(−) sequences are AF338324, AF378054 and AF338825, respectively. The transcription start point (tsp) for var is indicated. **b**, Luciferase activity from P. falciparum parasites after spontaneous transformation with original and modified pVLH and pHLH plasmids. Assays were carried out after cultivation of parasites for 72 h in the presence of plasmid-loaded erythrocytes\(^\text{6}\). Luciferase activity from homogenous plasmid populations before and after S phase in synchronized transfected parasites. Experiments were carried out in triplicate and repeated on at least three separate occasions. Error bars indicate calculated standard deviations.
brief communications

Animal behaviour

An unusual social display by gorillas

W

e have observed wild western low-

land gorillas (Gorilla gorilla gorilla) using water to generate spectacular ‘splash displays’. Most of these displays were

made by silverbacks in an agonistic context, and we propose that they are primarily linked to the intimidation of potential rivals for female acquisition. This unusual behaviour may have developed only in gorillas for female acquisition. This unusual behaviour may have developed only in gorillas for female acquisition.

W

e have observed wild western low-

land gorillas (Gorilla gorilla gorilla) using water to generate spectacular ‘splash displays’. Most of these displays were made by silverbacks in an agonistic context, and we propose that they are primarily linked to the intimidation of potential rivals for female acquisition. This unusual behaviour may have developed only in gorillas for female acquisition.

Almost nothing is known about the social behaviour of western lowland gorillas because of poor observation conditions and difficulties of habituation in dense forest. But the discovery that large numbers of gorillas feed in open, swampy clearings (baïs) in the forest of northern Congo means that their social interaction can be investigated.

Display behaviour incorporating objects in the environment occurs in agonistic encounters in all ape species, but manipulating water for communication has not been described in any wild primate and, with the possible exception of elephants, may be unique among terrestrial mammals.

We observed 124 gorillas over 32 months at the 12.9-hectare ‘Mbeli Bai’. Gorillas were visible for 27% (1,681 h) of the time that we were present at the clearing. Ninety splash displays, representing 57 independent bouts of social interaction, were produced by 19 individuals (4 unaffiliated, ‘solitary’ silverback males and 15 individuals from 9 groups). Ten display styles were seen, of which three were used in 75% of all displays. These were the ‘body splash’ (35% of displays), in which a gorilla runs or leaps into standing water of up to 1.5 m in depth (Fig. 1), and one-handed and two-handed splashes (40%), in which one or both arms are raised and then brought down forcibly, the open palms striking the water surface at a slight angle. Each of these three techniques generates large plumes of spray.

Of all splash-display bouts, 67% were produced in an agonistic context, 17.5% were made in play and 5% were directed towards other species; in 10.5% of cases the context was not evident. Silverbacks displayed the most frequently (68%), and almost exclusively in an agonistic context.

When the observed frequencies of splash bouts were compared with the expected frequencies (calculated from the proportion of visits made to the bai by each age/sex class), group silverbacks displayed twice as often as expected, whereas solitary silverbacks displayed more than four times as often. Solitary silverbacks were also the most frequent recipients of the display (six times more often than expected). Adult females, although they were the most frequently seen age/sex class, never produced splash displays and were rarely targeted.

Directly attracting the attention of females is not considered the prime purpose of splash displays, because solitary males displayed almost as frequently to other solitary males as to groups, and in more than half of these cases no females were in sight. The more likely purpose is to intimidate potential competitors for acquiring females.

Splash displays are an example of object-mediated behavioural plasticity in response to unusual circumstances. Although the three primary display styles resemble the dry-land charges and ground-slap displays seen in many gorilla populations, differences in context, execution and intra-dyad distance confirm that splash displays represent a distinct behavioural element in gorillas’ visual-display repertoire.

The bai offers visibility of up to 500 m, which is never encountered by gorillas in the forest, and so long-distance visual displays are clearly expedient. In the wild, the only other record of splash display comes from western lowland gorillas at M a ya Bai, 180 km from M beli (F. Magliocca, personal communication).

The paucity of data on western lowland gorillas has led to generalizations about their behaviour based on that of mountain gorillas (G. beringei beringei). But their feeding ecology is different, and our findings indicate that their social behaviour is too. We anticipate that gorillas, maligned as cognitively poor cousins to the other great apes, will emerge from further bai studies as adaptable, innovative and intelligent creatures that exploit a complex environment.

Richard J. Parnell†, Hannah M. Buchanan-Smith* (eds Russon, A. E., Bard, K. A. & Taylor Parker, S.)

Apes


Figure 1 A solitary silverback performing a ‘body splash’ display at Mbeli Bai, Nouabalé-Ndoki National Park, Congo (Brazzaville).
Identification of an angiogenic mitogen selective for endocrine gland endothelium

Jennifer LeCouter†, Joe Kowalski†, Jessica Foster‡, Phil Hass‡, Zemin Zhang†, Lisa Dillard-Telm§, Gretchen Frantz§, Linda Rangell¶, Leo DeGuzman*, Gilbert-Andre Keller§, Franklin Peale§, Austin Gurney†, Kenneth J. Hillan§ & Napoleone Ferrara*

Departments of † Molecular Oncology, ‡ Molecular Biology, § Protein Chemistry, ¶ Pathology and † Toxicology/Pathology, Genentech Inc., South San Francisco, California 94080, USA

The known endothelial mitogens stimulate growth of vascular endothelial cells without regard to their tissue of origin. Here we report a growth factor that is expressed largely in one type of tissue and acts selectively on one type of endothelium. This molecule, called endocrine-gland-derived vascular endothelial growth factor (EG-VEGF), induced proliferation, migration and fenestration (the formation of membrane discontinuities) in capillary endothelial cells derived from endocrine glands. However, EG-VEGF had little or no effect on a variety of other endothelial and non-endothelial cell types tested. Similar to VEGF, EG-VEGF possesses a HIF-1 binding site, and its expression is induced by hypoxia. Both EG-VEGF and VEGF resulted in extensive angiogenesis and cyst formation when delivered in the ovary. However, unlike VEGF, EG-VEGF failed to promote angiogenesis in the cornea or skeletal muscle. Expression of human EG-VEGF messenger RNA is restricted to the steroidogenic glands, ovary, testis, adrenal and placenta and is often complementary to the expression of VEGF, suggesting that these molecules function in a coordinated manner. EG-VEGF is an example of a class of highly specific mitogens that act to regulate proliferation and differentiation of the vascular endothelium in a tissue-specific manner.

Embryonic development, reproductive functions, tumorigenesis and many other proliferative processes are critically dependent on the development of a vascular supply. Over the last decade, much progress has been made in the identification of the regulators of blood vessel growth. Most notably, VEGF and the Tie2 ligands are widely distributed and have been shown to play a coordinated role in endothelial cell proliferation and in the assembly of the vessel wall in a variety of normal and abnormal circumstances.

Evidence also exists for local, tissue-specific regulation of endothelial cell growth and differentiation. Endothelial cells of different organs exhibit a striking degree of morphological and functional diversity to adapt to local needs. Whereas endothelial lineage is genetically determined, expression of a set of vascular-specific genes is regulated at the transcriptional level after interaction with local signalling pathways. The nature of the local regulatory signals is virtually unknown, although the specialized stroma is believed to be a major source. Conceivably, an integrated network of stimuli, which may include tissue-specific secreted proteins in addition to cellular and extracellular matrix components, functions to determine the structure and function as well as modulate growth of the resident endothelium.

The microvascular endothelium of endocrine glands is highly permeable to blood-borne substances and to specific secretory products, reflecting the need for rapid hormone release in responses to homeostatic changes. A major aspect of the high permeability in endocrine gland endothelium, as well as in some other organs, is the specialized discontinuities of the plasma membrane called fenestrae. On the basis that VEGF mRNA is highly expressed around fenestrated microvessels, it has been proposed that VEGF is uniquely involved in the induction and maintenance of such structures. However, expression of VEGF mRNA is also high in areas where capillaries are continuous (for example, myocardium or lung alveoli) and therefore it is unlikely that VEGF is the only determinant of a fenestrated phenotype.

Here we report on the identification of an angiogenic mitogen selective for one endothelial cell type, endocrine gland endothelium. Remarkably, the expression of this factor is largely restricted to steroidogenic glands. Although this protein does not show any structural homology to the VEGF family, it displays several striking biological similarities to VEGF, including hypoxic regulation and the ability to induce fenestration in the target cells. On the basis of such functional homologies, we propose the designation of endocrine-gland-derived vascular endothelial growth factor (EG-VEGF) for this factor.

Identification and structure of EG-VEGF

A library of purified human secreted proteins was screened for the ability to induce proliferation in primary bovine adrenal-cortex-derived capillary endothelial (ACE) cells. These cells respond only to a very restricted number of mitogens, including basic fibroblast growth factor (bFGF) and VEGF. We found that one factor, referred to herein as EG-VEGF, was capable of inducing a strong and reproducible mitogenic response. Mature EG-VEGF is a protein with a relative molecular mass of 8,600 (M, 8.6K) encoded by a complementary DNA cloned from a human ovary library. The 1.4-kilobase (kb) cDNA encodes a protein of 105 amino acids, with a well defined signal sequence. The mature protein is predicted to have 86 amino acids, including 10 cysteines, and an expected isoelectric point (pI) of 8.46 (Fig. 1a). EG-VEGF displays a high degree of homology (80%) to a non-toxic protein purified from the venom of the black mamba snake (Dendroaspis polylepis polylepis), venom protein A (VPRA) (Fig. 1b). The structure of the native VPRA was solved and the disulphide bridge partners were revealed on the basis of a floating assignment by nuclear magnetic resonance analysis. Notably, the number and spacing of cysteines for EG-VEGF are completely conserved. Thus, EG-VEGF is the human orthologue or a closely related homologue of VPRA. A human molecule closely related to a peptide isolated from the yellow-bellied toad (Bombina variegata), Bv8 (ref. 15), is 58% identical to the EG-VEGF mature protein and thus belongs to the same gene family. EG-VEGF displays...
EC-VEGF is mitogenic and chemotactic

EC-VEGF is mitogenic and chemotactic

EG-VEGF stimulated proliferation of ACE cells with a ED₃₀ (effective dose for a 50% response) of 0.2 nM and a maximal effect at ~2 nM (Fig. 2a). The fold increase in cell number was very similar to that induced by VEGF. The bacakulosis-produced His- and Fc-tagged EC-VEGF proteins behaved very similarly in all experiments. Consistent with a mitogenic function, EC-VEGF induces in ACE cells a rapid and significant phosphorylation of the MAP kinases ERK1 and ERK2, as well as of other proliferation and survival signalling molecules (manuscript in preparation). The mitogenic activity of EC-VEGF was not blocked by administration of a soluble VEGF receptor (mFDr-immunoglobulin-gamma (IgG)) tested at 50–100 ng ml⁻¹, indicating that such an effect is not mediated by release of these findings however do not rule out the possibility of reciprocal inductive effects between the two factors in more complex systems in vivo.

To determine the spectrum of cell types that proliferate in response to EC-VEGF, we tested a variety of other endothelial and non-endothelial cell types, including human umbilical vein (HUVFEC), human dermal microvascular (HMDVEC), bovine brain capillary (BBBC) and adult bovine aortic (ABAE) endothelial cells. BC cells showed about 10% of the proliferative response to VEGF, and the other endothelial cell types failed to show any effect after EC-VEGF administration, at all concentrations tested (Fig. 2b). EC-VEGF did not elicit any proliferative response in cultures of vascular smooth muscle cells (VSMCs), pericytes, fibroblasts or keratinocytes (Fig. 2c). Thus, EC-VEGF is a mitogen not only specific to endothelial cells, but also acting selectively on a defined type of endothelial cell. In contrast, other endothelial cell mitogens, such as VEGF and bFGF, do not show any significant selectivity for various endothelial cell types.

An essential aspect of the angiogenesis cascade is chemotaxis of VEGF and bFGF are able to act as chemotaxants and stimulate endothelial cell migration. To further define the biological activity of EC-VEGF, we assessed whether ACE cells display a chemotactic response to this molecule in a modified Boyden chamber assay. A strong and reproducible chemotactic response was elicited in ACE cultures, with a peak response at 0.5 nM EC-VEGF (Fig. 2d). The magnitude of the effect was in the same range as that induced by VEGF.

To extend our results to another primary cell type, we purified endothelial cells from baboon adrenal cortex (BAEC). In BAEC cultures, 0.5–5 nM EC-VEGF induced a chemotactic response of the same magnitude as that induced by VEGF. Furthermore, EC-VEGF induced migration in MS-1 endothelial cells to an even greater extent than VEGF. The MS-1 cell line, isolated from microvessels of murine endocrine pancreas, retains highly differentiated properties, such as VEGF receptor expression. However, no response was elicited in HUVFECs, even though these cells displayed a strong response to VEGF. Therefore, in addition to its mitogenic activity, EC-VEGF also acts as a chemotaxant, but only for a specific endothelial cell type. In agreement with these findings, binding studies with iodinated EC-VEGF revealed high-affinity, specific binding in ACE cells, which could be crosslinked to a membrane protein. No high-affinity binding was however, detected in non-responsive cells such as HUVFECs (manuscript in preparation).

Endothelial cells within the endocrine glands display a rather unique fenestrated architecture that is found in other restricted sites, such as kidney glomeruli, choroid plexus, median eminence, liver and many tumours. Fenestrae are specialized plasma membrane microdomains, or windows, about 60 nm in diameter, that are usually clustered. They are highly permeable to fluid and small solutes and are thought to facilitate large exchange of materials between interstitial fluid and plasma, such as that occurring in steroid-producing endocrine glands and others such as pancreatic
islets. We tested whether EG-VEGF could induce fenestration in endothelial cells, alone or in combination with VEGF. Notably, so far only VEGF has been reported to induce fenestration, in vivo and in vitro.

ACE cells were grown to confluence on extracellular matrix produced by bovine corneal endothelial cells and then treated with growth factors for 24 h. In agreement with previous studies, VEGF induced fenestration in 4.33 ± 1.53% (mean ± s.d) of ACE cell profiles (Fig. 2f). EG-VEGF had a very similar effect, inducing fenestration in 5 ± 1% of cell profiles (Fig. 2g). A combination of the two factors produced an additive or cooperative response, inducing fenestration in 11 ± 1% of the ACE cell profiles. No fenestrae were observed in the absence of VEGF or EG-VEGF, or in the presence of EGF. Very similar results were obtained with the islet-microvessel-derived MS-1 cell line (data not shown). These findings support the hypothesis that VEGF and EG-VEGF may cooperate in vivo, in settings such as the adrenal cortex or the ovary to induce the fenestrated phenotype and regulate permeability of the resident endothelial cells. It remains to be established whether EG-VEGF has any effects on fenestrated endothelia outside endocrine glands.

**Figure 2** EG-VEGF is a mitogen, chemoattractant, and inducer of fenestration for specific endothelial cells. In ACE cell cultures, EG-VEGF induced a maximal mitogenic response at 2 nM, with an ED₅₀ of 0.2 nM. C, control. B, c, proliferation assays with several endothelial (b, HUVECs, HMVECs, BBC cells, ABAE cells) and non-endothelial cell types (c, VSMCs, pericytes and fibroblasts (BHK21 and human neonatal fibroblasts, hFb) and keratinocytes). Basal media served as a negative control (C); bFGF (f), EGF (e) or VEGF (m), added respectively at 5, 5 and 10 ng ml⁻¹, served as positive controls. EG-VEGF was tested at 10 and 100 nM. d, EG-VEGF induced a chemotactic response in ACE cells, BAECs and MS-1 cells, but not in HUVECs. Each graph is a representative experiment. Data are mean ± s.d., and the proliferation or migration index is relative to the negative control, arbitrarily set to 1. e–g, EG-VEGF induces fenestration in adrenal cortex-derived capillary endothelial cells. ACE cells, grown to confluence on ECM, were treated with 2.5 nM VEGF, 10 nM EG-VEGF or both. Electron micrographs of ACE cells untreated (e), treated with VEGF (f) or EG-VEGF (g) revealed that both molecules are capable of inducing fenestration. Arrows indicate the location of fenestrae. Magnification is indicated (in μm for e and f, nm for g).

**EG-VEGF expression is induced by hypoxia**

Hypoxia is a key inducer of angiogenesis in both physiological and pathological conditions. By virtue of activating hypoxia-inducible factor-1 (HIF-1), a dimeric transcription factor composed of HIF-1α and HIF-1β subunits, low oxygen tension is known to induce expression of VEGF and Flt-1 (VEGFR-1), in addition to erythropoietin (EPO) and certain glycolytic enzymes, through cis-acting regulatory elements. We sought therefore to determine whether hypoxia regulates expression of EG-VEGF mRNA. Exposure of the human adrenal carcinoma cell lines SW13 and H295R to hypoxic conditions (~2% oxygen) resulted in a 2.75 ± 1.5% and 2.10 ± 1.2% increase in EG-VEGF mRNA levels above normal, respectively, whereas the VEGF mRNA increased 352 ± 30% and 266 ± 15%, respectively (Fig. 3a). Similar results were obtained in two additional independent experiments.

A database search of the EG-VEGF promoter sequence for the core HIF-1 binding site revealed a putative element within the first 2,450 nucleotides upstream of the transcription start site (Fig. 3b), based on a consensus sequence (TAAGTGCAGGC, bold represents the invariant sequence). After an 18–24-h incubation, a luciferase reporter construct containing the putative EG-VEGF element conferred a 3.3 ± 0.8-fold increase above normal oxygen conditions. This level is comparable to that conferred by the hypoxia-response element (HRE) Epo consensus, 3.4 ± 1.2-fold (Fig. 3c). Mutating the core sequence of either the consensus or the putative EG-VEGF HRE abolished the response to hypoxia, verifying the specificity of

The promoter region of the human EG-VEGF gene contains a hypoxia-responsive element (HRE) that is responsible for the induction of EG-VEGF expression under hypoxic conditions. This element is located within the first 2,450 nucleotides upstream of the transcription start site. When cells are exposed to hypoxia, the transcription factor HIF-1α is stabilized, leading to the induction of the HRE and subsequent upregulation of EG-VEGF expression. This finding supports the hypothesis that EG-VEGF plays a role in angiogenic processes under hypoxic conditions, particularly in tissues such as the adrenal cortex and ovary.
the response. Although we cannot rule out additional mechanisms, pending a detailed promoter analysis these findings indicate that in all likelihood HIF-1 is a mediator of the hypoxic regulation of the EG-VEGF gene. It is striking that, in spite of the lack of homology, EG-VEGF and VEGF not only have similar effects in responsive endothelial cells but may also be regulated through common mechanisms.

**EG-VEGF is restricted to steroidogenic tissues**

To elucidate the expression pattern of EG-VEGF, we performed dot blot analysis on tissue RNA arrays representing an extensive series of human tissues and cell lines. This evidenced expression only in ovary, testis, adrenal, placenta and, to a lesser extent, prostate (see Supplementary Information). To verify this finding, we performed northern blot analysis using RNA from a variety of human tissues. A single mRNA species of 1.4 kb was expressed, in decreasing order of intensity, in ovary, testis, adrenal and placenta, in agreement with the dot blot data. No expression was detected in any other tissue examined, with the exception of a very weak signal in the prostate (Fig. 4). Similar findings were obtained by in situ hybridization in human multi-tissue arrays (data not shown). These findings indicate that steroidogenic endocrine glands are the major site of expression of EG-VEGF mRNA. A much weaker expression occurs in a non-steroidogenic organ like the prostate.

To identify the cell types that express EG-VEGF, we examined a series of ovary and testis specimens from humans and other primates. In the testis, we detected a strong hybridization signal, which was virtually restricted to the Leydig cells, the testosterone-producing cells (Fig. 5a–d). It is noteworthy that the endothelium of the testis has a surprisingly high turnover: as many as 3% of endothelial cells are labelled with bromodeoxyuridine (BrDU), which is probably related to the intense metabolic activity associated with spermatogenesis and steroidogenesis. In this context, Leydig cells are known to be a source of angiogenic and permeability-enhancing factors such as VEGF. Notably, in our specimens the VEGF hybridization signal was considerably less intense than the EG-VEGF signal (data not shown). In the ovary, intense EG-VEGF mRNA expression was localized to the cortical stroma, the cumulus oophorus, theca, and granulosa of developing follicles (Fig. 5e–p). The strong expression of EG-VEGF mRNA within the theca is coincident with the development of a capillary network associated with follicular development and steroid hormone production, and is consistent with a pro-angiogenic role of EG-VEGF.

This expression pattern is essentially similar in the other primate species examined, the cynomolgus monkey (Macaca fascicularis) and the chimpanzee (Pan troglodytes) (Fig. 5g–j and data not shown). Whereas VEGF has a low expression in the specialized stroma, EG-VEGF is strongly expressed (Fig. 5e, f, k, l). Interestingly, such stroma becomes intensely hyperplastic and angiogenic, producing excessive amounts of androgens, in the polycystic ovary syndrome, and we observed strong expression of EG-VEGF mRNA in the hyperplastic stroma in a series of human polycystic ovary specimens (unpublished observations). The expression pattern of EG-VEGF in the cumulus oophorus is very similar to that of VEGF (Fig. 5m–p). EG-VEGF expression is not detected in the corpus luteum (data not shown). Whereas EG-VEGF is highly expressed in the ovarian stroma and within developing follicles, VEGF is expressed at the highest level in the corpus luteum (Fig. 5e–p and data not shown).

**EG-VEGF induces angiogenesis and ovarian cyst formation**

We examined the in vivo activity of EG-VEGF in a variety of model systems. In the rat corneal pocket assay, purified EG-VEGF failed to show a significant response in the eyes tested, whereas VEGF induced the expected angiogenic effects (Fig. 6a–c). Similar results were obtained in the rabbit cornea (data not shown). To achieve EG-VEGF delivery in a local and sustained fashion, we generated adenovirus vectors and injected these into various sites in athymic nude rats or mice. Recombinant adenovirus expressing IacZ or VEGF served as controls. When injected into the skeletal muscle of nude rats, the effects of IacZ and EG-VEGF adenovirus were essentially indistinguishable, with minimal inflammatory infiltrate and absence of angiogenesis (Fig. 6d–f). In contrast, the VEGF-expressing adenovirus induced frank angiogenesis, with mitotic endothelial cells, oedema and granulation tissue, in agreement with previous studies. However, the EG-VEGF protein was produced at high level in the skeletal muscle, arguing against the possibility that lack of EG-VEGF synthesis may explain the absence of angiogenic effects in this site (see Supplementary Information). Very similar results were obtained after injection of the same adenovirus vectors in the mouse ear (Fig. 6g, h) or skeletal muscle (data not shown).
In contrast, intra-ovarian delivery of either EG-VEGF or VEGF adenovirus resulted within a week in a dramatic enlargement of the injected ovaries, with grossly visible blood vessels as well as haemorrhagic areas (Fig. 6i). The ovarian wet masses in a representative experiment in nude rats were 33 ± 3 mg in the lacZ group, 489 ± 30 mg in the EG-VEGF group and 191 ± 91 mg in the VEGF group (n = 4). The mass of uninjected ovaries (37 ± 7 mg) was not different from the lacZ group. Although there was a trend towards larger mass in the EG-VEGF group compared with the VEGF group, the histological picture was very similar. There was, in both cases, ovarian enlargement with disruption of the normal ovarian stroma by large cystic spaces containing proteinaceous material and frank haemorrhage (Fig. 6k, l). High-power examination revealed intense angiogenesis at the periphery of the cysts (Fig. 6m, n). The picture is consistent with cyst formation secondary to leakage from the newly formed vessels. The morphology of the lacZ ovaries was normal (Fig. 6j). Thus, VEGF and EG-VEGF are able to trigger the same chain of events in a responsive tissue, including the ability to induce fluid extravasation, in addition to angiogenesis. This could have been expected, in view of the striking similarity of the two factors in their effect on responsive endothelial cells, including the ability to induce fenestration.

**Discussion**

VEGF was identified as a mitogen specific to endothelial cells more than ten years ago. Such selectivity, which distinguished it from other endothelial cell mitogens, led to the hypothesis that this molecule might play a unique role in the regulation of blood vessel growth. Subsequent studies indeed demonstrated the essential role of VEGF in a surprisingly broad variety of normal and abnormal angiogenic processes, including embryogenesis, postnatal growth, corpus luteum development, growth plate morphogenesis, tumorigenesis and ischemic retinal diseases. However, it was difficult to explain how a single, ubiquitously expressed protein could trigger such a diverse range of processes.

![Northern blot analyses of human RNA samples from different tissues with an EG-VEGF specific probe.](image)

**Figure 4** Northern blot analyses of human RNA samples from different tissues with an EG-VEGF specific probe. A single EG-VEGF transcript of ~1.4 kb was identified. Expression is highest in ovary and testis, followed by adrenal and placenta. A signal is detectable in prostate only after prolonged exposure. Equivalent RNA loading was assessed by hybridization with the control actin probe (data not shown). Contents of the lanes are indicated above the blots, and the size is indicated at the right. PBL, peripheral blood leukocytes.

![Expression of EG-VEGF mRNA in testis and ovary. Bright-field images and the corresponding dark-field images are shown.](image)

**Figure 5** Expression of EG-VEGF mRNA in testis and ovary. Bright-field images and the corresponding dark-field images are shown (a and b, c and d, and so on). a–d. In situ hybridization studies revealed that, in the human testis, EG-VEGF transcript is restricted to the testosterone-producing Leydig cells. e, f. EG-VEGF signal is prevalent in the human ovarian stroma in cells that can be described as perivascular (k, l). A very similar pattern was evidenced in the chimpanzee ovary (data not shown). g, h. A strong signal is detected in the human corpus haemorrhagicum. In developing follicles of humans (i, j) and chimpanzees (m–p), EG-VEGF RNA is expressed in the theca and granulosa cells—both steroidogenic cell types—and the cumulus oophorus. Arrows indicate the course of a blood vessel. Scale bars: a, b, g, h, 0.1 mm; c, d, i–l, 25 μm; e, f, 115 μm; m–p, 55 μm.
expressed molecule might account for the great diversity of phenotypic and growth characteristics of the endothelia in those circumstances. Conceivably, while VEGF signalling represents a rate-limiting step in many situations, local regulatory pathways modulate endothelial cell growth and function in a tissue- or organ-specific manner, synergistically with VEGF. Furthermore, recent data suggest the existence of VEGF-independent angiogenic pathways that remain to be identified. For example, early results of clinical studies with VEGF inhibitors in cancer patients have provided initial evidence of efficacy. However, progression eventually occurs in advanced malignancies, in spite of the anti-VEGF treatment.

Our results demonstrate that steroidogenic glands express a novel angiogenic and permeability-enhancing factor, EG-VEGF, with unique target cell selectivity. The partially complementary expression patterns of VEGF and EG-VEGF indicate that these molecules may function in a coordinated or complementary manner to regulate angiogenesis and permeability as well as to maintain a

Figure 6 Selectivity of the in vivo angiogenic effects of EG-VEGF. a–c, Rat corneal pocket assay. Note the strong angiogenic response induced by VEGF protein, whereas EG-VEGF has essentially no effect. d–f, Injection of adenovirus CMV-lacZ, CMV-VEGF, or CMV-EG-VEGF (5 × 10⁸ PFU) in the skeletal muscle (sm) of nude rats. Arrowheads indicate microspheres marking the injection site; arrows indicate new blood vessels. Note the angiogenic response, with abundant new vessel formation, induced by VEGF, whereas both lacZ and EG-VEGF adenovirus had no appreciable effects. g, h, VEGF and EG-VEGF adenovirus injection in the mouse ear. Again, VEGF adenovirus resulted in a strong angiogenic response, which was absent in the animals injected with EG-VEGF adenovirus. i, Gross appearance of ovaries (ov) following injection of lacZ (C, control), VEGF or EG-VEGF adenovirus, after 7 d. Note the much larger mass plus the presence of superficial vessels and haemorrhagic areas in both VEGF and EG-VEGF groups. j–l, Micrographs of ovaries injected with adenovirus vectors (5 × 10⁸ PFU), as indicated. Note the normal architecture and morphology of the lacZ-treated ovary in j. Corpora lutea (CL) are indicated. In contrast, the VEGF (k) and EG-VEGF (l) groups revealed very similar changes, with large fluid-filled or haemorrhagic cystic areas (asterisks). m, n, High-power micrographs of boxed areas in k and l, respectively. Areas of intense angiogenesis are evident at the periphery of cystic lesions. Arrows point towards blood vessels. Scale bars: d–h, 100 μm; i, 0.5 cm; j–l, 1 mm; m, n, 33 μm.
Recombinant Fc/His fusion proteins

A cDNA clone encoding human EG-VEGF was identified in the Lifeseq EST database (Incyte Genomics) on the basis of an algorithm that predicted the signal peptide and was sequenced in its entirety. The coding sequences of human EG-VEGF were amplified by polymerase chain reaction (PCR) and subcloned into the expression vector pVL1393 (PharMingen). The fusion proteins were expressed in H5 cells by recommended procedures (Invitrogen). In brief, the DNA constructs were cotransfected with BaculoGold baculovirus DNA (PharMingen) in a 7:1 ratio into adherent Sf9 cells. Cells were incubated at 28°C for 2–4 d, and the supernatant was collected. The transfection supernatant was amplified and subjected to affinity purification by either protein A Sepharose beads (Pharmacia) for Fc fusion proteins or Ni-NTA agarose beads (Qiagen) for His-tagged proteins. To examine the protein expression, SDS–PAGE analysis was performed on the affinity-purified recombinant proteins under non-reducing and reducing conditions, followed by staining. Proteins were routinely submitted for amino-terminal sequencing and endotoxin measurements, which were below 1 endotoxin unit (EU) per mg. CHO cells expressing the untagged EG-VEGF were established and the highest-producing clones were expanded. To determine whether EG-VEGF binds to heparin, we applied ~15 µg of untagged protein into a 1-ml Hi-Trap heparin—sepharose column (Pharmacia) in 10 mM Tris, pH 7.4, 0.1 M NaCl. The column was eluted with a linear gradient of NaCl (0.1–2 M). Human VEGF165 was used as a reference.

Proliferation assays

Media and other cell culture reagents were obtained from Life Technologies. ACE cell proliferation assays were performed as previously described23. HUVECs, HMVECs and human keratinocytes were purchased from Clonetics. Bovine pericardium and human neonatal fibroblasts were a gift of M. Gerritsen. For proliferation assays, 4,000–6,000 cells per well were plated in 12-well dishes. Controls included wells in the basic assay media with no other added factors, bFGF or VEGF, in the case of endothelial cell assays, served as positive controls. Cell numbers were determined after 5–7 d.

Migration assays

ACE cells, BAEcs, MS-1 cells and HUVECs were used for migration assays. Primary baboon endometrial endothelial cells were isolated from adrenal glands of premature or fetal baboons, which were a gift of R. Clyman. The capsule was removed and the remaining tissue was finely chopped with sterile razor blades into fragments ~2 mm³. The fragments were subsequently incubated at 37°C for 30–40 min in 0.1% collagenase (Sigma) in 50:50 Ham’s F10 : DMEM media, 10% FCS with the addition of D2Nase (Life Technologies). Single-cell suspensions were washed, re-suspended in PBS with 5% FCS, incubated with 1 µg of anti-KDR monoclonal antibody for 10 min, washed, and then incubated with goat anti-mouse antibody conjugated to fluorescein isothiocyanate (FITC, Sigma). The endothelial cell population was maintained in CSM medium containing 10% serum and growth factors (Cell Systems). Filter inserts (8.0 µm, Falcon 3907) were coated with 1% gelatin.

Endothelial cells were cultured before being assayed, as described above. Cells were digested with trypsin and transferred to endothelial basal media (EBM, Clonetics) with 0.1% BSA for the assay. Cells were plated at 1–5·10⁴ per upper chamber and growth factors were added in the lower chamber. The assay was routinely 16 h at 37°C. At completion, we removed cells from the upper side of the membrane by scraping with a polycarbonate swab; the remaining cells on the bottom of the membrane were then fixed with methanol. Cells were stained with propidium iodide and counted under low power with the Image-Pro program.

Fenestration assays

Extracellular matrix (ECM) was prepared as previously described25. Briefly, corneal endothelial cells were isolated from steer eyes (Pel Freez) and these were expanded in 50:50 Ham’s F10 : DMEM media supplemented with 15% FCS, penicillin–streptomycin, fungizone. To prepare the ECM-coated plates, 4·10⁴ cells were plated per well in 6-well dishes and cultured for ~10 d in low-glucose DMEM supplemented with 10% FCS, 2.5% dextran (Sigma) and penicillin–streptomycin. At the end of 10 d, the cells were quickly lysed in 0.02 M NH₄OH in water, rinsed several times with PBS and stored 4°C in PBS with antibiotics. ACE cells were plated at a density of 1–2·10⁴ and grown to confluence. No addition, 10 nM EGF, 2.5 nM VEGF, 10 nM EG-VEGF or 2.5 nM VEGF plus 10 nM EG-VEGF were added to individual wells, in at least duplicate. The fenestration assays were replicated three times. Cells were rinsed with PBS and fixed for 2 h in 2% formaldehyde, 2.5% glutaraldehyde in 0.1 M cacodylate buffer. After washing, the samples were post-fixed in aqueous 1% osmium for 2 h, washed in water, dehydrated through graded ethanol and propylene oxide, and embedded in EPONATE 12 (Ted Pella). Ultrathin sections were cut on a Reichert Ultracut E microtome, counterstained with uranyl acetate and lead citrate and examined in a Philips CM12 transmission electron microscope at 80 kV; images were captured with a GE Medical Systems digital camera. One hundred cell profiles were examined for each condition in triplicate.

Expression analyses

Human multiple tissue polyA⁺ RNA array and human polyA⁺ RNA multiple tissue northern blots were purchased from Clontech. cDNA probes were prepared using 30–50 ng of the human cDNA fragments with the Redi-Prime II kit (Amersham), using [³²P]-dCTP at 3,000 Ci mmol⁻¹ (Amersham). Probes were purified on Sephadex G50 spin columns (Pharmacia) and hybridizations were carried out at 68°C in ExpressHyb hybridization solution (Stratagene). Tissues were processed for in situ hybridization by a method described previously26. [³²P]-UTP-labelled RNA probes were generated as described27. Sense and antisense probes were synthesized from a cDNA fragment corresponding to nucleotides 219–958 of the human sequence.

Hypoxic regulation studies

For expression analysis, RNA isolates from replicate, matched samples of 5W13 and H295R cells (both from the ATCC) treated with normal-oxygen or hypoxic conditions were prepared using the Rneasy kit (Qiagen) as described by the manufacturer. For real-time quantitative PCR with reverse transcription (RT-PCR), 50 ng of total RNA was
assayed in triplicate with the Perkin Elmer Tagman kit reagents and an ABI prism 7700 Sequence Detector. Oligonucleotides and probes used were as follows: EG-VEGF forward, 5'-CGGCGAGGACCAAGTG-3'; EG-VEGF reverse, 5'-TGGCGCAAGGAAGAACGGACG-3'; EG-VEGF probe, FAM-5'CTCTTCAGAATAAGGAGCACA-3'TAMRA); VEGF forward, AATGAGGCGGCTCTGAGT; VEGF reverse, TGTAGCAGATATGCTCTGTA; VEGF probe, FAM-5'TTTGAGCAAGAAGACTTTGA-3'TAMRA); β-actin primers and probe were obtained from the manufacturer (Perkin-Elmer).

The luciferase reporter constructs were generated by cloning compatible, annealed oligonucleotides into the BglII site of the pG3-Luciferase vector (Promega). HRE sens/heatshock-luciferase reporter system (Promega) were transfected the recombined AdEasy plasmids into host HEK293 recombined, in BJ5183 electrocompetent bacteria (Stratagene), with the AdEasy vector and secured using non-traumatic forceps. Doses of 10^8 or 5×10^7 high-molecular-mass FITC-dextran to allow for visualization of the vasculature. Measurements of newvascular area in corelional whole mounts were performed using computer-assisted image analysis (Image-Pro Plus).

For adenoviral injections in the ovary, athymic nude rats were anaesthetized with isofluorane, the area prepared by transiently transfecting the recombined AdEasy plasmids into host HEK293 cells. Adenovirus stocks were further amplified in HEK293 cells and purified using the Virakt Adeno purification kit (Virapur). Adenovirus titres were obtained by agarose-overlaid plaque assays.

In vivo assays
Cornel polycyst angioigenesis assays in adult rats were performed as previously described44. Hydron-coated pellets containing 200 ng VEGF165 or 500 ng EG-VEGF, plus methylcel- lulose and aluminium sulphate, were inserted into the base of the pocket (n = 6). Control pellets contained 500 ng CD4-IGc. At day 6, animals were killed and injected with high-molecular-mass FITC-dextran to allow for visualization of the vasculature. Measurements of newvascular area in corelional whole mounts were performed using computer-assisted image analysis (Image-Pro Plus).

For adenoviral injections in the ovary, athymic nude rats were anaesthetized with isofluorane, and a 2–2.5 cm incision was made in the left dorsal area. The ovary was lifted and secured using non-traumatic forces. Doses of 10^6 or 5×10^7 plaque-forming units (PFU) in 5–10 μl saline were injected by a gas-tight Hamilton syringe fitted with a 31-gauge needle (Hamilton). All procedures were performed in a biosafety cabinet with biosafety level 2 practices. For skeletal muscle (left gastrocnemius) or subcutaneous ear injections, athymic nude mice or rats were anaesthetized with isofluorane, the area was cleaned, and doses of 5×10^6 PFU of each virus preparation in a 50–μl volume was injected per site. Animals in all adenovirus studies were killed 1 week after adenovirus administration. At necropsy, tissues were dissected and fixed or frozen for histological analysis.

Received 31 May; accepted 25 July 2001.

27. Supplementary information is available on Nature’s World-Wide Web site (http://www.nature.com/).
Characterization of extrasolar terrestrial planets from diurnal photometric variability

E. B. Ford†, S. Seager† & E. L. Turner*

*Princeton University Observatory, Peyton Hall; and †Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA

The detection of massive planets orbiting nearby stars has become almost routine2-3, but current techniques are as yet unable to detect terrestrial planets with masses comparable to the Earth's. Future space-based observatories to detect Earth-like planets are being planned. Terrestrial planets orbiting in the habitable zones of stars—where planetary surface conditions are compatible with the presence of liquid water—are of enormous interest because they might have global environments similar to Earth's and even harbour life. The light scattered by such a planet will vary in intensity and colour as the planet rotates; the resulting light curve will contain information about the planet's surface and atmospheric properties. Here we report a model that predicts features that should be discernible in the light curve obtained by low-precision photometry. For extrasolar planets similar to Earth, we expect daily flux variations of up to hundreds of per cent, depending sensitively on ice and cloud cover as well as seasonal variations. This suggests that the meteorological variability, composition of the surface (for example, ocean versus land fraction) and rotation period of an Earth-like planet could be derived from photometric observations. Even signatures of Earth-like planet life could be constrained or possibly, with further study, even uniquely determined.

NASA and ESA are now considering two ambitious space missions—TPF* and Darwin† respectively—to detect and characterize terrestrial planets orbiting nearby Sun-like stars. Although very different designs are being considered4,5,6, all have the goal of spectroscopic characterization of the atmospheric composition, and in particular the detection of gases that are important for or caused by life on Earth, including O$_2$, O$_3$, CO$_2$, CH$_4$ and H$_2$O (ref. 7). A mission capable of measuring these spectral features would necessarily obtain sufficient signal-to-noise ratios to measure photometric variability. Many other important properties of an extrasolar planet could be derived from photometric measurements; this is the motivation for the investigation presented here. Photometric variability would be especially valuable for studying many planets quickly or a planet that is too dim for spectroscopic studies. Moreover, the photometric variability could be monitored concurrently with a spectroscopic investigation, as was done for the transiting extrasolar giant planet of HD209458 (ref. 8).

We have developed a code which calculates the total light scattered by an extrasolar planet towards an observer. The code performs Monte Carlo integrations, with single scattering, over a spherical planet using a map that specifies the scattering surface type at each point on the sphere; the code also uses a set of wavelength-dependent bidirectional reflectance distribution functions that specify the probability that light incident from one direction will scatter into another direction for each type of scattering surface (refs 9-19; see also B. Rock and J. Salisbury, http://specilib.jpl.nasa.gov/archive/jhu/becknic/vegetation/txt/deciduous.txt). The observed flux also depends on the viewing geometry which is specified by the phase angle (the angle between the star, planet and observer), obliquity (the direction of the planet's rotation axis), orbital inclination relative to the line of sight, and time of day (t). Although the geography and climate of an Earth-like extrasolar planet are completely unknown, we can attempt to calculate what an optical or near-infrared light curve of Earth would look like if viewed from a very large distance. Simple variations about this basic reference model then allow us to explore reasonable possibilities for other Earth-like planets. We use a map of Earth from a one square degree satellite surface map which classifies each pixel as permanent ice, dirty/temporary ice, ocean, forest, brush, or desert9. We consider cloudy models separately, using the scattering properties of Earth clouds21. We focus our attention on quadrature (a phase angle of 90°) for which the planet–star separation is largest and the observational constraints thus least severe.

The diurnal light curve from our cloud-free Earth model has variation as high as 150% (Fig. 1). The significant intensity variation is due to the facts that surfaces have different albedos and a relatively small part of the visible hemisphere dominates the total flux from an unresolved planet. For example, at quadrature less than 10% of the surface often produces more than 50% of the total reflected light. In our cloud-free Earth model the variability is primarily due to land (lower specular reflection) and ocean (higher) rotating in and out of view. The peak in the light curve at t = 0.5 day is caused by the high albedo of sand from the Sahara desert, and the dip at t = 0.8 day is caused by South America rotating into the location where there is usually specular reflection off the ocean. Ice, sand, oceans and vegetation can all produce significant features in the rotational light curve, but distinguish themselves by their colours (for example, ice is very nearly grey, whereas sand has an albedo which rises by about 30% between 450 nm and 750 nm).

On Earth clouds are extremely important to the reflected light curve both because of their very high albedos and their short timescales of formation, motion and dissipation. Figure 2 shows the diurnal light curve of our Earth model with daily, seasonal and annual average cloud coverage maps from the ISCCP database (ref. 22 and see http://isccp.giss.nasa.gov/). Here the intensity variation as high as 150% (Fig. 1). The significant intensity variation is due to the facts that surfaces have different albedos and a relatively small part of the visible hemisphere dominates the total flux from an unresolved planet. For example, at quadrature less than 10% of the surface often produces more than 50% of the total reflected light. In our cloud-free Earth model the variability is primarily due to land (lower specular reflection) and ocean (higher) rotating in and out of view. The peak in the light curve at t = 0.5 day is caused by the high albedo of sand from the Sahara desert, and the dip at t = 0.8 day is caused by South America rotating into the location where there is usually specular reflection off the ocean. Ice, sand, oceans and vegetation can all produce significant features in the rotational light curve, but distinguish themselves by their colours (for example, ice is very nearly grey, whereas sand has an albedo which rises by about 30% between 450 nm and 750 nm.)

On Earth clouds are extremely important to the reflected light curve both because of their very high albedos and their short timescales of formation, motion and dissipation. Figure 2 shows the diurnal light curve of our Earth model with daily, seasonal and annual average cloud coverage maps from the ISCCP database (ref. 22 and see http://isccp.giss.nasa.gov/). Here the intensity variation as high as 150% (Fig. 1). The significant intensity variation is due to the facts that surfaces have different albedos and a relatively small part of the visible hemisphere dominates the total flux from an unresolved planet. For example, at quadrature less than 10% of the surface often produces more than 50% of the total reflected light. In our cloud-free Earth model the variability is primarily due to land (lower specular reflection) and ocean (higher) rotating in and out of view. The peak in the light curve at t = 0.5 day is caused by the high albedo of sand from the Sahara desert, and the dip at t = 0.8 day is caused by South America rotating into the location where there is usually specular reflection off the ocean. Ice, sand, oceans and vegetation can all produce significant features in the rotational light curve, but distinguish themselves by their colours (for example, ice is very nearly grey, whereas sand has an albedo which rises by about 30% between 450 nm and 750 nm.)

### Figure 1
Rotational light curve for a cloud-free Earth model. The light curve is sampled at four-minute intervals and the Poisson noise is due to Monte Carlo statistics in our calculations. The pink, red, green and blue curves correspond to wavelengths of 750, 550 and 450 nm, and their differences reflect the wavelength-dependent albedo of sand, ice, ocean and vegetation. The error bars in all figures show estimates of the photometric variability. Many other important properties of an extrasolar planet could be derived from photometric measurements; this is the motivation for the investigation presented here. Photometric variability would be especially valuable for studying many planets quickly or a planet that is too dim for spectroscopic studies. Moreover, the photometric variability could be monitored concurrently with a spectroscopic investigation, as was done for the transiting extrasolar giant planet of HD209458 (ref. 8).

![Figure 1](https://example.com/figure1.png)
variation is due to the contrast between clouds and land or ocean. Clouds tend to raise the overall brightness and variability, reducing the fractional variation in the reflected light curve compared to the cloud-free case (Fig. 1). On Earth cloud patterns cause a variation of about 20% within a typical day. Cloud patterns can be coherent over several days (Fig. 2a), allowing the rotational period of the Earth to be measured from its light curve and thus the light curve to be averaged over many days. Light curves of Earth with seasonal and annual cloud coverage maps (Fig. 2b) demonstrate that average cloud patterns for Earth vary with seasons. These seasonal light curves reflect the fact that there are locations that are almost always cloudy (such as the Amazon Basin) and regions that are virtually cloud free (such as the Sahara desert) at different times of the year.

We could learn about Earth's unresolved rotationally modulated flux from a satellite from Earth or by observing the reflection of the Earth from the dark side of the moon (Earthshine). Goode et al. have measured Earthshine for about 200 days since 1998. Our model agrees with their measurements for both the mean reflectivity (0.086 for the model versus 0.092 that we observed) and fractional variance (13% model versus 15% observed) at a phase angle of between 80° and 100°. Their observations are made from a single location, limiting their range of viewing geometries. Monitoring Earthshine from several different longitude locations will allow for more complete daily coverage.

While the Earthshine observations affirm the accuracy of our Earth model, they cannot address the light curves of extrasolar planets which will not be identical to those of Earth. The rotational period, the continental fraction and arrangement could be completely different. In addition, the cloud fraction and spatial and temporal distribution depends on many variables such as the location of continents and obliquity. Predicting cloud coverage is very difficult for global climate models, so we use cloud-free models to consider plausible Earth-like planets by varying the surface map (Fig. 3). Figure 3 shows daily light curves for planets with several different surface maps which illustrate that models with qualitatively different surfaces produce distinctly different light curves. For example, an Earth-like planet whose land was covered by ice (Fig. 3a) or thick forests (Fig. 3b) would have much larger amplitude variations than the Earth's. Varying the fractional ocean coverage (Fig. 3c) affects both the normalization and the variability of the light curve. Whereas Fig. 3c and d demonstrates that an extrasolar planet's diurnal light curve may contain information about the planet's surface, the full inverse problem of obtaining a unique determination of surface features from a light curve may be intractable and will certainly require much more investigation. For example, changing the obliquity of our cloudless Earth model also causes significant changes in the light curve (Fig. 3d). Nevertheless, a planet's light curve will clearly place constraints on its surface properties and climate in that many possible models would be ruled out by any specific set of photometric data.

In addition to studying 'geological' surface features we might learn about biological features encoded on the planet's surface. On Earth vegetation has a dramatic sudden rise in albedo by almost an order of magnitude at around 750 nm, known as the 'red edge'. Vegetation has evolved this strong reflection as a cooling mechanism to prevent overheating which would cause chlorophyll to degrade. Although the albedo of sand also increases towards the infrared, the red-edge vegetation signature is more rapid and may be detectable from the unresolved Earth. (See also the Galileo observations of part of Earth.) We cannot necessarily expect to find Earth-like

Figure 2 Rotational light curves for model Earth with clouds. Model daily light curves are shown at 550 nm for our Earth model with clouds, as viewed from quadrature. Each panel represents a different surface map: a. Theoretical light curves using cloud cover data from satellite measurements taken on six consecutive days (13–18 April 1986). b. Theoretical light curves for Earth using seasonal (dotted and dashed lines) and annual (solid line) average cloud cover (averaged over 8 years). Using actual cloud data allows us to accurately model the Earth, but is not applicable to extrasolar terrestrial planets.

Figure 3 Rotational light curves for cloudless Earth-like models. Model daily light curves are shown at 550 nm for an Earth-like planet, as viewed from quadrature. The light curves shown in a–c are generated using variations on our cloud-free Earth model, which is presented as the solid black line in each panel for reference (corresponding to the green curve in Fig. 1). These models do not include clouds because the cloud coverage would differ for each Earth-variant in a way not yet known. a. Light curves for Earth-like planets for which all the land is covered with ice (dashed line) and none of the land is covered with ice (dotted line). b. Light curves for Earth-like planets for which all the land is covered with thick forests (dashed line) and all of the land is covered with desert (dotted line). c. Light curves for Earth-like planets for which the land and oceans have been recursively expanded and contracted along their perimeters to achieve the desired fractional ocean coverage. d. Light curves for our cloudless Earth model for different obliquities. The north pole has been tilted towards the Sun by 0° (solid line), 30° (dotted line), and 90° (dashed line).
vegetation on extrasolar terrestrial planets, but photometric measurements in different colours may be able to detect a unique signature, different from any known surface features or atmospheric constituents on Earth or other solar system planets. We expect the diurnal rotational variation of an Earth-like planet to be lower in the mid-infrared (flux variation of a few per cent) than in the optical because the surface temperature does not vary as much as surface albedo across the Earth. For a planet with non-zero obliquity the mid-infrared seasonal flux variation should be larger than the mid-infrared rotational variation because of seasonal temperature variation. In addition, optical multicolour photometry and polarization as a function of phase angle (on an orbital timescale) could help constrain the type of atmospheric scatterers and the particle size distribution or detect a large smooth specular surface such as an ice- or an ocean-covered planet. Alternatively, the variation on an orbital timescale may be overwhelmed by variations indicating strong seasonal changes in the atmosphere.

Theoretical light curves of the unresolved Earth have a variation of 10–20%. A terrestrial planet finder that can measure 5% optical variation could detect weather, the rotational period, and seasonal changes in the cloud pattern of a planet like our Earth. If the surface were to contribute most of the scattered light rather than the clouds, then planets with different surface features would show very different diurnal light curves. This is in dramatic contrast to planets like Venus which would show almost no diurnal variability. Thus, we expect the diurnal light curve of an extrasolar Earth-like planet to contain detectable features encoding information about its physical and even biological properties.
which is a feature of laser operation. 

The set-up used to demonstrate stimulated entanglement is illustrated in Fig. 2. A 120-fs pulse pump at 390 nm wavelength (with a repetition rate of 80 MHz) passes through a β-barium borate (BBO) crystal and creates pairs of polarization-entangled photons in spatially distinct modes a and b. The experimental parameters are chosen such that (to first order) the singlet photon-pair state shown in equation (3) is created. Initially modes a and b are in the vacuum state and the photon pairs are spontaneously created. The fact that modes a and b geometrically diverge, and that horizontally and vertically polarized photons experience different crystal parameters, limits the useful crystal length and thereby prohibits an efficient stimulated emission process. To obtain significant stimulated emission we redirect the spontaneously created photon pairs into the crystal at the same time (tuned by a delay on mirror M3) as the reflected pump pulse passes through the crystal a second time. Provided that the feedback loop for the photon pairs is polarization independent, which is obtained by using a bow-tie folded geometry including a λ/2 waveplate that exchanges H and V polarizations, optimum conditions for stimulated emission of photon pairs can be established. As stimulated emission can be seen as a constructive multi-particle interference effect, and because the process of parametric down-conversion is sensitive to the phase of the pump, we should expect to observe an oscillation between stimulated and suppression of emission as function of the pump-pulse delay. The period of this oscillation corresponds to the optical frequency of the pump laser. In the region where the difference between the pump delay and the feedback loop is larger than the coherence length of the observed photons (determined by the 5-nm narrow-bandwidth filters in front of the single-photon detectors), no such interference pattern is expected.

To study the 2- and 4-photon entangled states we measure each term in equations (3) and (4) individually in two non-orthogonal polarization bases. The |1,0;0,1⟩ and |0,1;1,0⟩ terms are measured in the desired bases by using a polarizer in front of a single-photon detector in each of the spatial modes a and b. The |1,1;1,1⟩ term is detected by the introduction of polarizing beam splitters in the appropriate basis in each mode followed by four single-photon detectors (Fig. 2). As we do not use multi-photon detectors, we can demonstrate stimulated entanglement by measuring 2- and 4-photon properties of state (2) for increasing values of τ.

The fact that modes a and b are in different states and only a few experiments and proposals have addressed second-order terms. By analogy with a conventional laser, the idea of a tensor product of states for the gain medium, which enhances the emission of the higher-order terms in equation (2). The state shown in equation (2) has the following features.

First, modes a and b are entangled in photon number because, for any n, the number of photons in each mode is identical. The photon (pair) number distribution is shown in Fig. 1 for increasing average photon (pair) number output per pulse 〈n〉. The shifting of the maximum and broadening of the distribution for higher values of 〈n〉 resembles the coherent state photon-number distribution as produced by conventional lasers. These features are explained by the fact that stimulated emission—originating from the boson statistics of photons—favours amplification of higher over lower photon-number terms.

The second important property of the state in equation (2) is that the set of terms for each n form a maximally entangled state in polarization. The normalized 1-pair term is the rotationally symmetric Bell state (singlet spin-1/2):

\[
|\psi\rangle = \frac{1}{\sqrt{2}} (|1,0;0,1\rangle - |0,1;1,0\rangle)
\]

The normalized 2-pair term is given by

\[
|\psi\rangle = \frac{1}{\sqrt{3}} (|2,0;0,2\rangle - |1,1;1,1\rangle + |0,2;2,0\rangle)
\]

and represents the singlet spin-1 state. Similar to the spin-1/2 case, the rotational symmetry arises from the relative phase relations and the equal weights of the terms. In general, the n-pair term has the properties of a singlet spin-n/2 state. The rotational symmetry of the full state can be shown by expressing the n-pair terms in any other basis, and verifying that the same expressions are obtained. The crucial role of stimulated emission is to provide for each n equally weighted terms. In principle, a photon counting measurement on state (2) (either in mode a or b) performs a projection onto a certain singlet spin-n/2 state. Subsequently, this maximally entangled state can be explored for quantum information tasks. In practice, in quantum optics experiments where the fragile photons are in general destroyed by any measurement, the projection and the exploration of the state are performed simultaneously. This procedure, usually referred to as post-selection, has proved to be very useful: for example, for demonstrations of quantum teleportation,

quantum cryptography, and three-particle Greenberger–Horne–Zeilinger (GHZ) correlations, and for novel optical quantum computation schemes. Here we use post-selection to demonstrate stimulated entanglement by measuring 2- and 4-photon properties of state (2) for increasing values of τ.

The photon number (pair) distribution, P(n), arising from stimulated parametric down-conversion shifts its peak and broadens as the mean number of photons increases. This indicates that for increasing interaction strength (gain), terms with higher numbers of photons obtain a larger amplification factor compared to lower terms, which is a feature of laser operation.
only measure the $|2,0;0,2\rangle$ and the $|0,2;2,0\rangle$ term with a 0.25 probability using a combination of a polarizer, a 50-50 beam splitter and two single-photon detectors in each mode.

Quantitative predictions for the amplification of the individual terms in equations (3) and (4) resulting from the double-pass configuration are obtained by expanding the unitary evolution of the created light fields in the polarization-entangled photon-pair creation operator $\hat{K}$. To second order in $\hat{K}$ we obtain

$$\hat{U} = \hat{U}_2 \hat{U}_1 = 1 + e^{i\theta} \hat{K}^\dagger + r \hat{K}^\dagger_2 + \frac{1}{2} e^{i\theta} \tau \hat{K}^\dagger_2 + \frac{1}{2} \tau^2 \hat{K}^\dagger_2 + e^{i\theta} \tau \hat{K}^\dagger_2 \hat{K}^\dagger_2$$

where subscripts 1 and 2 refer to the first and second pass through the crystal. The relative phase $\theta$ between the first and second pass of the pump pulse is tunable by the translation of mirror M3 (Fig. 2). We discuss two limits: when $K_2^\dagger = K_1^\dagger$, and when $K_2^\dagger$ is distinguishable from $K_1^\dagger$. The first case applies at zero delay where efficient phase-sensitive stimulated emission occurs. From equation (5) it follows that doubling the value of the interaction parameter $r$ results in an increase in probability for the 2-photon terms from $r^2$ to $4r^2$, and in an increase for the 4-photon terms from $r^4$ to $16r^4$. We note that the 4-photon state has a four times larger amplification than the 2-photon states, which is characteristic of stimulated emission. The second case applies if the reflected pump-pulse delay between the two passes is not equal to the delay of the entangled photons in the feedback loop. In this case, there are simply two independent contributions to the 2-photon detection events, but there are several distinct contributions to the 4-photon detection events. Each single pass has a small probability (of $r^4$) to create state (4). In addition, as current single-photon detectors do not have a high enough time resolution to distinguish between photons arriving from the first or second pass, there are spurious 4-fold coincidences from a combination of 2-photon states created in both passes. The spurious contributions to the $|2,0;0,2\rangle$ and the $|0,2;2,0\rangle$ detections will be $2r^4$, and for the $|1,1;1,1\rangle$ detection will be $2r^4$.

We scan from the region where $K_2^\dagger$ is completely distinguishable from $K_1^\dagger$ into the region where $K_2^\dagger = K_1^\dagger$, while observing the intensity of the 2- and 4-photon terms. From the considerations above we expect—in the case of ideal stimulated emission—the terms in equation (3) to show a 2-fold increase, the middle term in equation (4) to show a 4-fold increase, and the other two terms in equation (4) to increase by a factor of 16/3 = 5.33. Owing to the rotational symmetry of states (3) and (4), these predictions are basis independent.

Figure 3 shows our experimental data for the detection of the $|1,0;0,1\rangle$ (panel a) and the $|2,0;0,2\rangle$ (b) terms measured in the 45°
rotated basis, and the $|1,1,1,1\rangle$ (c) term in the H/V basis. The solid curves are the envelopes of the oscillating functions giving the maximum and minimum theoretical values for the coincidence rates. The experimental data show an increase of $1.95 \pm 0.10$ for $|1,0,0,1\rangle$, $5.3 \pm 0.6$ for $|2,0,0,2\rangle$ and of $4.1 \pm 0.3$ for $|1,1,1,1\rangle$. These results are in good agreement with the predictions discussed above. Similar results have been obtained in the other bases and for the $|0,1,1,0\rangle$ and $|0,2,2,0\rangle$ terms, demonstrating the rotational invariance—that is, the spin-1/2 and spin-1 singlet structure—of states (3) and (4). Additional data indicate an amplification due to the second pass of 3.95 $\pm 0.10$ for the 2-fold coincidences, and of $17 \pm 2$ for the 4-fold coincidences. This demonstrates the shifting of the photon-number pair distribution towards terms with higher photon numbers, a characteristic of stimulated emission. A final comparison of obtained results are in good agreement with the predictions discussed above.

Using multi-pass amplification pumped by higher-intensity pulses it should be possible to produce rotationally symmetric Bell state; we have also demonstrated that stimulated emission of polarization-entangled photons can be achieved experimentally. Both the characteristic shifting of the photon pair distribution towards higher photon numbers, and the rotational symmetry of the 2- and 4-photon contribution, have been observed. The good agreement between the experimental data and the theory shows that the stringent indistinguishability requirements to obtain entanglement in a stimulated process using external resonators have been met. Although related theory and experiments on interference-enhanced emission of photon pairs and on photon injection into nonlinear crystals have been reported,$^2^2,^2^3,^2^7,^2^8$ our results constitute (to our knowledge) the first experimental demonstration of the onset of laser-like operation for entangled photons.

Using multi-pass amplification pumped by higher-intensity pulses it should be possible to produce rotationally symmetric multi-photon entangled states with an average photon (pair) number of the order of 100. As exploration of such states is based on post-selection, the challenge of creating them should go in parallel with the challenge of constructing low-loss transmission lines and high-efficiency multi-photon detectors. Although there are encouraging developments in low-loss optical fibres$^2^2$ and highly efficient multi-photon detectors,$^2^0$ we will always have to face the situation of losing photons in the process of creating, transporting and analysing the desired state of equation (2). As this state is one large, complex, entangled state, one might think that the loss (or a measurement) of a single photon would destroy all the interesting properties of the state. On the contrary, the complex entanglement has the remarkable feature that the loss or measurement of one (or more) particles does not eliminate all the entanglement between the remaining particles. To illustrate this point, we focus our attention on the 4-photon state of equation (4), and consider a measurement of a photon in mode $a$ in the H/V basis, with the measurement result being H. The state of the remaining three particles will be $(1/\sqrt{2})|1,0,0,2\rangle – |0,1,1,1\rangle$, which still contains non-maximal entanglement between modes $a$ and $b$. A generalization to the actual loss of several photons from the full state (2) is currently under study.

The increased sharpness of the peaks in the interference pattern for the singlet of spin-1 (see Fig. 4) is potentially of interest for applications in metrology. An interferometric measurement of relative distance, for example, would benefit from the more precise location of the maxima obtained by looking at the 4-fold (or higher) coincidence rates. Furthermore, the increased amount of entangled terms made available by stimulated emission and post-selection offer new possibilities for higher-bit-rate quantum cryptography. We consider that entanglement robustness—together with the rotational symmetry of the state created by stimulated polarization entanglement—opens the way to many applications in quantum information, and provides a powerful tool for studying the almost unexplored area between the discrete and the macroscopic optical quantum correlation experiments.

Received 2 May; accepted 3 July 2001.


Acknowledgements

We thank C. Mikkelsen, P. Varisco, W. Irvine, A. Ekert and J. Rarity for suggestions and experimental support. This work was supported by the EPSRC, the UK Defence Evaluation and Research Agency, and the European QuComm project.

Correspondence and requests for materials should be addressed to A.-L. (e-mail: a.lamas@qubit.org).
Mechanism for the destruction of $\text{H}_3^+$ ions by electron impact

V. Kokouline*, Chris H. Greene* & B. D. Esry†

* Department of Physics and JILA, University of Colorado, Boulder, Colorado 80309-0440, USA
† Department of Physics, Cardwell Hall, Kansas State University, Manhattan, Kansas 66506, USA

The rate at which the simplest triatomic ion ($\text{H}_3^+$) dissociates following recombination with a low-energy electron has been measured in numerous experiments. This process is particularly important for understanding observations of $\text{H}_3^+$ in diffuse interstellar clouds. But, despite extensive efforts, no theoretical treatment has yet proved capable of predicting the measured dissociative recombination rates at low energy, even to within an order of magnitude. Here we show that the Jahn–Teller symmetry-distortion effect—almost universally neglected in the theoretical description of electron–molecule collisions—generates recombination at a much faster rate than any other known mechanism. Our estimated rate constant overlaps the range of values spanned by experiments. We treat the low-energy collision process as a curve-crossing problem, which was previously thought inapplicable to low-energy recombination in $\text{H}_3^+$. Our calculation reproduces the measured propensity for three-body versus two-body breakup of the neutral fragments, as well as the vibrational distribution of the $\text{H}_3^+$ product molecules.

Despite the key role of the $\text{H}_3^+$ ion in many astrophysical and chemical environments, the dissociative recombination (DR) of $\text{H}_3^+$ with a low-energy electron:

$$\text{H}_3^+ + e^- \rightarrow \text{H}_2 + \text{H} \quad \text{or} \quad \text{H} + \text{H} + \text{H}$$

has remained poorly understood despite experimental progress since the early 1990s. The high DR rate measured in laboratory storage-ring experiments seems incompatible with the large amount of $\text{H}_3^+$ observed in diffuse interstellar clouds. Yet one model suggests that the high storage-rate can be understood in terms of the ‘clumpy’ nature of the intervening interstellar clouds responsible for the absorption of infrared radiation at the $\text{H}_3^+$ resonance lines. Another problem is that flowing-afterglow Langmuir probe (FALP) measurements obtained a DR rate that is one order of magnitude smaller (or even less) than rates extracted from storage-ring experiments—and there is an orders-of-magnitude discrepancy between the storage-ring experiments and theoretical estimates. No satisfactory theoretical model has yet been able to predict a DR rate anywhere near as large as experiments measure for $\text{H}_3^+$ at sub-eV electron energies.

Here we develop a method that includes the non-perturbative Jahn–Teller coupling between the electronic and nuclear motions. Although a complete quantitative implementation of this method remains to be made, a preliminary implementation presented here allows us to understand at least the large value of the FALP experimental cross-section measurements. The theory also correctly predicts other details (within the approximations discussed below), such as the DR branching ratio ($\text{H} + \text{H} + \text{H}$)/($\text{H}_2 + \text{H}$), and the most probable vibrational quantum ($\nu \approx 5$) of the $\text{H}_2$ products.

We treat the ionic vibrational dynamics on the $\text{H}_3^+$ electronic potential surface within the adiabatic hyperspherical approximation. The hyperspherical radius $R$ is a useful collective coordinate for any few-body system: for three atoms of equal mass, the squared hyperradius is defined by $R^2 = \sqrt{3r_1^2 + r_2^2 + r_3^2}$, where $r_1$, $r_2$, and $r_3$ are the distances of each nucleus from the centre of mass. The adiabatic hyperspherical potential curves $U_i(R)$ for the $\text{H}_3^+$ ion are energy eigenvalues of the fixed-$R$ ionic nuclear motion Schrödinger equation:

$$U_i(R) = H_{\text{ionic}}(\Phi_{i}(\theta, \phi; R)) = U_i(R)\Phi_{i}(\theta, \phi; R)$$

Here $H_{\text{ionic}}$ is the vibrational Hamiltonian, $\theta$ and $\phi$ are two hyperangles, and $\Phi_{i}$ is the eigenfunction. The ionic potential curves are displayed in Fig. 1.

To describe the electron–ion interaction, a set of ab initio $\text{H}_3^+$ potential surfaces and their mutual couplings are first parameterized at fixed values of the nuclear positions using multichannel quantum defect theory (MQDT). It includes the Jahn–Teller coupling for all principal quantum numbers and the electronic continuum. The adiabatic hyperspherical potential curves of $\text{H}_3^+$ are then extracted by searching for autoionizing resonances of the $e$–$\text{H}_3^+$ system. Each potential curve $U(R)$ (that is, energy as a function of the hyperradius $R$) generally exhibits a nonzero autoionization width $\Gamma_i(R)$. The initial electron-capture process, followed by the subsequent two-body or three-body fragmentation, can then be described by the DR “curve-crossing formalism.”

Our treatment can be viewed as an application of the vibrational frame transformation of MQDT that has successfully treated the diatomic Rydberg state interactions, however, the role of the intermediate nuclear distance in diatomic systems is replaced here by the hyperradius $R$. The quantum-mechanical scattering amplitude for a Jahn–Teller-type symmetry-changing $p$-wave collision, with the nuclei clamped, will be written as $S_{\text{MT}}(R, \theta, \phi)$. Here $(R, \theta, \phi)$ are the nuclear coordinates, rewritten in terms of a hyperradius $R$ and two hyperangles $\theta, \phi$. In an earlier study of photoabsorption by a metastable Rydberg state of $\text{H}_2^+$, the vibrational frame transformation was employed to obtain the electron–ion scattering matrix, excluding dissociation channels, as a matrix element of $S_{\text{MT}}(R, \theta, \phi)$ connecting two rovibrational eigenfunctions of $\text{H}_3^+$ integrated over all nuclear coordinates $(R, \theta, \phi)$. The fixed-nuclei scattering matrix $S_{\text{MT}}(R, \theta, \phi)$ represents an amplitude for the reaction:

$$\text{e}(l\Lambda') + \text{H}_3^+(R, \theta, \phi) \rightarrow \text{e}(l\Lambda) + \text{H}_2^+(R, \theta, \phi)$$

where the orbital angular momentum $l = 1$. That is, the electron

---

**Figure 1** The first 200 adiabatic potential curves of $\text{H}_3^+$ in hyperspherical coordinates. The neglect of mutual coupling between these adiabatic channels is a good approximation here, as shown by the fact that the root-mean-square error of the lowest vibrational levels in the first seven potential curves is 32 cm$^{-1}$. This error is much smaller than the vibrational spacing of 2,500 cm$^{-1}$. Note that in the notation $\nu \Phi_{\nu}$, the lowest such potential curve contains the 00$^1$ vibrational level, and all of the symmetric-stretch vibrational excitations of that level. The energy is in atomic units.
scatters from one irreducible representation \( \Lambda' \) of the \( D_{3h} \) symmetry group into another, \( \Lambda \), while the nuclei do not have time to move.

For a \( p \)-wave electron, the \( p\sigma \)-component corresponds to the lone \( D_{3h} \) irreducible representation \( \Lambda = 1 \), while the \( p\pi \)-components \( \Lambda = \pm 1 \) are associated with the Jahn–Teller active modes of irreducible representation \( E \). The Jahn–Teller-induced coupling between the \( p\pi \)-components can be represented by a 2 \( \times \) 2 reaction matrix.

\[
K = \left( \begin{array}{cc}
\delta p^2 & \lambda \rho \exp(i\phi) \\
\lambda \rho \exp(-i\phi) & \delta p^2
\end{array} \right)
\]  

Here \((\rho, \phi)\) are radial and azimuthal coordinates of the doubly degenerate vibrational modes, and \((\lambda, \delta)\) are the Jahn–Teller coupling constants.\(^\dagger\)\(^\ddagger\). In equation (4), the constant, equilibrium geometry quantum defect \( \mu' = \mu_{\Lambda = \pm 1} \) is not shown. It is included into the treatment in the same way as in ref. 19. The Jahn–Teller and MQDT parameters \((\lambda, \delta, \mu_{\Lambda})\) are taken to be independent of the nuclear configuration, which tests show is a good approximation over the range of the vibrational ground state of \( \text{H}_2^+ \). We adopt Jahn–Teller parameters determined in ref. 28.

The short-range scattering amplitude for this process is now calculated by a generalized vibrational frame transformation\(^\ddagger\)\(^\dagger\) as an integral:

\[ S_{\Lambda', \Lambda}(\mathbf{r}) = \langle \Phi_{\Lambda'} | S_{\Lambda \Lambda}(\mathbf{r}, \theta, \phi) | \Phi_{\Lambda} \rangle \]  

Here the double brackets denote an integral over the hyperangles only, at a fixed \( \mathbf{r} \). The dimension of this \( S \)-matrix is \( 3N \times 3N \), where \( N \) is the number of adiabatic eigenstates of \( \text{H}_2^+ \) that are retained. A standard MQDT closed-channel ‘elimination’, plus a time-delay analysis, determines the resonance positions and partial widths at each hyperradius\(^\ddagger\)\(^\dagger\).

Our treatment differs from that of refs 18 and 19, however, in that we treat the quantization of the vibrating nuclei in two stages, by using hyperspherical coordinate techniques. We deliberately delay the solution of the Schrödinger equation in the hyperradial degree of freedom, in order to gain insight into the relevant DR pathways. This provides an immediate link with the theory of DR in diatomic molecules, with the role of the internuclear distance in diatomics generalized to the hyperradius in polyatomic species. In principle, this generalization will also permit the application of more sophisticated treatments developed for diatomics\(^\ddagger\)\(^\dagger\) that incorporate the competition between direct and indirect Rydberg state pathways.

In general, each \( \text{H}_2^+ \) adiabatic potential curve supports an infinite number of autoionizing Rydberg states, at every value of \( R \). The direct contributors to DR are the \( 2p \) components, which are the only ones that lead diabatically to dissociation. All other higher-excited Rydberg states are energetically closed dissociation channels in this energy range, but nevertheless they can contribute as indirect DR pathways. Once such a Rydberg state is excited during an autoionization process, the system can oscillate in \( R \) many times, and can then undergo a Landau–Zener transition into one of the \( 2p \) dissociative pathways that are open. The \( 2p \) symmetry has three different components (with \( \Lambda = 0, \pm 1 \)). The coupled \( 2p_1 \) and \( 2p_2 \) components form two diabatic potential surfaces of \( \text{H}_2^+ \) that correlate asymptotically either to \( \text{H}_2^+ + c \) or to \( \text{H}_1 + c + \text{H}_2 \). The \( 2p_3 \) state does not lead to a dissociation channel that is open in this sub-eV energy range. Yet an initially captured \( np_\ell \) electron can be scattered into the \( np_{\ell+1} \) states by Coriolis coupling, which gives another indirect pathway to DR although it is a minor contributor to the final rate.

Figure 2 shows how the dissociative recombination process occurs within our framework. To describe the DR process quantitatively, we use the cross-section expression\(^\ddagger\)\(^\dagger\):

\[ \sigma = \sum \frac{\pi^2}{\hbar^2} \Gamma_{\ell}(R) |U(R)|^2 \]  

Here the sum is over all resonant curves crossing the ground-state hyperradial potential curve for \( \text{H}_2^+ \), \( \Gamma_{\ell}(R) \) is the partial auto-

\[
\begin{array}{|c|c|c|c|c|c|}
\hline
\ell & \ell' & N_{\text{ion}} & R_0 (\text{a.u.}) & \Gamma_\ell (\text{cm}^{-1}) & |U|_\ell (\text{a.u.}) \\
\hline
2 & u & 53 & 1.920 & 0.18 & 0.325 \times 10^{-2} \\
3 & l & 12 & 1.951 & 0.10 & 0.187 \times 0.42 \\
4 & u & 5 & 1.957 & 0.15 & 0.138 \times 0.45 \\
4 & u & 6 & 1.967 & 0.60 & 0.133 \times 0.50 \\
5 & u & 5 & 2.097 & 0.74 & 0.076 \times 1.42 \\
5 & l & 4 & 2.091 & 0.61 & 0.078 \times 1.37 \\
5 & l & 5 & 2.107 & 0.86 & 0.073 \times 1.50 \\
5 & l & 6 & 2.108 & 0.62 & 0.073 \times 1.51 \\
6 & u & 5 & 2.245 & 0.96 & 0.065 \times 2.21 \\
6 & l & 4 & 2.234 & 0.58 & 0.072 \times 2.20 \\
6 & l & 5 & 2.233 & 1.03 & 0.048 \times 0.76 \\
6 & l & 6 & 2.485 & 0.50 & 0.043 \times 1.07 \\
6 & l & 6 & 2.477 & 0.51 & 0.043 \times 1.12 \\
\hline
\end{array}
\]  

\( \ell \) and \( \ell' \) mean upper and lower states of the Jahn–Teller couple \( \Lambda = \pm 1 \). Note that states \( \ell = 0 \), \( \ell = 1 \), \( \ell = 2 \) cross the horizontal line \( E_\ell = 0.025 \text{eV} \) at two different times. \( N_{\text{ion}} \) is the approximate principal quantum number of the resonance, while \( |U|_\ell \) is the eigenvalue index for the \( \text{H}_2^+ \) hyperradial potential curves, starting from the lowest energy state.
ionization width in the incident channel, $E_d$ is the energy of the capture electron in atomic units, $\Psi(R_0)$ is the normalized hyper-radial wavefunction of the lowest $H_3^+$ vibrational level, shown at the inset of Fig. 2, $R_d$ is the Condon point (that is, the hyperradius at which the 8th curve lies at an energy $E_d$ above the $H_3^+$ ground state energy), and $U'_\beta(R_0)$ is the derivative of the excited-state $H_3^+$ hyper-spherical potential curve into which capture occurs. For cases where capture occurs into a potential curve that has a minimum near the Condon point, we have implemented a modification to the approximate overlap integral that replaces its contributions in equation (6) by a coherent sum of the values from both Condon points at the lowest vibrational eigenenergy in that potential $(\pi^2/E_d)\sum \Psi(R_\alpha)|\Gamma(\beta,R_\alpha)U'_\beta(R_\alpha)|^2$. This modification eliminates the unphysical divergence in the cross-section that would otherwise result at those minima. We have verified that this approximation is reasonable. Table 1 shows the dominant contributing pathways at an incident electron energy of 0.025 eV, and also the leading 2p 'direct' contribution. Note that the largest resonance width obtained in our calculations is 1.8 times smaller than the largest width (2.8 cm$^{-1}$) due to vibrational autoionization that emerges from the theoretical calculations$^{18,19,28}$ and from experiment$^{18}$. Although this is not fully quantitative agreement, the comparison gives us some confidence that the electron–ion interaction is described at a reasonable level of approximation.

Figure 3 compares our calculated DR cross-section with the newest storage-ring experiment measurement$^7$. The branching ratio has also been calculated, but for the direct 2p pathways only; our theoretical results for the fraction of DR events that leads to three-body fragmentation is $\sigma(H + H + H)/\sigma_{DR} = 0.70 \pm 0.07$. A coupled-channel calculation that can track the recombination flux excited by indirect np pathways is beyond the scope of the present study. Consequently, our results only suggest what the final branching ratio might be after the indirect pathways are incorporated more satisfactorily. Nevertheless, it is at least encouraging that the theoretical branching ratio overlaps with the experimental result$^7$ of 0.75 $\pm$ 0.08. Moreover, this direct-pathways calculation results in a peak vibrational distribution for the channels $H_2^+$($\nu, j$) that peaks at $\nu = 5$–6, which is in accord with recent experimental results$^3$. Here Vand j are the vibrational and notational quantum numbers for $H_2^+$.

We have conducted two different calculations of the total recombination rate, in order to assess the role of indirect processes. The first is a 'lower-bound' cross-section that includes only the direct dissociation pathways (2p) in the summation of equation (6). The second is an 'upper-bound' cross-section that includes all pathways in equation (6). The latter calculation will be accurate if the coupling between 2p and np channels is strong enough for predissociation to occur before the excited state can autoionize—that is, during a few vibrations of the np states. This is plausible because the autoionization lifetimes are so long, of the order of thousands of vibrational periods, but it is far from proved. It will be important for future studies to test this assumption quantitatively by actually mapping out the detailed Landau–Zener crossings between indirect and direct DR pathways. Alternatively, an experiment could help to assess this key issue by measuring the branching ratio between predissociation and np Rydberg levels, for $n$ in the range 3–7. We stress that our findings are consistent with previous theoretical work$^{14,15,20}$ which has already suggested and documented a large enhancement of the DR rate through indirect Rydberg state pathways, by as much as two orders of magnitude.

For a thermal distribution of incident electron energies, the DR rate coefficient $\alpha = \langle np \rangle$ is defined as

$$\alpha(kT) = \frac{8\pi}{(2\pi kT)^{3/2}} \int_0^\infty \alpha(E) e^{-E/kT} dE$$

(7)

Table 2 compares our calculations with the results of three recent experiments and with a recent calculation (ref. 14). A more complete comparison with earlier experiments can be found in refs 1 and 8. The 300 K rate constant calculated by ref. 14 is close to our 'lower-bound' rate constant, but this agreement is fortuitous, because our lower-bound value includes only the contributions from the 2p 'direct' DR pathways, whereas the result quoted for ref. 14 includes an estimate for the indirect pathways. For that reason, the value from ref. 14 should be compared with our 'upper-bound' result in Table 2; the 'direct' DR rate$^{14}$ is another two orders of magnitude smaller.

Some questions still remain to be answered by future studies of the low-energy DR process. We have found that Jahn–Teller coupling is more important by far in the sub-eV energy range
Parasitic computing

Albert-László Barabási†, Vincent W. Freh‡, Hawoong Jeong* & Jay B. Brockman†

Reliable communication on the Internet is guaranteed by a standard set of protocols, used by all computers. Here we show that these protocols can be exploited to compute with the communication infrastructure, transforming the Internet into a distributed computer in which servers unwittingly perform computation on behalf of a remote node. In this model, which we call parasitic computing, one machine forces target computers to solve a piece of a complex computational problem merely by engaging them in standard communication. Consequently, the target computers are unaware that they have performed computation for the benefit of a commanding node. As experimental evidence of the principle of parasitic computing, we harness the power of several web servers across the globe, which—unknown to them—work together to solve an NP complete problem.

Sending a message through the Internet is a sophisticated process regulated by layers of complex protocols. For example, when a user selects a URL (uniform resource locator), requesting a web page, the browser opens a transmission control protocol (TCP) connection to a web server. It then issues a hyper-text transmission protocol (HTTP) request over the TCP connection. The TCP message is carried via the Internet protocol (IP), which might break the message into several packages, that navigate independently through numerous routers between source and destination. When an HTTP request reaches its target web server, a response is returned via the same TCP connection to the user’s browser. The original message is reconstructed through a series of consecutive steps, involving IP and TCP; it is finally interpreted at the HTTP level, eliciting the appropriate response (such as sending the requested web page). Thus, even a seemingly simple request for a web page involves a significant amount of computation in the network and at the computer endpoints. The success of the Internet rests on the cooperation and trust between all involved parties.

Received 4 April; accepted 4 July 2001.

Here we demonstrate that the trust-based relationships between machines connected on the Internet can be exploited to use the resources of multiple servers to solve a problem of interest without authorization. In essence, a ‘parasitic computer’ is a realization of an abstract machine for a distributed computer that is built upon standard Internet communication protocols. We use a parasitic computer to solve the well known NP-complete satisfiability problem, by engaging various web servers physically located in North America, Europe, and Asia, each of which unknowingly participated in the experiment. Like the SETI@home project (see http://www.seti.org), parasitic computing decomposes a complex problem into computations that can be evaluated independently and solved by computers connected to the Internet; unlike the SETI project, however, it does so without the knowledge of the participating servers. Unlike ‘cracking’ (breaking into a computer) or computer viruses, however, parasitic computing does not compromise the security of the targeted servers, and accesses only those parts of the servers that have been made explicitly available for Internet communication.

To solve many NP complete problems, such as the travelling salesman or the satisfiability problem, a common technique is to generate a large number of candidate solutions and then test the candidates for their adequacy. Because the candidate solutions can be tested in parallel, an effective computer architecture for these problems is one that supports simultaneous evaluation of many tests. An example of such a machine is illustrated in Fig. 1a. Here, the computer consists of a collection of target nodes connected to a network, where each of the target nodes contains an arithmetic and logic unit (ALU) that is capable of performing the desired test and a network interface (NIF) that allows the node to send and receive messages across the network. A single home parasite node initiates the computation, sends messages to the targets directing them to perform the tests, and tabulates the results.

Owing to the many layers of computation involved in receiving and interpreting a message, there are several Internet protocols that, in principle, could be exploited to perform a parasitic computation. For example, an IP-level interpretation could force routers to solve the problem, but such an implementation creates unwanted local bottlenecks. To truly delegate the computation to a remote target computer, we need to implement it at the TCP or higher levels. Potential candidate protocols include TCP, HTTP, or encryption/decryption with secure socket layer (SSL).

During package transfer across the Internet, messages can be corrupted, that is, the bits can change. TCP contains a checksum that provides some data integrity of the message. To achieve this, the sender computes a checksum and transmits that with the message. The receiver also computes a checksum, and if it does not agree with the sender’s, then the message was corrupted (see Fig. 2). One property of the TCP checksum function is that it forms a sufficient logical basis for implementing any Boolean logic function, and by extension, any arithmetic operation.

To implement a parasitic computer using the checksum function we need to design a special message that coerces a target server into performing the desired computation. As a test problem we choose to solve the well known ‘satisfiability’ (or SAT) problem, which is a common test for many unusual computation methods. The SAT problem involves finding a solution to a Boolean equation that satisfies a number of logical clauses. For example, \((x_1 \lor x_2 \land \neg x_3)\) AND \((\neg x_2 \land x_3)\) in principle has 2 potential solutions, but it is satisfied only by the solution \(x_1 = 1, x_2 = 0, x_3 = 1\). This is called a 2-SAT problem because each clause, shown in parentheses, involves two variables. The more difficult 3-SAT problem is known to be NP complete, which in practice means that there is no known polynomial-time algorithm which solves it. Indeed, the best known algorithm for an \(n\)-variable SAT problem scales exponentially with \(n\), following \((1.33)^n\) (ref. 6). Here we follow a brute-force approach by instructing target computers to evaluate, in a distributed fashion, each of the 2 potential solutions.

A possible parallel approach to the SAT problem is shown schematically in Fig. 1. The parasite node creates \(2^n\) specially constructed messages designed to evaluate a potential solution. The design of the message, exploiting the TCP checksum, is described in Fig. 3. These messages are sent to many target servers throughout the Internet. Note that while we choose the simpler 2-SAT case to illustrate the principle behind the coding, the method can be extended to code the NP complete 3-SAT problem as well, as explained in the Supplementary Information (see also http://www.nd.edu/~parasite). The message received by a target server contains an IP header, a TCP header, and a candidate solution (values for \(x_i\)). The operators in the Boolean equation determine the value of the checksum, which is in the TCP header. The parasite node injects each message into the network at the IP level (Fig. 1), bypassing TCP. After receiving the message, the target server verifies the data integrity of the TCP segment by calculating a TCP checksum and checks the received message against the candidate solution. The construction of the check sum is such that an invalid solution fails the TCP checksum and is dropped, which is illustrated by the black path labelled ‘failure’. Consequently, only valid solutions propagate up through the TCP protocol layer to HTTP, illustrated by the red path labelled ‘success’. Targeted web servers respond to all requests that reach HTTP, even invalid HTTP requests. Thus, the parasite node sends out a message for each possible solution (black arrow in a), but only receives a response for each solution that is valid (red arrow in a).

**Figure 1** Schematic diagram of our prototype parasitic computer. **a,** A single parasite node (green) coordinates the computations occurring remotely in the Internet protocols. It sends specially constructed messages to some number of targeted nodes (blue boxes), which are web servers consisting of an arithmetic and logic unit (ALU) and a network interface (NIF). **b,** Levels of communication between the parasitic node and the target in our proof-of-principle implementation. The packet carrying the problem to be solved is inserted into the network at the IP level, bypassing the parasitic node’s TCP.
The construction of the message (Fig. 3) ensures that the TCP checksum fails for all messages containing an invalid solution to the posed SAT problem. Thus, a message that passes the TCP checksum contains a correct solution. The target server will respond to each message it receives (even if it does not understand the request). As a result, all messages containing invalid solutions are dropped in the TCP layer. Only a message which encodes a valid solution ‘reaches’ the target server, which sends a response to the ‘request’ it received.

We have implemented the above scheme using as a parasitic node an ordinary desktop machine with TCP/IP networking. The targeted computers are various web servers physically located in North America, Europe, and Asia, each of which unwittingly participated in the experiment. As explained in Fig. 1, our parasite node distributed 2^N messages between the targets. Because only messages containing valid solutions to the SAT problem pass through TCP, the target web server received only valid solutions. This is interpreted as an HTTP request, but it is of course meaningless in this context. As required by HTTP, the target web server sends a response to the parasitic node, indicating that it did not understand the request. The parasite node interprets this response as attesting to the validity of the solution. As expected and by design, incorrect solutions do not generate responses for the web server. A typical message sent by the parasite, and a typical response from a target web server are included in the Supplementary Information.

Our technique does not receive a positive acknowledgement that a solution is invalid because an invalid solution is dropped by TCP. Consequently, there is a possibility of false negatives: cases in which a correct solution is not returned, which can occur for two reasons. First, the IP packet could be dropped, which might be due to data corruption or congestion. Normally TCP provides a reliability mechanism against such events, but our current implementation cannot take advantage of this. Second, because this technique exploits the TCP checksum, it circumvents the function the checksum provides. The TCP checksum catches errors that are not caught in the checks provided by the transport layer, such as errors in checksum and its complement are shown in the figure. The value of the checksum \( S \) calculated by the target is

\[
P = (x_0 \oplus x_1) (x_2 \oplus x_3) (x_4 \oplus x_5) (x_6 \oplus x_7) (x_8 \oplus x_9) (x_{10} \oplus x_{11}) (x_{12} \oplus x_{13}) (x_{14} \oplus x_{15})
\]

The logical table of XOR, AND, and the binary sum (+). In order to get a TRUE answer for \( P \), each clause shown in separate parentheses in \( a \) needs to be independently TRUE. \( b \). To evaluate the value of \( P \) we generate a 32-bit message \( M \) that contains all 16 variables, each preceded by a zero. As an illustration, we show a possible solution. \( c \). TCP groups the bits of the received message in two 16-bit segment and adds them together (Fig. 2a). As shown in \( d \), this will result in adding each \((x_i, x_{i+1})\) pair together where \( i \) is odd. The sum can have four outcomes. Comparing the sum with the \( e \) column in the table in \( b \), we notice that a TRUE answer for the XOR clause \((x_i \oplus x_{i+1})\) coincides with the (01) result of the \((x_i, x_{i+1})\) sum. Similarly, if the clause has an AND operator, \((x_i \land x_{i+1})\) is true only when the result is (10). This implies that for a set of variables \( \{x_i, x_{i+1}, \ldots, x_{15}\} \) that satisfies \( P \) the checksum will be determined by the corresponding operators only (that is, a \( 2 \) should give (01) for the sum check, and for \( \land \) the sum is (10)). For illustration, in \( d \) we show the formal lineup of the variables, while in \( b \) we show an explicit example. The correct complemented checksum for \( E \) should be \( S = 1011010101010101 \). In contrast, the parasitic computer places in the header the transmitted checksum \( T = 1001101001010101 \), which is uniquely determined by the operators in \( P \), as shown in \( d \). To turn the package into a parasitic message the parasitic node prepares a package, shown in \( f \), preceded by a checksum \( T_e \) and continued by a 32-bit sequence \( \{S_1, S_2\} \), which represent one of the 2^16 potential solutions. If \( S_1 \) and \( S_2 \) do not represent the correct solution, then the checksum evaluated by the target TCP will not give the correct sum (111 ... 1). The TCP layer at the target concludes that the message has been corrupted, and drops it. However, if \( S_1 \) and \( S_2 \) contain the valid solution, the message is sent to HTTP. The web server interprets the solution as an HTTP request; however, because it is not a valid HTTP request, the web server responds with a message saying something like “page not found” (red arrow in Fig. 1a). Thus, every message to which the parasite node receives a response is a solution to the posed SAT problem (see Fig. 2d).
intermediate routers and the end points. Measurements show that the TCP checksum fails in about 1 in $2^{10}$ messages. The actual number of TCP checksum failures depends on the communication path, message data, and other factors. To test the reliability of our scheme, we repeatedly sent the correct solution to several host computers located on three continents. The rate of false negatives with our system ranged from 1 in about 100 to less than 1 in 17,000.

The implementation offered above represents only a proof of concept of parasitic computing. As such, our solution merely serves to illustrate the idea behind parasitic computing, and it is not efficient for practical purposes in its current form. Indeed, the TCP checksum provides a series of additions and a comparison at the cost of hundreds of machine cycles to send and receive messages, which makes it computationally inefficient. To make the model viable, the computation-to-communication ratio must increase until the computation exported by the parasitic node is larger than the amount of cycles required by the node to solve the problem itself instead of sending it to the target. However, we emphasize that these are drawbacks of the presented implementation and do not represent fundamental obstacles for parasitic computing. It remains to be seen, however, whether a high-level implementation of a parasitic computer, perhaps exploiting HTTP or encryption/decryption, could execute in an efficient manner.

It is important to note that parasitic computing is conceptually related but philosophically different for cluster computing, which links computers such that their cumulative power offers computational environments comparable to the best supercomputers. A prominent example of cluster computing is the SETI program, which has so far enlisted over 2.9 million computers to analyze radio signals in search of extraterrestrial intelligence. In cluster computing, the computer’s owner willingly downloads and executes software, which turns his computer into a node of a vast distributed computer. Thus, a crucial requirement of all cluster computing models is the cooperation of the computer’s owner. This is also one of its main limitations, as only a tiny fraction of computer owners choose to participate in such computations. In this respect, parasitic computing represents an ethically challenging alternative for cluster computing, as it uses resources without the consent of the computer’s owner.

Although parasitic computing does not compromise the security of the target, it could delay the services the target computer normally performs, which would be similar to a denial-of-service attack, disrupting Internet service. Thus, parasitic computing raises interesting ethical and legal questions regarding the use of a remote host without consent, challenging us to think about the ownership of resources made available on the Internet. Because parasitic computation exploits basic Internet protocols, it is technically impossible to stop a user from launching it. For example, changing or disrupting the functions that are exploited by parasitic computing would simply eliminate the target’s ability to communicate with the rest of the Internet.

In summary, parasitic computing moves computation onto what is logically the communication infrastructure of the Internet, blurring the distinction between computing and communication. We have shown that the current Internet infrastructure permits one computer to instruct other computers to perform computational tasks that are beyond the target’s immediate scope. Enabling all computers to swap information and services they are needed could lead to unparalleled emergent behaviour, drastically altering the current use of the Internet.
observed microcolumns form along a single primary crack and have one free and one fixed end rather than forming between the wing cracks of two adjacent primary cracks and having two fixed ends. We suggest that the onset of the localization of failure, which limits compressive strength, is triggered by the bending-induced failure of these fixed-free microcolumns. The microcolumns bend and fail owing to frictional drag along their free ends, caused by sliding along the primary crack, rather like the breaking of teeth in a comb under a sliding thumb. Once triggered, we assume that the incipient zone of localized damage is sufficient to focus additional microcracking ahead of itself in the manner of a process zone 7, although we have not attempted to quantitatively model this aspect of the failure process. As the zone advances, it creates, we imagine, a discrete band of intense damage inclined by \( \sim 30\degree \) to the direction of highest compressive stress that becomes the macroscopic fault.

To quantify this process, consider an optimally oriented primary crack inclined at 45\degree to the most compressive stress \( \sigma_1 \) (Fig. 1). As compression increases, frictional sliding along the primary crack will occur if the confinement ratio \( R = \sigma_1/\sigma_1' < (1 - \mu)/(1 + \mu) \), where \( \mu \) is the coefficient of friction along the sliding segment of the crack and \( \sigma_1' \) is the least compressive principal stress. Higher ratios suppress frictional sliding and lead to ductile failure. Even when \( R \) is less than this value, ductile behaviour will occur, we suggest, when the applied strain rate is sufficiently low to allow creep deformation to relax stress concentrations along the primary crack. Then microcolumns no longer form and thus are not available to trigger the localization of damage.

Sliding along the primary crack generates a stress concentration at the end of the sliding segment. This stress concentration results in a region around the sliding segment tip where the stress exceeds the yield stress of the material, resulting in inelastic deformation. If the applied strain rate is high enough, creep deformation will not have sufficient time to relax the stress concentration and a secondary crack will initiate. Secondary cracks do not initiate at lower strain rates because the stress concentration is relaxed by creep. Creep relaxation is governed by a power-law relationship \( \dot{\varepsilon} = B\sigma^n \), where \( \dot{\varepsilon} \) is the creep rate caused by the driving stress \( \sigma \), \( B \) is a temperature and material dependent constant, and \( n \) is a material constant. We suggest that the brittle–ductile transition occurs when the size of the region of creep deformation just equals the size of the region where the stress exceeds the yield stress of the material. The creep zone size is estimated from the Riedel–Rice model of creep zones around non-interacting cracks 8 and the elastic zone estimated using Irwin’s approximation 10. Accordingly, the brittle–ductile transitional strain rate \( \dot{\varepsilon}_t \) can be expressed in non-dimensional form as

\[
\dot{\varepsilon}_t = \frac{\dot{\varepsilon}_0 c^{1.5}}{B K_1 c} \approx \frac{8}{\pi[(1 - R) - \mu(1 + R)]} \quad \text{where} \quad R < \frac{1 - \mu}{1 + \mu}
\]

Here \( 2c \) is the primary crack size, which in laboratory samples is set by the dimensions of the grain boundary facets, and \( K_{IC} \) is the critical stress intensity factor (Table 1). Details on the derivation of this expression are given in Part A of the Supplementary Information. Implicit in the formulation is \( n = 3 \), a commonly observed value. Figure 2 demonstrates the excellent agreement between the predictions of equation (1) and the deformation behaviour versus confinement ratio of laboratory specimens of numerous crystalline solids (ice, rock and silicate minerals). If the variation of the parameters \( B \) and \( K_{IC} \) as a function of temperature are known, a similar prediction of the transitional strain rate as a function of temperature is possible. For example, between \(-40\degree C \) and \(0\degree C \), the parameter \( B \) for freshwater ice increases by a factor of about 200 (ref. 11), while \( K_{IC} \) decreases by a factor of two 12. The predicted increase in transitional strain rate is in reasonable agreement with the order of magnitude increase observed in the laboratory 11. Unfortunately, the data are insufficient to similarly test model predictions for rock.

Next we consider the brittle compressive strength. Equation (1) defines the limits of the confinement ratio and strain rate required for the initiation of secondary cracks, which we consider to be the critical step in the sequence of events leading to brittle failure. Once secondary cracks exist, terminal failure is governed by the load required to trigger the bending-induced failure of the slender microcolumns created between them. Failure of a microcolumn occurs when the stress concentration at the tip of a secondary crack results in its out-of-plane extension. The initiation of out-of-plane extension is related to the compressive loading along the column, the net bending moment imparted by frictional drag along the free end of the microcolumn, and the microcolumn geometry. In previous work we have suggested that this situation is similar to the out-of-plane extension of an edge crack in a brittle plate 13. Here we extend this earlier model to include the effects of friction between columns. In this case the failure stress \( \sigma_f \) can be approximated, in dimensionless form, as

\[
\sigma_f = \frac{\sigma_1' \sqrt{\dot{\varepsilon}_t}}{K_{IC}} = \frac{2}{\left(1 + \left(1 - \mu \frac{1 + R}{1 - R}\right)^{0.5}\right)^{0.5}} \left(1 + 3 \mu \alpha \dot{\varepsilon}(1 - R)^{3/2}\right)
\]
where \( \alpha = h/w \) is the aspect, or slenderness, ratio of the microcolumns. The more slender the microcolumns, the weaker the material. A complete derivation of this expression is provided in Part B of the Supplementary Information.

The geometry of the microcolumns depends on the length and spacing of the secondary cracks. For \( R > 0 \), the extension of the secondary cracks is stable, with the length asymptotically approaching the limiting value

\[
\frac{h}{c} = \frac{1 - R - \mu(1 + R)}{4.3R}
\]

(3)

Experimental data for ice\(^{14}\) and rock\(^{15}\) suggest that axial splitting is suppressed when the length of the secondary cracks is less than the size of the primary crack \( h < 2c \), implying that \( R \approx 0.05 \) for shear faulting.

The column width is set by the spacing of the secondary cracks. This, we suggest, is related to the extent of the shear stress concentration just ahead of a sliding segment along the primary crack. The sliding segment of the primary crack is defined by the region of the primary crack that has already separated (fully decohered) and along which sliding is governed by friction. Ahead of the sliding segment, the primary crack has not separated, but sliding may occur by a viscous process that results in partial decohesion\(^{16}\). The rate of viscous sliding and decohesion immediately ahead of the sliding segment is accelerated because the shear stress is intensified there. As a result, non-uniform sliding occurs along the primary crack which induces tensile stress within material on one side of the boundary\(^{17}\) and results in the nucleation of a secondary crack if the applied strain rate is above the brittle–ductile transition. At the same time, the accelerated viscous sliding ahead of the sliding segment weakens the primary crack over a distance equal to the extent of the zone of shear stress concentration, eventually resulting in the extension of the sliding segment into this region. This process repeats itself, with the sliding segment extending, nucleating a secondary crack, and then extending again, culminating in a set of secondary cracks. Their spacing is controlled by the extent of the zone of shear stress concentration ahead of the sliding segment. Here we approximate the boundary of this zone as the location where the magnitude of the gradient of the normalized shear stress induced by a mode II crack\(^{18}\) exceeds unity, yielding

\[
\frac{w}{c} = \left[ 1 + \left(1 - \frac{1 + R}{1 - R}\right)^{\frac{1}{2}} \right] - 1
\]

(4)

The normalizing shear stress is the shear stress that would exist in the absence of any cracks. The complete derivation is provided in Part B of the Supplementary Information.

Figure 3 compares the brittle compressive strength predicted by the above model to the experimental measurements from ice and a variety of rocks deformed over a range of confinement. The

Figure 2 Deformation style as a function of dimensionless strain rate and confinement ratio. Black symbols indicate ductile behaviour, white symbols indicate brittle behaviour. The curve shows the predicted transition assuming \( \mu = 0.5 \). The temperatures at which these data were collected are shown in Table 1. Data were taken from refs 15, 20–25, 28 and 30.

Figure 3 Brittle compressive strength of ice and various rock types as a function of confinement ratio. The curve shows the predicted compressive strength assuming \( \mu = 0.5 \). The temperatures at which these data were collected are shown in Table 1. Data were taken from refs 14, 15 and 25–30.
secondary, or ‘comb’, crack model fits the data quite well over the regime of moderate confinement for which it was created, consistently matching the lower bound. Comparison of our model with previous models is difficult because of the use by other investigators of fitting parameters. For example, failure strengths predicted by the model of Horii and Nemat-Nasser are generally a factor of two to three greater than those predicted by our model. However, their model has three free parameters that describe the length and spacing distributions of the primary cracks at failure. Similarly, the models of Peng and Johnson and of Bazant and Xiang use the aspect ratio of microcolumns created by wing cracks emanating from adjacent primary cracks as a fitting parameter. None of these parameters is easily predicted a priori. If the microcolumn aspect ratio predicted by equations (3) and (4) (which is consistent with laboratory observations) is used in the models of Peng and Johnson and Bazant and Xiang, which assume that microcolumn failure occurs via elastic buckling, the predicted compressive strengths are two to three orders of magnitude greater than measured values.

The agreement between prediction and measurement (Figs 2 and 3) provides support for the major role of secondary cracks in brittle compressive failure. The confinement ratio and strain rate required to initiate secondary cracks control the brittle–ductile transition, whereas the bending-induced failure of the microcolumns created between secondary cracks controls strength. In addition, the agreement between theory and experiment for a variety of materials suggests that micro-mechanical processes similar to those described here are active in a variety of crystalline solids. The new model, therefore, should have broad applicability.

Received 20 December 2000; accepted 9 July 2001.


Supplementary information is available on Nature's World-Wide Web site (http://www.nature.com/) or as paper copy from the London editorial office of Nature.

Acknowledgements

This work was supported by the US National Science Foundation, the Army Research Office, and the Office of Naval Research.

Correspondence and requests for materials should be addressed to C.E.R. (e-mail: carl.renshaw@dartmouth.edu) or to E.M.S. (e-mail: erland.schulson@dartmouth.edu).

Mount Etna lies near the boundary between two regions that exhibit significantly different types of volcanism. To the north, volcanism in the Aeolian island arc is thought to be related to subduction of the Ionian lithosphere. On Sicily itself, however, no chemical or seismological evidence of subduction-related volcanism exists, and so it is thought that the volcanism—including that on Mount Etna itself—stems from the upwelling of mantle material, associated with various surface tectonic processes. But the paucity of geological evidence regarding the primary composition of magma from Mount Etna means that its source characteristics remain controversial. Here we characterize the trace-element composition of a series of lavas emitted by Mount Etna over the past 500 kyr and preserved as melt inclusions inside olivine phenocrysts. We show that the compositional change in primary magmas from Mount Etna reflects a progressive transition from a predominantly mantle-plume source to one with a greater contribution from island-arc (subduction-related) basalt. We suggest that this is associated with southward migration of the Ionian slab, which is becoming juxtaposed with a mantle plume beneath Sicily. This implies that the volcanism of Mount Etna has become more calc-alkaline, and hence more explosive, during its evolution.

Volcanoes in eastern Sicily (Mount Etna and the volcanosedimentary Iblean plateau) are situated on the suture between the converging European and African plates, adjacent to the subduction-related Aeolian island arc (Fig. 1). There is, however, no clear evidence that this volcanism is directly connected with active subduction beneath the Aeolian arc; that is, there is no subduction-
related seismic trace beneath Sicily. It has therefore been proposed that Sicily is not located directly above the active subduction zone but at the lateral edge. Volcanic rocks of the Iblean Mountains range from quartz-tholeiites to undersaturated alkaline lavas, and volcanism at Mount Etna is composed for the most part of intermediate alkaline lavas. On the other hand, a wide range of rock types occurs within the nearby Aeolian islands, from typical island-arc calc-alkaline series to shoshonitic potassic lavas. The coexistence of alkaline-type and calc-alkaline-type volcanism places a number of constraints on any geodynamic picture of the southern Tyrrhenian Sea. It has for example been proposed that the Aeolian calc-alkaline series is associated with subduction of Ionian lithosphere, whereas alkaline-type volcanism in eastern Sicily is related to extensional tectonics of a corresponding backarc marginal basin (see, for example, ref. 1). Alternatively, Mount Etna has been recently attributed to sideways suction of asthenospheric material from beneath the African plate, induced by rollback motion of the Ionian slab.

The samples in which the primitive olivine-hosted melt inclusions were analysed come from Vulcano island (in the Aeolian island arc), the Iblean Mountains and Mount Etna (Fig. 1). Etnan samples represent different periods of volcanism. They include tholeiitic hyaloclastites from the Aci Castello series, which represents samples of the oldest Etnan volcanics so far identified (520 ± 40 kyr before present, ref. 6), basalts from the oldest part of the Etnan alkaline sequence (Santa Venera deposits dated at 120 kyr ago), alkali basalts from Mount Maletto (the most primitive of all Etnan holocene lavas), which represent recent (5 kyr BP) eccentric flank eruptions, and present-day (February 1999 eruption) hawaiite scoria from the southeast crater on the summit of Mount Etna. Basanite samples collected in the breccia deposits from Villasmundo (Iblean Mountains) have been dated to about 7.38 Myr ago (ref. 12) and high-K basaltic samples from a scoria cone on La Sommata (Vulcano island) were erupted ~50 kyr ago (ref. 13). Except for the 1999 eruption of the southeast crater on the summit of Mount Etna, the selected samples are the most primitive (in terms of MgO content) basalts and contain highly magnesian olivine crystals (85–91% forsterite; Fo95–91) as the main phenocryst type.

The major- and trace-element compositions of melt inclusions (after homogenization or correction for post-entrapment crystallization, see Methods) are given in Supplementary Information. The inclusions range in composition from nepheline-normative alkaline basalt to highly tholeiitic transitional basalt. They show a large range of concentrations for light rare-earth elements and high-field-strength elements (HFSE), suggesting a large range of parental magma compositions (Fig. 2). However, correlations between ratios involving highly incompatible elements (with bulk solid/melt partition coefficients, $D \ll$ melt fraction; Fig. 3) indicate that the trace-element variations among melt inclusions reflect a mixing relationship between two mantle sources, where one end-member is a mantle component represented by the Iblean Mountains and the other is a mantle component represented by Vulcano island.

We now consider these two mantle components (defined by the mixing arrays of Fig. 3) in more detail. Olivine-hosted melt inclusions from Vulcano have high concentrations of large-ion-lithophile elements, high ratios of light-rare-earth elements to heavy-rare-earth elements, and relatively low concentrations of HFSE (Fig. 2). Such trace-element features are typical of arc basalts, and are thought to reflect melting of mantle sources that have been metasomatized by fluid and sediment components derived from the downgoing slab. In contrast, melt inclusions from the Iblean Mountains lack the negative HFSE anomalies considered to be evidence for a subduction-related origin. They require the existence of a mantle source enriched in incompatible elements relative to normal mid-ocean-ridge basalts (N-MORB). Their trace-element compositions are remarkably similar to the ocean island basalts (OIB) showing HIMU-type isotopic signatures (very high Pb and relatively low Sr isotope ratios) (Figs 2 and 3); that is, they display a continuous decrease in the normalized abundances of Ba, K and Rb, and enrichments of Nb relative to similarly incompatible elements. The involvement of a HIMU-type mantle component in the sources of basalts from the Iblean Mountains and Mount Etna is also confirmed by recent Pb–Sr–Nd–Hf isotopic data. HIMU-type basalts are generally associated with mantle plumes, and their trace-element and isotopic features are thought to be derived from a source that may incorporate large segments of subduction-zone-processed recycled oceanic crust (see, for example, ref. 15).

The characterization of a HIMU-type component, which is generally attributed to deep-rooted mantle plumes, in the source of early Etnan eruptions of Aci Castello gives additional constraints on the origin of Mount Etna. In particular, involvement of a plume-type mantle component cannot be reconciled with shallow-level extensional tectonics (see ref. 17 and references therein). In contrast, it supports the hypothesis that early Etnan emissions (and the volcano-sedimentary Iblean plateau) are caused by a mantle plume located in a globally convergent geodynamical setting (see, for example, ref. 5). Note, however, that this model explains how plume-type and arc-type mantle sources can co-exist at the scale of the southern Tyrrhenian–Sicilian region, but does not clarify their interaction (as illustrated by the mixing trends of Fig. 3).

Volcanic activity of Mount Etna is characterized by a time-stratigraphic progression from primitive lavas with tholeiitic affinities (for example, the submarine Aci Castello series) to more alkaline rocks (including basalts from Santa Venera and Mount Maletto). This could be a similar transition to that observed on Hawaiian volcanoes. However, the tholeiitic-alkaline transition in Hawaii is marked by a gradual decrease in the degree of melting of a relatively uniform source (see, for example, ref. 18). In contrast, the different trace-element signatures for olivine-hosted melt inclusions in Aci Castello, Santa Venera and Mount Maletto basalts (Fig. 2) point towards a change of source material during the
magmaic evolution of Mount Etna. Melt inclusions from Aci Castello and Santa Venera display the trace-element characteristics of plume-derived HIMU basalts, whereas melt inclusions from Mount Maletto have trace-element compositions close to the primitive melt inclusions from subduction-related Vulcano island. Moreover, the mixing trends of Fig. 3 show that Mount Maletto melt inclusions represent liquids between typical OIB-type and island-arc-basalt type primitive magmas, represented respectively by melt inclusions from the Iblean Mountains and Vulcano island.

These observations provide strong evidence that the magma source beneath Mount Etna was originally plume-derived and has progressively become more arc-like. The melt inclusion data presented here therefore have implications for subduction-related magmatism—in particular, they are relevant to the debate about a possible contribution of plume to island-arc lavas (see, for example, refs 19, 20). Not only does the plume–arc transition process recorded by the Etnean primitive melt inclusions represent, to our knowledge, the first documentation of such a source evolution beneath a single volcano, but the respective signatures of Santa Venera and Mount Maletto melt inclusions indicate that the evolution has occurred in less than 100 kyr. It should also be noted that, at first sight, trace-element data for Mount Maletto melt inclusions can be accounted for by melting an OIB-like mantle wedge, provided that a HFSE-rich phase is residual during partial melting21. But we consider this hypothesis unlikely, because Nb remains a highly incompatible element (that is, the bulk $D$ for Nb is equivalent to that of Ce, see Fig. 2 legend) during the melting process recorded by the melt inclusions. Therefore, the low Nb concentrations in the Mount Maletto melt inclusions do not reflect the presence of HFSE-rich residual phases in an enriched mantle wedge, but rather the low abundance of Nb in the mantle sources of the melts.

The recent (post-1970) magmatic activity of Mount Etna is characterized by a higher frequency of eruption and higher effusion rates22. Moreover, the composition of the lavas shows anomalous enrichments in alkalies23 with an accompanying increase of $^{87}$Sr/$^{86}$Sr ratios (see, for example, ref. 10). Olivine (Fo 81)-hosted melt inclusions in hawaiite scoria of the 4 February 1999 eruption from the

Figure 2 Trace-element diagrams. Olivine-hosted melt inclusions in Etna basalts belonging to suites of different age are compared with olivine-hosted melt inclusions in basalts from other volcanoes of the southern Tyrrhenian–Sicilian region (Vulcano island and the Iblean Mountains). a, Representative trace-element patterns for melt inclusions from Mount Etna (Aci Castello, Santa Venera, Mount Maletto and the 1999 southeast crater series), Vulcano island and the Iblean Mountains. The order of element corresponds to the incompatibility sequence (taken as the order of their bulk solid/melt partition coefficients), defined by considering the element concentrations in the magmatic series and using standard deviations of the element concentration averages28. Compositions are normalized to the primitive mantle composition of ref. 29. Also shown for comparison is the average N-MORB (normal mid-ocean-ridge basalt) composition from ref. 29.

b, Ranges of trace-element compositions (normalized as in a) for olivine-hosted melt inclusions in basalts from Mount Etna (Aci Castello, Santa Venera, Mount Maletto and the 1999 southeast crater series), Vulcano island and the Iblean Mountains.
southeast crater\textsuperscript{11} were then analysed for comparison. When plotted in the diagrams of Figs 2 and 3, the 1999 melt inclusions have strong similarities with Mount Maletto melt inclusions; that is, their trace-element patterns and highly incompatible element ratios are intermediate between Vulcano and Iblean Mountains melt inclusions. This suggests that the plume–arc interaction process recorded by the olivine-hosted primitive melt inclusions from Mount Maletto basalt can also account for the geochemical peculiarities of the present-day Etna lavas. Therefore, the K and Rb enrichments and the gradual \( ^{87}\text{Sr}/^{86}\text{Sr} \) increase in Etna historical lavas could simply reflect increasing participation of the arc-source component. This hypothesis is consistent with the high H\textsubscript{2}O and S concentrations of present-day Etna lavas, which are more akin to subduction-related volcanoes than to oceanic hotspot volcanoes\textsuperscript{22}.

It is important to relate the geochemical evidence for a general transition from plume-type to arc-type mantle sources beneath Mount Etna to the regional setting. A simple model would be that subduction-related magmas originate from the nearby Aeolian island arc and migrate along tectonic features, such as the northwestern–southeastern Tindari–Giardini–Etna–Malta fault system (A–A in Fig. 1; see also ref. 25) to interact with Etna plume-related magmas. This hypothesis, however, is weakened by the observation that, to our knowledge, Etna lavas with plume signature have only been identified in the earlier succession. An alternative hypothesis is that the southwestern lateral limit of the Ionian slab beneath Mount Etna is progressively shifting. This could juxtapose subduction zone and mantle plume, allowing their respective melts to provide the mixed-source magmas. Finally, the plume–arc interaction process beneath Mount Etna may explain the characteristic features of its recent volcanic activity when compared with a typical hotspot volcano. In particular, Mount Etna displays an unusual eruptive style, characterized by frequent Strombolian explosions sometimes culminating in lava fountains (see, for example, ref. 26), and abundant gas emissions. Because subduction-related volcanoes are highly explosive, understanding the triggering mechanism of the plume–arc transition process identified here (and anticipating its evolution) is thus important when assessing the nature of future explosive events at Mount Etna.

**Methods**

**Melt inclusions**

Because olivine-hosted melt inclusions in hyaloclastites (from Aci Castello) and pyroclastic scoria (from Vulcano island and the southeast crater on the summit of Mount Etna) were naturally quenched during eruption, post-entrapment reactions that occurred inside the inclusions are restricted to slight precipitation of host phases on inclusion walls. Correction for this post-entrapment crystallization was made by incrementally adding olivine in Fe/Mg equilibrium until the liquidus olivine is identical to the host olivine. In all cases the amount of host olivine added to the glass inclusions was less than 6% by weight. Note, however, that the extent of the correction is not critical for the results, as this procedure simply induces a dilution by a constant factor of all the incompatible trace elements discussed here—so without correction, the relative trace-element patterns of the melt inclusions would be unchanged. For olivine-hosted melt inclusions in effusive flows from Mount Maletto, Santa Venera, and the Iblean Mountains, cooling rates were slow enough to induce both precipitation of the host olivine at the inclusion walls and nucleation of additional minerals inside the inclusions. The original melt compositions were produced in the glass samples after heating each sample up to the temperature at which the inclusion content homogenizes to an uniform single melt phase.

**Analytical techniques**

Major-element techniques of homogenized and unheated glass inclusions were measured with a Cameca SX 50 electron microprobe at the Cermics Center (University of Paris VI), using an accelerating voltage of 15 keV, sample current of 10 nA and a defocused beam.

---

**Figure 3** Plots of various ratios involving highly incompatible elements. Melt inclusions in olivine phenocrysts in dated basalts representative of the major episodes of Mount Etna volcanism are compared with melt inclusions in olivine phenocrysts in basalts from other volcanoes of the southern Tyrrenian–Sicilian region. Also shown for comparison is the estimated\textsuperscript{23} composition of the sources of HIMU (see text), EM1 and EM2-type (the two other isotopically extreme signatures) OIB, and the estimated\textsuperscript{23} average N-MORB composition.
beams, and trace-element compositions were measured with a Cameca IMS 4f ion microprobe at CNR-CSCC (Pavia, Italy), following the techniques described in ref. 27. Moreover, Sr and Cr were corrected for interferences by 26Si and 49Ti, at mass 45 and 52 (a.m.u.), respectively.

Received 5 February; accepted 5 July 2001.


Supplementary information is available on Nature’s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of Nature.

Genetic linkage of ecological specialization and reproductive isolation in pea aphids

David J. Hawthorne* & Sara Via†

*Department of Entomology; and †Department of Biology, University of Maryland, College Park, Maryland 20742, USA

The evolution of ecological specialization generates biological diversity and may lead to speciation4,5. Genetic architecture can either speed or retard this process. If resource use and mate choice have a common genetic basis through pleiotropy or close linkage, the resulting genetic correlations can promote the joint evolution of specialization and reproductive isolation, facilitating speciation6,7. Here we present a model of the role of genetic correlations in specialization and speciation, and test it by analysing the genetic architecture of key traits in two highly specialized host races of the pea aphid (Acyrthosiphon pisum pisum; Hemiptera: Aphididae). We found several complexes of pleiotropic or closely linked quantitative trait loci (QTL) that affect key traits in ways that would promote speciation: QTL with antagonistic effects on performance on the two hosts are linked to QTL that produce assortative mating (through habitat choice). This type of genetic architecture may be common in taxa that have speciated under divergent natural selection.

Natural selection is increasingly recognized as the main cause of speciation, yet the detailed ecological and genetic processes remain obscure2,8. To understand how adaptation to different environments drives speciation we need to re-examine the extent and mechanism of genetic correlations among characters within and across environments9,10. Genetic correlations (owing to pleiotropy or linkage disequilibrium) can either facilitate or constrain adaptive evolution10,11. For example, theory suggests that genetic correlations between resource use and mate choice can facilitate both specialization and speciation under divergent selection12,13. However, measures of genetic correlations among key traits have not been much used in previous empirical studies of the genetics of speciation, which have focused primarily on mechanisms of hybrid sterility and inviability12,13. Consequently, we lack empirical evidence of the genetic architectures that theory has identified as favourable for speciation.

Two classes of genetic correlations are particularly important for speciation under divergent selection. First, fitness trade-offs (negative genetic correlations between traits that are under divergent selection in different environments) suggest one mechanism for ecological specialization11,12,13,14. Secondly, positive genetic correlations between such fitness traits within environments and the characters that affect mate choice (including choice of mating habitat) can produce assortative mating among individuals specialized to the same environment, promoting reproductive isolation4,14. We integrated the genetic correlations of traits affecting specialization and mate choice into a single network. This revealed the potential for a set of complementary correlated responses to divergent selection that could facilitate specialization and reproductive isolation, even if the values of the individual genetic correlations are modest (Fig. 1a). Because every individual carries genes that affect traits in both environments, selection on resource use within environments produces correlated responses in traits that would be expressed in the alternative environment14. These correlated responses combine to increase the divergence and isolation of populations. Further direct and correlated responses are added if divergent selection also acts directly on the traits that lead
Acyrthosiphon pisum is a species of aphid that feeds on alfalfa and red clover (both in the subspecies A. p. pisum) in different environments, regardless of whether divergence takes place in sympatry or allopatry. Positive genetic correlations between resource use and mate choice (dotted boxes) speed up the evolution of specialization and cause reproductive isolation between divergent populations. Correlations between performance and mate choice across environments (solid box) indicated next to arrows facilitate ecological specialization and reproductive isolation. A negative genetic correlation between performance in two environments can align the two maps (‘Codom XXX’ on Fig. 2). For these four traits, the identified QTL together explain 10–57% of the genetic variance, which suggests that other QTL of small effect are also involved in both ecological specialization and habitat acceptance in pea aphids.

To probe the effects of individual chromosome segments on ecological specialization and assortative mating, and to distinguish the causes of the observed genetic correlations in Fig. 1b, we used quantitative trait loci (QTL) analysis. We examined QTL for ecological specialization (fecundity on the two hosts) and mate choice (host acceptance). We asked whether these QTL are in close enough proximity on the linkage map and of the appropriate directionality to suggest that they could contribute to fundamental genetic correlations.

The pea aphid linkage map (Fig. 2) contains four linkage groups, which is consistent with cytological observations of four chromosomes in pea aphids. We found QTL on all four linkage groups (Fig. 2). For these four traits, the identified QTL together explain only 10–57% of the genetic variance, which suggests that other QTL of small effect are also involved in both ecological specialization and habitat acceptance in pea aphids.

On linkage groups I and II, we found four complexes of QTL that affect several of the four key traits and are close enough together to indicate either pleiotropy or linkage. Two of these QTL complexes (W and Y, Fig. 2) are found only on the chromosome from the alfalfa parent, whereas the other two complexes (X and Z) involve chromosomes from both parents. We identify QTL in complexes X and Z as potentially homologous in the two genomes on the basis of their positions relative to the seven codominant markers used to align the two maps (‘Codom XXX’ on Fig. 2). The QTL in these complexes all have positive effects on characters expressed on the ‘native host’, that is, QTL on Ia (the alfalfa homologue for linkage group II) increase fecundity/acceptance of alfalfa (Fig. 2); or they have negative effects on characters expressed on the alternate host, that is, the QTL on Ia decrease fecundity on clover. Moreover, the directionality of effects of the closely adjacent QTL are completely consistent with the pattern of correlations shown in Fig. 1. Within each genome, QTL have antagonistic effects on fecundity in the two environments or on fecundity on one host and acceptance of the other. In contrast, the adjacent QTL affect fecundity and acceptance of the same host in the same direction. The probability of such coordinated directionality occurring by chance alone is $P < 0.00006$ (ref. 22).

Previous work suggests a functional link between host acceptance and fecundity on the host23. Thus, it is possible that each QTL dramatically affect both the dynamics and trajectory of evolution2,19,20. We term these ‘fundamental’ genetic correlations. In contrast, genetic correlations that are due to linkage disequilibrium among unlinked loci can only be maintained under very strong selection or in a subdivided population structure19,20, and should therefore be considered an outcome rather than a cause of the evolution of specialization (‘derived’ genetic correlations).

Depending on the underlying genetic architecture, these correlations could be either the cause or a consequence of the specialization and divergence of these incipient species. Genetic correlations due to linkage disequilibrium of closely linked genes or pleiotropy can be built up with little or no selection (respectively)19,20 and may
complex represents a single gene with pleiotropic effects, although a higher-resolution analysis is required to verify this. However, even if the QTL are separate genes or gene groups, they are clearly closely linked (Fig. 2), and, for several reasons, they may function in genetic correlations almost as if they were pleiotropic. First, physical linkage greatly increases the probability that linkage disequilibrium will arise by genetic drift and decrease the eroding effects of recombination. For example, although linkage disequilibrium between unlinked loci declines by 50 % in one generation, linkage disequilibrium between two loci that are 10 centimorgan (cM) apart would decline by only 9% (from equations (5.13) and (14.4) in ref. 20). Also, inbreeding, and especially cyclical parthenogenesis, further reduces the effects of recombination, with the effective map distance reduced from that under random mating, linkage disequilibrium between two loci that are 10 cM apart would decline by only 9% (from experimentwise significance levels of P < 0.05 from permutation tests: symbols marked with an asterisk suggest 0.08 < P < 0.12; 95% confidence intervals for the location of each QTL were based on the permutation tests. The plus or minus sign within each symbol indicates the directionality of the QTL effect. Letters W, X, Y and Z mark complexes of QTL that affect several traits each. 

Our analyses provide evidence that the genetic architecture of host use could have facilitated divergence and reproductive isolation of these incipient species. Although ecological specialization can occur in several different ways, none are as parsimonious as the complementary network of genetic correlations between resource use and mate choice (habitat acceptance) that is demonstrated here. Because this type of genetic architecture could significantly increase the probability of speciation for populations that evolve to utilize different resources, it may be a hallmark of ecological speciation by natural selection, especially in sympatric populations.

Methods

Crosses

Pea aphids are cyclically parthenogenetic, so parental and F1 genotypes can be clonally propagated. The mapping population was generated from pairwise reciprocal crosses between a single pair of clones, one a typical alfalfa specialist and the other a typical clover specialist. Two F2 genotypes from this cross were reciprocally crossed to obtain 194 F2 progeny. Each of these sexually produced genotypes was then maintained clonally, which permitted replicated phenotypic testing and increased the power of our QTL analysis.

Linkage mapping

Each F1 was genotyped for 173 informative dominant AFLP markers, at which parents were determined to be homozygous for different alleles: all F1 have a band (that is, they are heterozygous), and F2 are 3:1 banded: unband. Separate linkage maps were made from recessive markers for each parental genotype using MapMaker/Exp® software. The two parental framework maps, made with a subset of the markers, were aligned for three of the four linkage groups using sequence-tagged AFLP fragments.

Demographic performance of F2

The 194 F2 progeny were tested in three or four replicates on each of the two hosts. For each trial, a winged adult of a given genotype was allowed to produce offspring for 24 h in an enclosure containing one of the hosts. The adult was then removed, and after 5 days, one of the offspring was moved to a new enclosure in which its fecundity in the first 9 days of adult life was measured (variables FecA or FecC, depending on the test host).
Behavioural acceptance of alfalfa and clover by \( F_2 \)

For a given trial, four winged forms (alates)\(^{17}\) of one \( F_2 \) genotype were placed onto a small enclosure containing one potted alfalfa or clover plant. Alates were all in their adult migratory period\(^{17}\). After 70 h, the percentage of alates in the enclosure that were on the host with offspring was used as the measure of host acceptance (AccA or AccC, for alfalfa and clover, respectively; all values were arcsine square-root transformed).

**QTL mapping**

Mapping analyses were performed on the best linear unbiased predictors (BLUPs, Proc Mixed, SAS\(^{28}\)) for each trait from the replicated trials: 9-day fecundity on each host (FecA, FecC), and behavioural acceptances of each host (AccA, AccC). We used composite interval mapping in QTL Cartographer\(^{29}\) software, choosing markers as covariates for the analysis by stepwise regression. QTL for each character were placed on the linkage map if they were significant experimentwise at \( P < 0.05 \) by permutation test\(^{30}\). A few putative QTL with lower significance levels (0.08 < \( P < 0.12 \)) are also plotted (clearly marked on the figures by asterisks).

Received 15 February; accepted 5 July 2001.


Acknowledgements

We thank G. Churchill, D. Schemske, D. Schluter and Z.-B. Zeng for discussions. M. Arnold, T. Bradshaw, I. Conner, K. Shaw and J. Wilkinson provided useful comments on the manuscript. Z.-B. Zeng provided the calculation from ref. 22, and R. Lande pointed out the effects of cyclical parthenogenesis on recombination rates. A. Bouch, S. Skillman, C. Olson and A. Badgley provided expert technical assistance. This work was supported by NSF and USDA grants to S.V. and D.J.H.

Correspondence and requests for materials should be addressed to S.V. (e-mail: sv47@umail.umd.edu).

---

**Dynamics of travelling waves in visual perception**

Hugh R. Wilson\(^*\), Randolph Blake\(^†\) & Sang-Hun Lee\(^‡\)

\(^*\) Biology and Centre for Vision Research, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada

\(^†\) Department of Psychology, Vanderbilt University, Nashville, Tennessee 37203, USA

Nonlinear wave propagation is ubiquitous in nature, appearing in chemical reaction kinetics\(^1\), cardiac tissue dynamics\(^2,3\), cortical spreading depression\(^4\) and slow wave sleep\(^5\). The application of dynamical modelling has provided valuable insights into the mechanisms underlying such nonlinear wave phenomena in several domains\(^2,3,5\). Wave propagation can also be perceived as sweeping waves of visibility that occur when the two eyes view radically different stimuli. Termined binocular rivalry, these fluctuating states of perceptual dominance and suppression are thought to provide a window into the neural dynamics that underlie conscious visual awareness\(^2,6\). Here we introduce a technique to measure the speed of rivalry dominance waves propagating around a large, essentially one-dimensional annulus. When mapped on to visual cortex, propagation speed is independent of eccentricity. Propagation speed doubles when waves travel along continuous contours, thus demonstrating effects of collinear facilitation. A neural model with reciprocal inhibition between two layers of units provides a quantitative explanation of dominance wave propagation in terms of disinhibition. Our model provides a new tool for investigating fundamental cortical dynamics.

On first experiencing binocular rivalry, people often comment not only on the remarkable disappearance of one monocular stimulus for several seconds at a time, but also on the highly ordered transitions in dominance as one stimulus sweeps the other out of conscious awareness. These dominance waves are particularly prominent with larger rival patterns subtending many degrees of visual angle\(^9\). To study these dominance waves we simplified pattern geometry by using annular stimuli, thereby restricting wave propagation effectively to the one dimension around the annulus. Readers capable of free-fusion may experience salient traveling waves using the rival patterns in Fig. 1a or the anaglyphs on the website (see Fig. 1 legend).

To measure the dynamics of wave propagation, we devised a technique allowing us to control the location at which dominance waves originate and, subsequently, to estimate travelling speed around the annulus. Our technique capitalizes on the fact that an abrupt contrast increment in a suppressed pattern reliably triggers its immediate dominance\(^9,10\). The observer depressed and held the spacebar of a keyboard when the low-contrast radial grating was completely suppressed by the high-contrast spiral. This action triggered a brief contrast increment in the suppressed pattern at one of eight points (cardinal axes and diagonals). The spacebar was released only when the suppressed pattern became dominant at a monitored location, marked by nonius lines (see Methods). Data were consistent across all observers and demonstrated a linear
increase in propagation time with distance around the annulus (Fig. 1b). At the greatest distance around the annulus, two observers showed a flattening of their radial data, which is attributable to spontaneous reappearance of the suppressed pattern before arrival of the triggered dominance wave. For each observer propagation times, $T_p(x)$, were therefore fit with an equation incorporating constant-speed wave propagation ($v$) along with the gamma probability, $P(t)$, of spontaneous release from suppression at the target site before wave arrival:

$$T_p(x) = T_0 + \frac{x}{v} \left(1 - \int_0^v P(t) \, dt\right) + \int_0^v (\int_0^v P(t') \, dt') \, dt$$

(1)

where $x$ is the travel distance and $T_0$ is a constant response latency evident in the zero-distance data. The second and third terms in the equation are: (wave arrival time, $x/v$) $(\times$ (probability of no prior spontaneous reappearance) $(+ \times$ (expected time of spontaneous reappearance) $(\times$ (probability of prior spontaneous reappearance).

Least-mean-squares fitting of $v$ and $P(t)$ parameters to the radial data revealed an average propagation speed across observers of $3.65 \pm 0.54$ degrees s$^{-1}$.

Recurrent excitatory connections in primate visual cortex preferentially interconnect cells with similar preferred orientations and receptive fields that are roughly collinear$^{12,13}$. Both psychophysical$^{14}$ and transcranial magnetic stimulation$^{15}$ studies provide supporting evidence for collinear facilitation in humans. Therefore, we repeated our measurements using a low-contrast concentric target in place of the radial target. The same spiral pattern was again used, as it had the same local orientation difference (45°) from both radial and concentric contours. All observers again showed a linear increase in propagation time with distance, but the slopes were much shallower, signifying a greater propagation speed (Fig. 1b).

Collinearity of the suppressed target contours increased speed approximately twofold in three observers and even more in the fourth, averaging 9.60 ± 4.76 degrees s$^{-1}$. This increase in speed is consistent with previous evidence for facilitation between collinear gratings during the dominance phase of rivalry$^{16}$.

We investigated three additional aspects of dominance wave propagation. First, a low-contrast, spiral target pattern was used with a pitch angle orthogonal to the high-contrast spiral mask. This manipulation produced a speed of 5.8 degrees s$^{-1}$, intermediate between radial and concentric patterns (Fig. 1b, bottom left). Second, we tested whether dominance waves could propagate across a gap in the suppressed stimulus. Accordingly, a permanent gap (0.92° wide) in visual angle (three grating cycles) was introduced into the radial annulus at a point that was 67.5° distant from the marked arrival point. Dominance waves that were triggered 67.5° beyond the gap (that is, 135° from the arrival point) were blocked by the gap for both of the observers tested, and propagation times rose from 1.60 ± 0.055 s (S.L.) or 1.55 ± 0.056 s (R.B.) without the gap, to 2.71 ± 0.17 s (S.L.) or 2.28 ± 0.10 s (R.B.) in the presence of the gap. Both of these differences were highly significant ($t_{16} > 45.0; P < 10^{-6}$ for each subject), and the increased times correlate with the longer pathway (by 67%) in the opposite direction around the annulus when the shorter pathway is blocked by the gap. Very small gaps, however, can be traversed by dominance waves: a gap width of only 0.31° (one radial grating cycle) yielded equivalent propagation times for gap and no-gap conditions. Third, we determined whether eye movements would disrupt wave propagation. At the moment of wave initiation, observers shifted fixation from the central region (bull’s-eye) of the target to the marked arrival point itself. Arrival times were now independent of distance around the annulus, implying that eye movements effectively abolished retinotopically-based wave propagation.

To learn how wave speed varies with eccentricity, we scaled our entire stimulus so that the mean annular radius doubled from 1.8 to 3.6°, spatial frequency being halved to compensate for reduced resolution at the greater eccentricity. Complete data using the radial

---

**Figure 1** Rivalry stimuli and data of propagation times for dominance waves. a. In all experiments one eye viewed the high-contrast spiral grating (middle), while the other eye viewed either the lower-contrast radial (left) or concentric grating (right). Viewers can experience dominance wave propagation by free-fusing the bull’s-eyes. (Anaglypic versions of these stimuli and demonstrations of triggering are available at http://www.psy.vanderbilt.edu/faculty/blake/rivalry/waves.html.) Typically, when the radial grating is pitted in rivalry against the spiral, one small portion of radial grating achieves local dominance, and this propagates around the annulus. b. Propagation times for four observers (H.R.W., top left; R.B., top right; S.L., bottom left; K.S., bottom right) as a function of distance in degrees of visual angle around the annulus. Propagation times were significantly longer for the radial grating (filled circles) than for the concentric grating (open circles), and times for the spiral grating (open triangles) were intermediate. Lines are the best fits of equation (1), and standard errors are indicated.

**Figure 2** Dependence of propagation times on cortical distance. a. Best-fitting complex, logarithmic approximation (dashed lines) to a flattened retinotopic map of human V1 reported previously$^{17}$. Thick lines plot the mapping of half annuli with radii of 1.8 and 3.6°. Distance around the annulus was converted into centimetres across cortex using the formula: 1° = 0.6 cm (1.8° radius); 1° = 0.3 cm (3.6° radius). b. Radial pattern data for two subjects and two eccentricities indicate that propagation times are roughly constant in cortical coordinates. The best fit of equation (1) (thick line) produced an estimate of cortical speed of 2.24 cm s$^{-1}$. 

---

© 2001 Macmillan Magazines Ltd
Figure 3 Neural model for dominance wave propagation. a, Units driven by the spiral (S) or target (T) monocular pattern (radial, concentric, or spiral) generate mutual inhibition driven by interneurons (grey circles $I_T$ and $I_S$). The spatial spread of this inhibition produces dominance wave propagation through recurrent disinhibition. Collinear facilitation was simulated through recurrent excitation (dashed arrows). b, Dominance wave propagation generated by equations (2—4) (see Methods). Left, radial pattern dominance has just been triggered locally by a contrast pulse at the arrow. After 400 ms (right) the dominance wave has propagated over 40% of the annulus (arrows indicate propagation direction).
magnetic resonance imaging has revealed signatures of rivalry both in V1 (ref. 29) and in the extrastriate fusiform face area20. One possible resolution is that rivalry may be mediated by a complex interplay involving several cortical areas. In this scheme rivalry waves reflect activity propagation across V1, while more complex stimulus8,30 or spatio-temporal conditions2 accentuates aspects of rivalry dependent on higher cortical levels. In any case, our data demonstrate that the site of dominance wave propagation is retinotopically organized, has cortical magnification similar to V1, and exhibits collinear facilitation.

Our technique for measuring the speed of dominance wave propagation should prove useful in future studies of rivalry and other dynamic cortical phenomena. For example, the technique might be applied to behaving primates using spatial electrode arrays either in V1 or higher centres where rivalry has been reported. In addition, the enhanced speed of propagation along collinear contours provides a window into cortical excitability, as demonstrated by our model. It is tempting to suppose that individual and age-related differences in rivalry alternation rates may reflect differences in such collinear facilitation. These and other possibilities can now be tested using dominance wave speed as a sensitive probe.

Methods

Experiments

Four observers (one naive) participated in one or more of these experiments. Annual rival patterns were generated on a 21-in NEC monitor (1,024 × 768 resolution; 100 Hz frame rate) controlled by a Power Macintosh computer. The two annular patterns, each 1.8° in mean radius (except for the eccentricity experiments), were viewed through a mirror stereoscope with the head stabilized by a chin and head rest. Width of the spiral annulus was 0.95°, while the magnetic half-width of the radial or concentric pattern annulus was 0.46°. The spiral pattern had a pitch angle of 45° for the high-contrast mask so as to be at the same relative orientation with respect to both radial and concentric annular patterns (Fig. 1a). (Spiral target patterns had a 45° pitch so as to be locally orthogonal to the spiral mask.) Radial and concentric cosine grating contrast was adjusted for each observer to a value at which the spiral (100% contrast) was completely dominant for most of the viewing period—grating contrast varied among observers from 15–25%. During the experiment, observers maintained strict fixation on the identical bull’s-eye fusion targets in the centre of the pattern. Once the low-contrast grating was completely suppressed, the spiral alone being dominant, the observer depressed a switch. This produced a 100-ms contrast increment in the suppressed grating at one of eight equally spaced cardinal locations. The increment comprised a gaussian spatial envelope with half-width along the annulus of 18 arcmin; the magnitude of the contrast increment was 70%, a value sufficient to penetrate suppression locally on every trial. With fixation always maintained on the bull’s-eye, the observer monitored the phenomenal status of the grating region, demarcated by two nonius lines. Once that portion of the grating became dominant, the observer released the switch, thereby recording transit time to the arrival point. Trials were run in blocks of 32 with rest periods as needed.

Model

Model simulations were conducted in Matlab software on a Macintosh G4 computer using a Runge-Kutta routine with constant step size (0.25 ms). The equations are29:

\[ \tau \frac{dV}{dt} = -T_n + \tau \left( 100P^0 \right) + P \quad \text{where} \quad \tau = 20 \text{ ms} \]

and

\[ P = \frac{E_c - 0.27}{\tau} \sum_i \exp \left( -\tau_c / \tau_c^0 \right) + \frac{g}{\tau} \sum_i \exp \left( -\tau_c / \tau_c^0 \right) \]

\[ \tau_c \frac{dI_c}{dt} = -I_c + T_n \quad \text{where} \quad \tau_c = 11 \text{ ms} \]

\[ \tau_n \frac{dH_n}{dt} = -H_n + T_n \quad \text{where} \quad \tau_n = 900 \text{ ms} \]

where \( T_n \) is the firing rate of an excitatory neuron driven by the target stimulus, \( I_n \) and \( I_n \) are firing rates of inhibitory neurons driven by the respective excitatory neurons, and \( H_n \) is the spike frequency adaptation variable for \( T_n \). The input \( P \) to each neuron includes a constant \( E_c \) representing the stimulus strength, subtractive inhibition from a spatially weighted sum of \( I_n \) cells (\( \sigma = 1.0 \text{ mm} \)), and a term embodying collinear facilitation. The 900-ms time constant for the dH/dt equation is based on slow after-hyperpolarizing (AHP) potentials in human excitatory neurons24.

A comparable set of equations describes activity of S neurons driven by the spiral. For concentric patterns, the collinear facilitation parameter \( g = 0.04 \), but \( g = 0 \) for radial patterns. The intermediate case of spirals was approximated by \( g = 0.02 \) and reducing the spatial spread of excitation by half. Parameters were chosen to reflect available anatomical and physiological data. Thus, the time constant for inhibitory neurons (described by a linear equation for simplicity) was faster than that for excitatory cells, reflecting properties of cortical fast-spiking neurons. In equation (2) \( P_n \) was 0.93 so that negative inputs drive the firing rate to zero. The maximum firing rate of the S and T neurons was chosen to be 100 by convention, and the Naka–Rushton nonlinearity has been related to cortical physiology elsewhere25,26. Effective stimulus strengths of the rivaling monocular patterns were chosen to be \( E_c = 30 \) for the high contrast grating and \( E_c = 24 \) for the lower-contrast target stimulus. Variations of these numbers changed wave propagation speed by no more than 10%. Maximum strength of collinear excitation was constrained, using standard analytical techniques, so there could be no uniformly excited equilibrium state in the network27. Although our model is deterministic, addition of noise can easily generate a gamma function distribution of intervals29.

Received 6 April; accepted 1 July 2001.
Two-step process for photoreceptor formation in Drosophila

Bertrand Mollereau*, Maria Dominguez†‡, Rebecca Webel§, Nansi Jo Colley§, Benison Keung*, Jose F. de Celis†‡ & Desplan*‡

* Department of Biology, New York University, New York, USA
† Instituto de Neurociencias, UMH-CSIC, Campus de San Juan Alicante 03550, Spain
‡ Departments of Ophthalmology and Visual Science and Genetics, University of Wisconsin, Madison, Wisconsin 53792, USA
§ Centro de Biología Molecular ‘Severo Ochoa’, Universidad Autonoma de Madrid, Madrid 28049, Spain
† These authors contributed equally to this work

The formation of photoreceptor cells (PRCs) in Drosophila serves as a paradigm for understanding neuronal determination and differentiation. During larval stages, a precise series of sequential inductive processes leads to the recruitment of eight distinct PRCs (R1–R8)1. But, final photoreceptor differentiation, including rhabdomere morphogenesis and opsin expression, is completed four days later, during pupal development2,3. It is thought that photoreceptor cell fate is irreversibly established during larval development, when each photoreceptor expresses a particular set of transcriptional regulators and sends its projection to different layers of the optic lobes. Here, we show that the spalt (sal) gene complex4–7 encodes two transcription factors that are required late in pupation for photoreceptor differentiation. In the absence of the sal complex, rhabdomere morphology and expression of opsin genes in the inner PRCs R7 and R8 are changed to become identical to those of outer R1–R6 PRCs. However, these cells maintain their normal projections to the medulla part of the optic lobe, and not to the lamina where outer PRCs project. These data indicate that photoreceptor differentiation occurs as a two-step process. First, during larval development, the photoreceptor neurons become committed and send their axonal projections to their targets in the brain. Second, terminal differentiation is executed during pupal development and the photoreceptors adopt their final cellular properties.

The two zinc finger proteins of the sal gene complex are expressed in distinct subsets of PRCs throughout eye development8–10. sal major (salm) and sal related (salr) have almost identical expression patterns in most tissues, including the imaginal disc, and thus are likely to have similar or overlapping roles. In eye imaginal discs, the sal genes are expressed in a very dynamic pattern including R3 and R4 PRCs and cone cells. However, no obvious defects have been reported in salm mutant eye discs (U. Gaul and M. Mlodzik, personal communications). In the adult, salm is no longer expressed in outer PRCs but is restricted to the inner PRCs, R7 and R8 (ref. 9). To determine when this transition happens, we analysed salm expression during pupal life. After 24 h of pupation, salm was expressed in R3, R4 and cone cells (Fig. 1a and b). After 48 h, salm expression was strongly diminished in the cone cells with only weak labelling detected at 72 h in these cells (Fig. 1c and d). Between 48 h and 60 h of pupation, salm expression was turned off in R3 and

Figure 1 Expression of Salm during eye pupal maturation. Twenty-four hour pupal retina showing Salm expression in R3 and R4 (a) and in cone cells (b). Cone cells are located in a focal plane above R3 and R4. After 48 h, Salm expression is lost in cone cells but remains in R3 and R4 (c). After 72 hours, Salm expression is limited to R7 and R8 where it remains to adulthood. Weak expression of Salm is also detected in cone cells at 72 h (d). R7 cells were identified with anti-Prospero antibodies (shown in Fig. 4).

Figure 2 salm/salr mutant clones reveal transformation of inner photoreceptor cells (PRCs) into outer PRCs. a–e, Phase-contrast images of tangential sections through mosaic salm/salr mutant eye tissue. Wild-type cells are pigmented (close-up in a) and salm/salr cells are unpigmented. Whenever an R7 (arrowhead in b) or R8 (arrowhead in c) cells is mutant, it forms a large rhabdomere typical of outer PRCs. In addition, a small proportion of mutant ommatidia have additional outer PRCs (asterisk in e). Scale bars, 5 μm (a, c); 4 μm (b); 3.3 μm (d, e). f–h, Electron microscopy analysis of salm/salr mutant ommatidia; ey-FLP GMR hid CL/F40Df(2L)5. One-week-old salm/salr mutant ommatidia contains a photoreceptor with duplicated rhabdomeres (arrows in f) (scale bar, 1.5 μm), or display nine individual rhabdomeres connected to separate cell bodies (scale bar, 2.0 μm) (g). Eight-week-old salm/salr mutant ommatidia containing PRCs with reduced rhabdomeres or internalized rhabdomeres (arrows in h) (scale bar, 2.2 μm), as well as cells lacking rhabdomeres. R, rhabdomere. Numbers 1–8 stand for photoreceptor cell numbers, defined by their position and size.
Expression of rhodopsins is altered in salm/salr (Df(2L)32FP5) mutant eyes. In wild-type flies, the six outer PRCs, R1–R6, express rhodopsin1 (rh1) and mediate image formation and dim light vision. The inner PRCs R7 and R8 express distinct rhodopsins (rh3 or rh4 in R7; rh5 or rh6 in R8) that mediate perception13–15. To define more precisely the cell fates adopted by sal mutant ommatidia, we examined rhodopsin expression in eyes that were completely mutant for salm/salr (Fig. 3). All rhabdomeres contained rh1 (Fig. 3b) but none of the inner rhodopsins (Rh3, Rh4, Rh5 and Rh6) were detected at significant levels (Fig. 3d, f and h), strongly supporting the model proposed above that both R7 and R8 were transformed into outer R1–R6 PRCs. A similar, but weaker and less penetrant phenotype was observed with a single mutant for salm (salm3 allele) (data not shown), probably owing to a partially redundant function for this family of related transcription factors. Because salm is normally only expressed in non-Rh1–positive PRCs, and because rh1 expression is expanded in salm/salr mutants, it is possible that rh1 is repressed by salm/salr. However, the extent to which Sal proteins regulate rhodopsin promoters is as yet unknown.

The fairly late timing of salm expression in inner PRCs during pupal life (that is, much later than the time PRCs send out their axons in third instar larvae) suggested that its role in photoreceptor differentiation was not related to early photoreceptor specification or axon pathfinding. Consistent with this hypothesis, we found that, in the salm/salr mutants, the transformed R7 and R8 cells projected their axons to the medulla, which is their normal site of projection (Fig. 4b and d). Therefore, it appears that the determination of R7 and R8 is correctly initiated, but that these cells later adopt features typical of outer R1–R6 PRCs. Furthermore, the expression of prospero (pros), an early cell-marker for R7 neurons17 whose expression is maintained in the adult and controls aspects of R7 differentiation (T. Cook and C.D., manuscript in preparation; and

Figure 4 R7 and R8 projections and pros expression remain unaltered in salm/salr mutant eyes. R7 and R8 projections to the medulla are revealed with 24B10 antibodies in WT (a) and salm/salr mutants (b). The two levels of projections of R7 and R8 are distinguishable (arrows). (c, e) In WT, rh1-lacZ staining spans the whole retina and outer PRCs project only to the lamina. (d, f) In salm/salr mutants, rh1-lacZ reveals additional projections to the medulla, demonstrating that R7 and R8 express rh1 but still project to the medulla. L, lamina; M, Medulla. Prospero, an early R7 marker, is present in R7 nuclei in WT (e) and remains in transformed R7 cells in salm/salr mutants (f). Staining in the upper part of the retina reveals the cone cells.
Mouse anti-Rh5 (1/20) antibodies were a gift from S. Britt. Rabbit anti-Rh6 antibodies and 50% pupation) were isolated and stained directly, while late-pupal eye discs (60 and (1/5,000) were generated in our laboratory (D. Killian and P. Beaufils). Rat anti-Salm parentheses above) were incubated overnight at 4°C or a secondary Cy3 goat anti-mouse or anti-rabbit (diluted 1/400) (Jackson) were used. Sections. After several washes, a secondary FITC goat anti-rabbit antibody (diluted 1/200), ey GMR-hid

To generate marked salm/salh homologous mutant clones in the eye, w; FRT40 Df(2L)123F5-5/Cyo males were crossed with w-FLP; FRT40 P[w+]/30C/Cyo females. To generate salm/salh or salm (salm1) clones in the GMR-hid background FRT40 Df(2L)123F5-5/Cyo or FRT40 salm males (a gift from U. Gaul) were crossed with eYFP;FRT40MRaid CL females. The wild-type tissue was eliminated by the combination of the GMR-hid transgene and a cell lethal gene, which together induce apoptosis in tissues that are non-homologous for salm/salh or salm.


Acknowledgements

We are indebted to R. Barrio for her contribution to this work. We thank S. Britt, R. Khurlein, C. Doe, C. Zuker, P. Beaufils and K. Basler for fly stocks and antibodies, the Desplan and Treisman laboratories for support and discussion, and I. Tan for help with ultrathin section analysis. We are also grateful to U. Gaul and M. Mlodzik for allowing us to report unpublished results, J. Treisman for support, and T. Cook and F. Pichaud for comments on the manuscript. We would like to thank Developmental Study Hybridoma Bank for antibodies. B.M. was supported by the Human Frontier Science Program (HFSP). This work was supported by grants from the National Eye Institute (NEI) to C.D., from HHMI, Research to Prevent Blindness (RPB) and the Retina Research Foundation to N.L.C., and the Dirección General de Investigación from Ministerio de Ciencia y Tecnología (MCYT) to M.D.

Correspondence and requests for materials should be addressed to C.D. (e-mail: claudes.desplan@mvu.ee) or J.E.C.D. (e-mail: jildeclos@tisto.chbm.uam.es).
A major goal of molecular oncology is to identify means to kill cells lacking p53 function. Most current cancer therapy is based on damaging cellular DNA by irradiation or chemicals. Recent reports support the notion that, in the event of DNA damage, the p53 tumour-suppressor protein is able to prevent cell death by sustaining an arrest of the cell cycle at the G2 phase. We report here that adenovirus-associated virus (AAV) selectively induces apoptosis in cells that lack active p53. Cells with intact p53 activity are not killed but undergo arrest in the G2 phase of the cell cycle. This arrest is characterized by an increase in p53 activity and p21 levels and by the targeted destruction of CDC25C. Neither cell killing nor arrest depends upon AAV-encoded proteins. Rather, AAV DNA, which is single-stranded with hairpin structures at both ends, elicits in cells a DNA damage response that, in the absence of active p53, leads to cell death. AAV inhibits tumour growth in mice. Thus viruses can be used to deliver DNA of unusual structure into cells to trigger a DNA damage response without damaging cellular DNA and to selectively eliminate those cells lacking p53 activity.

Saos-2 cells (a p53-null, pRb-null osteosarcoma line) and U2OS cells (which are wild type for p53 and pRb) were infected with AAV-2. Saos-2 cells died, while U2OS cells did not die but enlarged (see Supplementary Information). To exclude replication of the AAV, ultraviolet-inactivated viruses were used and found to give similar results. Ultraviolet-inactivated AAV was used for all subsequent experiments. Flow cytometry revealed that Saos-2 cells, when infected with AAV, accumulated briefly with 4n DNA content. Cell death occurred soon after (Fig. 1a). Many of these cells contained a quantity of DNA less than 2n (Fig. 1a), which together with Annexin V staining (data not shown) showed that they underwent apoptosis. However, most infected U2OS cells arrested with 4n DNA content and re-entered the cell cycle after several days (Fig. 1b). A minority of cells arrested with 2n DNA content. When higher amounts of AAV were used, most of the U2OS cells arrested for a prolonged period (Fig. 1c). Whereas lower amounts of virus were sufficient to kill Saos-2 or transiently arrest U2OS cells (Fig. 1a and b), higher amounts were used to prolong the arrest and facilitate analysis of the arrested cells. To investigate the contribution of p53 activity in the response of cells to AAV infection, we inactivated the p53 protein in U2OS by ectopically overexpressing p53DD, a trans-dominant negative p53 mutant. Infection of these cells with AAV resulted in a transient G2 arrest followed by cell death, as seen with Saos-2 cells (Fig. 1d).

When normal human osteoblasts (NHOs) were infected with AAV, they also enlarged and arrested with 4n DNA content (Fig. 1e). When the p53 protein in NHOs was degraded by expression of the human papilloma virus 16 (HPV-16) E6 protein before infection with AAV, the cells arrested at G2 for a short period before dying (Fig. 1f). These observations show that the effect of AAV is not unique to osteosarcomas but is also observable in normal osteoblasts, and that the outcome of this infection is dependent on p53 activity. Consistent with this, the p53 protein in U2OS was stabilized after AAV infection (Fig. 2a). A similar increase was also observed for the p21 protein (Fig. 2a), which is indicative of an increase in p53 activity and a possible increase in p21 activity.

To analyse further the cell cycle block imposed by AAV, we assayed the activity of the cyclin B–cdc2 kinase complex, which triggers the G2/M transition of the cell cycle. Cells blocked in mitosis by nocodazole exhibited high cyclin B–cdc2 kinase activity (Fig. 2b). However, AAV-infected U2OS and Saos-2 cells, despite having a 4n DNA content, possessed cyclin B–cdc2 kinase activity that was even lower than that of the unsynchronized control population, indicating that the AAV-induced block was at the G2 phase. Following infection of U2OS with AAV, a substantial fraction of the cdc2 protein migrated on gel electrophoresis at a slower rate than the control, indicating that it might be phosphorylated (Fig. 2c). Cells treated with nocodazole did not show the reduced migration of cdc2 protein, further suggesting that AAV does not induce an arrest in mitosis. Upon checking the protein level of CDC25C, a phosphatase crucial for activating cdc2 (ref. 8), we found that it decreased dramatically in AAV-infected U2OS (Fig. 2d) but not in AAV-infected U2OS+p53DD cells (Fig. 2e). This disappearance of CDC25C protein was prevented by N-acetyl-leu-leu-norleucinal (NaLLN), involving inhibition of the proteasome complex (Fig. 2f). This degradation was specific as the level of CDC25B protein (Fig. 2d) and that of many other proteins tested were unchanged.

To determine whether AAV could elicit these effects on other cell
types, we infected a pair of isogenic human colon carcinoma cell lines, HCT116p53+/+ and HCT116p53−/− (ref. 1), with AAV. Both the HCT116p53+/+ and HCT116p53−/− cells arrested at the G2 phase of the cell cycle (Fig. 1g and h). However, the HCT116p53−/− cells underwent cell death after several days (Fig. 1h). In AAV-infected HCT116p53−/− cells, the levels of p53, p21 and 14-3-3σ proteins increased whereas that of CDC25C decreased (Fig. 2g). Interestingly, the HCT116p53+/+ and HCT116p53−/− cells, like that in the U2OS+p53DD cells, was unchanged after AAV infection (Fig. 2h), once again revealing the coupling of CDC25C protein degradation to p53 activity. To test the relative importance of p21 and 14-3-3σ, HCT116 cells that lacked either of these proteins were infected with AAV. Cells lacking p21 failed to sustain the G2 arrest and died, while those lacking 14-3-3σ were still able to sustain the arrest with minimal cell death (Fig. 1i and l), underlining the importance of p21 in the G2 arrest.

The cellular responses to AAV and to DNA damage bear many similarities. In both cases, cells can respond either by establishing an arrest at G2 if p53 is present, or by pausing briefly at G2 before dying, when p53 is absent13. We treated U2OS cells with etoposide, a DNA-damaging drug10, in place of AAV infection. When the levels of p53, p21 and CDC25C proteins were analysed, they were found to change in a manner similar to that caused by AAV (Fig. 2i). We know that, in response to DNA damage, nuclear localization of CDC25C is inhibited, inducing a G2 arrest11. Our results reveal another mechanism for inactivation of CDC25C phosphatase: proteasomal degradation. An early step in DNA damage signalling can be mediated by the ATM protein12. U2OS cells infected with AAV in the presence of the ATM inhibitor caffeine13 failed to arrest at the G2 phase, but continued to proliferate (Fig. 1k). Consistent with this, ATM-null cells (AT5BI, transformed with simian virus 40 (SV40)) were not affected by AAV. The SV40-transformed control cells (GM847 and MRC5-SV2) accumulated in G2 with cell death presumably depending on the level of residual p53 activity (Fig. 1l–n). These findings suggest that AAV affects the cell by inducing an ATM-dependent DNA damage response. Failure to induce this in ATM-null cells renders AAV ineffective in causing G2 arrest or apoptosis.

To test whether the incoming viral proteins were responsible for these effects, we used AAV-like particles without DNA we made using recombinant baculoviruses14, or empty AAV particles containing Rep, but not AAV DNA. When exposed to ultraviolet irradiation and added to cells, none of these affected cell growth (data not shown). Retroviral-mediated expression of the Rep proteins in cells did not induce the effects we saw with AAV15. The ultraviolet-inactivated AAV used here is unable to support the synthesis of viral proteins or DNA. As the effect of AAV on cells was not diminished by ultraviolet treatment, we conclude that viral replication or gene expression is not required for these effects. Ultraviolet-induced lesions in DNA were not responsible, as equivalent amounts of ultraviolet-inactivated adenovirus had no effect in parallel experiments (data not shown). Instead the evidence indicates that AAV DNA, which is single stranded with hairpin loops at both ends, can be sensed as abnormal DNA by the cell16 and trigger a DNA damage response. We tested this by injecting Saos-2, U2OS and U2OS+p53DD cells with an oligonucleotide corresponding to the AAV hairpin and containing no AAV coding sequence. This killed

Figure 2 Biochemical analyses of proteins in AAV-infected cells. a, c–i, Western blotting analyses of various proteins in the indicated cell types. b, Cyclin B–cd2 kinase assay of cells with and without AAV infection. H1, histone 1.

Figure 3 Cell viability following co-injection of AAV hairpin DNA (the inverted terminal repeat, ITR) and plasmids encoding green fluorescence protein (GFP) into different cell types.
the Saos-2 and U2OS+p53DD cells but not the U2OS cells (Fig. 3), supporting the conclusion that it is indeed the DNA of AAV that triggers a DNA damage response in the cells.

We tested whether this mechanism may be used to prevent the establishment of tumours. HCT116p53<sup>+/−</sup> or HCT116p53<sup>−/−</sup> cells were injected under the skin of nude mice, followed by injection of AAV, or PBS as a control, 2 days later. All of the control HCT116p53<sup>−/−</sup> injections gave rise to tumours. Tumour incidence was reduced to 17% when AAV was used. HCT116p53<sup>−/−</sup> cells were more resistant to AAV, as 80% of injections still produced tumours when AAV was present (Table 1 and see Supplementary Information). No non-specific toxicity was observed. This result is consistent with the observation<sup>17</sup> that defective AAV particles that contained sub-genomic-length DNA with terminal hairpins protected hamsters against adeno-virus-induced tumours. DNA fragments extracted from these defective particles were also able to prevent tumorigenesis. We then asked whether AAV is able to prevent the growth of pre-existing tumours lacking p53. Injection of AAV into established HCT116p53<sup>−/−</sup> tumours inhibited their growth, resulting in tumours smaller than those injected with PBS. AAV injected into tumours derived from HT29, another p53-null colon carcinoma cell line, caused complete regression of 60% of the tumours and suppressed the growth of the remainder (Table 1 and see Supplementary Information).

How human tumours that lack p53 activity die in response to therapy remains puzzling<sup>18</sup>. Activation of p53 is often considered to lead to cell death, notably in lymphocytes<sup>19</sup>. However, there are cell types and situations where the presence of p53 prevents cell death by DNA damage<sup>2,10</sup>. As irradiation and genotoxic drugs cause physical damage to cellular DNA, it was proposed that when damaged p53-null cells attempt to divide, they undergo mitotic catastrophe<sup>2</sup>. In the experiments described above, we induced a DNA damage response using AAV instead of damaging the DNA of the cell, and observed that cells which lack p53 activity arrest transiently and die, suggesting that they possess a checkpoint that triggers apoptosis if they fail to sustain a G2 arrest in the face of DNA damage signalling. The absence of damaged cellular DNA may also explain why the ATM<sup>−/−</sup> cells did not establish a delayed G2 arrest, as reported for certain ATM<sup>−/−</sup> cells following irradiation<sup>20</sup>.

Whereas further work will elucidate what kind of DNA damage AAV DNA is mimicking and details of which pathway is being triggered, the immediate importance of these findings is the introduction of the principle of using viruses to efficiently deliver DNA with unusual or modified structures into cells to eliminate those cells that lack p53 activity.

**Methods**

**Cell culture and inactivation of p53 in vivo**

U2OS and Saos-2 cells were cultured in DMEM supplemented with 10% fetal calf serum. NHO cells were purchased from Clonetics and were cultured in osteoblast Growth Medium (Clonetics) supplemented with 10% fetal calf serum and ascorbic acid. DNA encoding the p53DD protein was cloned into the retroviral vector plabhupero. To inhibit ATM activity, cells were treated with 2 mM caffeine. Annexin V analysis was performed according to the instructions of the manufacturer (Boehringer).

**Inactivation of AAV and infection of cells**

AAV prepared as described<sup>12</sup> was diluted in 0.5 ml PBS in a small plastic dish and exposed to 2,400 J/m<sup>2</sup> of ultraviolet irradiation (wavelength 254 nm) from a Stratalinker. The inactivated viruses were further diluted in 2.5 ml DMEM (10% FCS) before adding them to cells. After 3 h, fresh medium was added to make up 10 ml.

**Flow cytometry**

Cells were gathered and fixed in 70% ethanol. Fixed cells were centrifuged, resuspended and incubated in 100 μg/ml RNase at PBS at 37°C. After 30 min, propidium iodide was added to a concentration of 100 μg/ml<sup>2</sup>. DNA content was measured with a flow cytometer.

**Western blotting and cyclin B–cdc2 kinase assay**

Cells were washed twice with PBS and collected with a rubber-covered rod. After centrifugation cell pellets were resuspended in two volumes of Reporter Lysis Buffer (Promega) supplemented with a cocktail of protease inhibitors (Calbiochem). After incubation on ice for 30 min with occasional swirling, the samples were centrifuged at 16,100 g for 10 min. Supernatants were collected and protein concentrations measured by the Bradford assay (BioRad). Protein samples of 30 μg were resolved on SDS-polyacrylamide gels, transferred to nylon membranes (Hybond) and analysed with antibodies against p53, p21, CDC25C, CDC25B, actin, cyclin B and cdc2 (all from Santa Cruz). Cyclin B–cdc2 kinase assays were performed as described<sup>22</sup>.

**Injection of cells**

DNAs used in these experiments were first passed through a 0.2-μm filter. pCieGFP is a plasmid that contains a CMV promoter which controls the expression of the green fluorescent protein gene. The AAV hairpin oligonucleotide was synthesized (Microsynth) based on the sequence of the AAV-2 inverted terminal repeat (nucleotide positions 1–145). DNAs (either pCieGFP at 400 μg ml<sup>−1</sup> or pCieGFP at 200 μg ml<sup>−1</sup> with hairpin DNA at 200 μg ml<sup>−1</sup>) were injected into cells with an Eppendorf Microinjector. Green cells were visible 4 h after injection and were then counted on successive days. Received 30 March; accepted 11 June 2001.

---

**Table 1 Effects of AAV on growth of tumours in athymic mice**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Tumour incidence (%)</th>
<th>Relative tumour volume (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PBS</td>
<td>12 of 12</td>
<td>100</td>
</tr>
<tr>
<td>AAV</td>
<td>10 of 10</td>
<td>100</td>
</tr>
<tr>
<td>PBS</td>
<td>8 of 8</td>
<td>80</td>
</tr>
</tbody>
</table>


---

© 2001 Macmillan Publishers Ltd NATURE | VOL 412 | 30 AUGUST 2001 | www.nature.com
Cyclic nucleotide-gated (CNG) channels are crucial components of visual, olfactory and gustatory signalling pathways. They open in response to direct binding of intracellular cyclic nucleotides and thus contribute to cellular control of both the membrane potential and intracellular Ca²⁺ levels. Cytosolic Ni²⁺ potentiates the rod channel (CNG1) response to cyclic nucleotides and inhibits the olfactory channel (CNG2) response. Modulation is due to coordination of Ni²⁺ by channel-specific histidines in the C-linker, between the S6 transmembrane segment and the cyclic nucleotide-binding domain. Here we report, using a histidine scan of the initial C-linker of the CNG1 channel, stripes of sites producing Ni²⁺ potentiation or Ni²⁺ inhibition, separated by 50° on an α-helix. These results suggest a model for channel gating where rotation of the post-S6 region around the channel’s central axis realigns the Ni²⁺-coordinating residues of multiple subunits. This rotation probably initiates movement of the S6 and pore opening.

The mechanism of Ni²⁺ modulation is simple: Ni²⁺ binding is state dependent. Ni²⁺ is an intermediate Lewis acid and prefers association with intermediate Lewis bases, such as the imidazole of histidine. In CNG1 channels, naturally bearing H420, the open channel has a higher affinity (>100 times) for Ni²⁺ than the closed channel, and Ni²⁺ promotes channel opening. In CNG2 channels, naturally bearing a histidine at the 417 equivalent position, the closed channel has higher affinity for Ni²⁺ than the open channel, and Ni²⁺ promotes channel closing (Fig. 1a) [5]. Ni²⁺ affects the opening transition, but not the intrinsic ligand binding, as Ni²⁺ modulation is seen even in the presence of saturating concentrations of partial agonists where binding sites are maximally occupied [12,13]. Individual histidines bind Ni²⁺ poorly (~1 mM), but multiple histidines, especially in a square planar geometry, can coordinate Ni²⁺ with sub-micromolar affinity. Ni²⁺ coordination in CNG1 channels occurs between two adjacent channel subunits [10].

Predictions of the secondary structure [14] suggest that the S6 and the C-linker of CNG1 channels are α-helical in nature until position 420. The KcsA K⁺ channel has been crystallized [12], providing a structural model for the pores of other P-loop-containing cation channels, including CNG channels. To create a working model, we threaded the aligned amino-acid sequence of the S6 and P-loop of the CNG1 channel through the KcsA structure and extended the α-helix of the CNG1 channel S6 to position 420 (Fig. 1b). Residue 417 (Fig. 1b, red) points in the direction of the central axis, which allows Ni²⁺ coordination, while residue 420 (Fig. 1b, green) points away from the central axis. Rotation of the helix on channel opening could change the orientation of the residues such that residue 420 would point towards, and 417 away, from the central axis, thereby explaining state-dependent histidine coordination of Ni²⁺.

This model predicts that other positions near 417 and 420 in the post-S6 region might also be able to coordinate Ni²⁺. We completed a histidine scan between positions 403 and 420 in a Ni²⁺-insensitive CNG1 channel, H420Q (Fig. 2c right) [2,5]. Each residue between position 403 and 420 was individually replaced with histidine, and cGMP dose-response curves were measured in the absence and presence of Ni²⁺.

We found three Ni²⁺ phenotypes: Ni²⁺-inhibited, Ni²⁺-potentiated and Ni²⁺-insensitive. Ni²⁺ inhibited three channels, Q409H, D413H and Q417H (Fig. 2a). For Q409H and D413H channels, Ni²⁺ caused a decrease in current at sub-saturating cGMP concentrations, causing the cGMP dose-response relationship to shift right (Fig. 2a, right). Q417H currents were completely inhibited by 1 μM Ni²⁺, even at saturating concentrations of cGMP, which made the measurement of cGMP dose responses in Ni²⁺ impossible. Ni²⁺ potentiated two channels, K416H and H420Q (Fig. 2b). Addition of Ni²⁺ to these channels increased the current, causing the cGMP dose-response relationship to shift left (Fig. 2b, right). Ni²⁺ had no effect on 11 out of 18 channel constructs (Fig. 2c). Two mutants (I415H and M419H) failed to produce currents in either the absence or presence of Ni²⁺.

To quantify the effect of intracellular Ni²⁺, we fit the cGMP dose-response relationships for all mutants (except D413H) in the absence and presence of Ni²⁺ with a simple allosteric model containing two ligand-binding events and an allosteric opening transition (see Methods). Although the model is oversimplified, it provides a reasonable fit of the data and a straightforward analysis.
D413H channels produce a shallow dose-response relationship (Hill slope near 1) that cannot be fit with the two-binding-step model, and thus a single-binding-step model was used for D413H. Assuming a fixed value for the intrinsic binding affinity (K), the fits yield a measure of the equilibrium constant of allosteric opening (L). Large L values indicate energetically favoured opening transitions. Channels insensitive to Ni\(^{2+}\) had very similar L values in the presence and absence of Ni\(^{2+}\) (Fig. 3a). Mutants inhibited by Ni\(^{2+}\), Q409H and Q413H, had lower L values in Ni\(^{2+}\) than in its absence. Presumably Q417H would behave similarly if L in the presence of Ni\(^{2+}\) could be measured, as is the case when histidine is present in the equivalent position of CNG2 channels. Ni\(^{2+}\)-poletivated mutants, K416H and H420, showed opposite results—L values in Ni\(^{2+}\) were higher than in its absence. The ratio of L values in the absence and presence of Ni\(^{2+}\) were used to calculate the Ni\(^{2+}\)-induced change in free energy of the allosteric transition, \(\Delta G_{Ni}\) (Fig. 3b; see Methods). Ni\(^{2+}\)-insensitive mutants had \(\Delta G_{Ni}\) values near 0. Ni\(^{2+}\)-inhibited mutants had positive \(\Delta G_{Ni}\) values, as Ni\(^{2+}\) favoured the closed state; Ni\(^{2+}\)-poletivated mutants had negative \(\Delta G_{Ni}\) values, as Ni\(^{2+}\) favoured the open state. Ni\(^{2+}\) sensitivity was periodic, with inhibiting sites every four positions and potentiating sites every four positions.

All of the amino acids studied are well outside of the cyclic nucleotide-binding domain and were not expected to change the intrinsic histidine-binding affinity. However, it was possible that the changes in cGMP dose-responses were due to changes in intrinsic ligand binding. To test this, we measured current evoked by saturating concentrations of the full agonist cGMP and the partial agonists cAMP and cIMP with and without Ni\(^{2+}\) (Fig. 4a). Saturating concentrations of any cyclic nucleotide should maximally occupy the cyclic nucleotide-binding sites; therefore, changes in the intrinsic ligand affinity would not affect L values measured in saturating ligand. We tested all Ni\(^{2+}\)-sensitive mutants from our initial experiments plus the control channel (H420Q). Assuming that the mutations do not have cyclic nucleotide-specific effects on the L value (that is, mutant and Ni\(^{2+}\) effects on \(\Delta G_{Ni}\) are equivalent for all cyclic nucleotides), we determined L values for the mutants independent of a ligand-binding model (Fig. 4b; see Methods). As before, channels inhibited by Ni\(^{2+}\) had smaller L values in the presence of Ni\(^{2+}\), whereas channels potentiated by Ni\(^{2+}\) had larger L values in Ni\(^{2+}\). Ni\(^{2+}\) had no effect on the L value of the H420Q channel. \(\Delta G_{Ni}\) values calculated from saturating cyclic nucleotide data were similar to \(\Delta G_{Ni}\) values calculated from cGMP dose-responses (Figs 4c and 3b, respectively). Small differences in the \(\Delta G_{Ni}\) values calculated by these two methods may be due to the simplified nature of the gating model.

Ni\(^{2+}\)-inhibited positions were found every fourth amino acid (409, 413, 417). Potentiating positions (416, 420) were also four amino acids apart. Mapping the results of our scan on the homology model (Fig. 1b) generates Fig. 5a. Residue S399 (Fig. 5a, yellow) is at a position equivalent to the bottom of the helical bundle of KCa1 channels. A S399C mutation results in intersubunit disulphide bonds in closed, but not open, CNG1 channels, suggesting that 399 residues are closer together in closed than in open channels. The model in Fig. 5a assumes that the S helix continues unbroken through the initial C-linker. Inhibiting (red) and potentiating (green) positions fell along two vertical stripes on the helix. The stripe of inhibiting residues is offset 50° from the potentiating stripe. Clockwise rotation of the stripes relative to the central axis (as viewed extracellularly) on opening would move potentiating residues towards the place occupied by inhibiting residues in closed channels.

Although the model in Fig. 5a is intuitive, there are potential complications. The distances between coordinating residues are large (20–40 Å), too large to facilitate coordination of Ni\(^{2+}\) (~10 Å)\(^{13}\). Another problem is the orientation of residues on the helix. This model predicts reactive stripes that are parallel to the helical axis if the coordinating residues are to be in the same orientation relative to the central axis of the channel. This is not the observed pattern of every four residues, where reactive stripes...
cross the helical faces 35° off the helical axis. The observed pattern of Ni²⁺ reactivity is exactly as would be expected if these residues are at a right-handed, helical bundle crossing—analogue to the helical bundle crossing of KcsA. A possibility that would overcome both problems of distance and exposure angle would be if the helix were kinked between positions 399 and 409. Inserting a kink by rotating the ϕ- and ψ-angles of the peptide backbone at position 403 results in a 'kinky' model (Fig. 5b). The kink re-orientates inhibiting residues to directly face the central axis and gives them a shorter, more consistent intersubunit distance. Kinking also results in a second helical bundle crossing such that the pattern of diagonal stripes across the helical axes is explained.

Our data suggest that channel opening involves rotation of the initial C-linker around the central axis without translating radially (Fig. 5b, c). Reactive amino acids would effectively rotate clockwise relative to the central axis during opening. This movement would

![Figure 4](image_url)

**Figure 4** | Values calculated from saturating cyclic nucleotide concentrations. **a**, Currents induced by saturating cGMP (2 mM, green), cIMP (16 mM, blue), and cAMP (16 mM, red) at +60 mV are shown for Ni²⁺-sensitive channels without and with 1 μM Ni²⁺. **b**, L values calculated from saturating ligand currents without (filled circles) and with (open circles) 1 μM Ni²⁺. **c**, ΔΔG lié values for the opening transition are plotted versus histidine position for Ni²⁺-sensitive mutants. H420Q (control) values are circled on the right in **b** and **c**. One micromolar of Ni²⁺ abolished Q407H currents, thus neither L in Ni²⁺ nor ΔΔG lié could be determined for this mutant (question marks). Error bars indicate s.e.m.

![Figure 5](image_url)

**Figure 5** | Models showing how Ni²⁺ coordination could change CNG1 gating. **a**, Ni²⁺-sensitivity results are shown mapped on the model of Fig. 1a. Histidine substitution at the red positions produced Ni²⁺-inhibited channels. Histidine substitution at the green positions produced Ni²⁺-potentiated channels. Position 399 is yellow. Blue residues produced Ni²⁺-insensitive channels. Mutating grey residues yielded non-functional channels. The kinky model of channel gating is shown in the closed **(b)** and open **(c)** states.
allow the inhibiting positions to coordinate Ni$^{2+}$ when the channel is closed and the potentiating positions to coordinate Ni$^{2+}$ when the channel is open. It would also increase radial separation of 399 residues and widen the S6 bundle crossing—movements that are associated with channel opening$^1$. The movement of the C-linker is likely to control, or at least report, movement of the associated S6 transmembrane domain. Neither Ni$^{2+}$ inhibition nor potentiation was voltage dependent and the Ni$^{2+}$ concentration used does not cause appreciable block of the channel. We therefore expect that Ni$^{2+}$-binding residues are more cytoplasmic than the permeation pathway and that permeant ions flow through ‘windows’ below the S6 bundle crossing in the model.

Our data provide indications of the nature of movement in the CNG channel C-linker. The C-linker is critical to the coupling of ligand binding to channel opening$^{2,3,15–19}$. We show the C-linker bundle as a four-fold symmetrical structure, but our data do not eliminate the possibility that the bundles actually associate as a dimer of dimers. It is known that histidine-Ni$^{2+}$ coordination by H420 is mediated by adjacent but not opposite subunits$^{10}$, and there is evidence that the channel gates as a dimer of dimers$^{19}$.

The kinky model (Fig. 5b, 5c) is reminiscent of a model recently proposed for KcsA based on EPR data, where channel opening involves a translation and rotation of TM2 (refs 21, 22). More recently KcsA was suggested to be kinked just after TM2, causing a second, intracellular bundle crossing$^{23}$. Our data suggest that CNG1 channels also contain a second bundle crossing, which is cytoplasmic to S6, in the C-linker. We propose that a rotational movement at this bundle crossing is associated with channel opening. The CNG channel C-linkers probably act as lever arms, coupling conformational changes of the ligand-binding domains to pore opening.

Methods

Mutagenesis

All channel constructs were made from a modified version of CNG1 in which all endogenous cysteine codons have been mutated$^2$, and H420 was replaced with glutamine. Mutagenesis was accomplished by standard methods$^1$. The PCR-amplified regions of all constructs were fit by the linear model in scheme 1.

This model includes two independent cyclic nucleotide-binding steps followed by a concerted allosteric opening event. $K$ is the association equilibrium constant for binding of cGMP; $I$ is the equilibrium rate constant of the allosteric transition. The assumption of two binding steps is required to fit the slope of the dose-response data. The dose-response relationships for cGMP were fit by equation (1).

$$I = \frac{I_{\text{max}}}{1 + \left(\frac{[cGMP]}{K_{\text{cGMP}}} + \frac{[cGMP]}{K_{\text{cIMP}}}ight)^K}$$

(1)

where $I$ is the current and $I_{\text{max}}$ is the current when the open probability is 1. $K$ was determined by fitting multiple mutants simultaneously and $K$ was then held fixed at a value of 16,000 M$^{-1}$; $I_{\text{max}}$ and $L$ were allowed to vary in the fit. Equation (1) could not fit the D413H data as the cGMP dose-response relationship for that mutant was shallower than that of the other channel constructs. D413H channel data were therefore fit with the model in scheme (2).

This model includes a single cyclic nucleotide-binding step followed by a concerted allosteric opening event and is described by equation (2).

$$I = \frac{I_{\text{max}}}{1 + \left(\frac{[cGMP]}{K_{\text{cGMP}}} + \frac{[cGMP]}{K_{\text{cIMP}}}ight)^K}$$

(2)

The data at saturating cyclic nucleotide concentrations were fit by equation (3).

$$I_{\text{max}} = \frac{L_{\text{cGMP}}}{L_{\text{cGMP}} + L_{\text{cIMP}}} \times \left(\frac{I_{\text{max}}}{L_{\text{cGMP}}} + 1ight)$$

(3)

where $L_{\text{cGMP}}$ is the current of wild type or the experimentally determined macroscopic current with a saturating concentration of cyclic nucleotide (2.0 mM for cGMP and 16 mM for cAMP and cIMP); $I_{\text{max}}$ is the current if the channel open probability was 1. $L_{\text{cGMP}}$ is the current if the channel open probability was 0. $L_{\text{cIMP}}$ is the current if the channel open probability was 0.004 that were then held constant for fitting $L$. Using these ratios and the measured currents for $L_{\text{cGMP}}$, $L_{\text{cIMP}}$ and $L_{\text{cIMP}}$ we fit currents to equation (3) using a least squares algorithm.

The change in free energy of allosteric opening on addition of Ni$^{2+}$, $\Delta G_{\text{GTP}}$, was calculated from the experimentally determined $I$, values using equation (4).

$$\Delta G_{\text{GTP}} = -RT\ln\frac{I}{I_{\text{max}}}$$

(4)

Secondary structure prediction and analysis

Secondary structure was predicted using the nPred software (http://www.cmpharm.ucsf.edu/~nomi/npredict-instruc.html)$^2$. The dose-response data (except D413H) were fit by the linear model in scheme 1.

Oocyte expression and electrophysiology

CNG1 complementary DNA s in the pGEMHE plasmid were linearized with NheI restriction enzyme, and RNA was synthesized in vitro and injected into defolliculated Xenopus laevis oocytes as described$^2$. Recordings were made in the excised, inside-out patch clamp configuration using an Axopatch 200A patch-clamp amplifier (Axon Instruments) and Pulse acquisition hardware (HEKA Elektronik)$^2$. Patch pipettes were pulled from borosilicate glass and had resistances of 0.25–1 $M\Omega$. Cyclic nucleotides wereSigmaPlot (Jandel Scientific). The extracellular pipette solution for all recordings was (in mM): 130 NaCl, 0.2 EDTA, 3 HEPES, pH 7.2. Cyclic nucleotides were pulled from borosilicate glass and had resistances of 0.25–1 $M\Omega$. After fire polishing. We

© 2001 Macmillan Magazines Ltd NATURE | VOL 412 | 30 AUGUST 2001 | www.nature.com
Ablation of XRCC2/3 transforms immunoglobulin V gene conversion into somatic hypermutation

Julian E. Sale1*, Daniella M. Calandrini1, Minoru Takata1, Shunichi Takeda1 & Michael S. Neuberger*

* Medica Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
† Immunology and Molecular Genetics, Kawasaki Medical School, 577 Matsushima, Kurashiki, Okayama 701-0192, Japan
‡ CREST Research Project, Radiation Genetics, Faculty of Medicine, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan

After gene rearrangement, immunoglobulin V genes are further diversified by either somatic hypermutation or gene conversion1. Hypermutation (in man and mouse) occurs by the fixation of individual, non-templated nucleotide substitutions. Gene conversion (in chicken) is templated by a set of upstream V pseudogenes. Here we show that if the RAD51 paralogues XRCC2, XRCC3 or RAD51B are ablated the pattern of diversification of the immunoglobulin V gene in the chicken DT40 B-cell lymphoma line1 exhibits a marked shift from one of gene conversion to one of somatic hypermutation. Non-templated, single-nucleotide substitutions are incorporated at high frequency specifically into the V domain, largely at G/C and with a marked hotspot preference. These mutant DT40 cell lines provide a tractable model for the genetic dissection of immunoglobulin hypermutation and the results support the idea that gene conversion and somatic hypermutation constitute distinct pathways for processing a common lesion in the immunoglobulin V gene. The marked induction of somatic hypermutation that is achieved by ablation the RAD51 paralogues is probably a consequence of modifying the recombination-mediated repair of such initiating lesions.

A small proportion of the gene-conversion events occurring at the immunoglobulin loci in chicken B cells can lead to the generation of non-functional immunoglobulin genes—most frequently through introduction of frameshift mutations4–6. Thus, the generation of surface immunoglobulin-μ (slgM)-loss variants in DT40 cells can be used to give an initial indication of immunoglobulin variable region (V) gene-conversion activity. Compared with the parental DT40 line, a mutant that lacked RAD54 showed a considerably diminished proportion of slgM loss variants (Fig. 1). A fluctuation analysis performed on multiple clones revealed that the ΔRAD54 line generates slgM-loss variants at a frequency nearly tenfold less than that of parental DT40. In contrast, a ΔRAD52 line generates slgM-loss variants at a similar frequency to wild-type cells (Fig. 1). These observations support earlier findings concerning gene conversion in ΔRAD54- and ΔRAD52-DT40 cells7–8.

We extended our analysis to DT40 cells that lacked XRCC2 and XRCC3. We expected that these RAD51 paralogues, which (as with RAD52 and RAD54) are thought to have an involvement in the recombination-dependent pathway of DNA damage repair2,9–12, might also be implicated in immunoglobulin V gene conversion. Rather than seeing a diminished abundance of slgM-loss variants, the ΔXRCC2 and ΔXRCC3 lines showed a much greater accumulation of gene conversion variants than the parental line (Fig. 1). (The contrasting effects of RAD54- and XRCC2/3-deficiency on immunoglobulin V gene conversion suggests that, although these proteins are all implicated in recombination-mediated DNA repair, they probably have different functions, which is consistent with the distinct patterns of sensitivity of the deficient cells to different forms of DNA damage2,9–12.) In the case of ΔXRCC2-DT40, transfection of the human XRCC2 complementary DNA under control of the human β-globin promoter caused the frequency of generation of slgM loss variants to revert to values close to those of the wild type (Fig. 1).

As deficiency in both XRCC2 and XRCC3 is associated with chromosomal instability13–15, we wondered whether the increased frequency of slgM-loss variants could reflect gross rearrangements or deletions within immunoglobulin loci. However, Southern blot analysis of 24 slgM subclones of ΔXRCC3-DT40 did not reveal any loss or alteration of the 6-kilobase (kb) Sall/BamHI fragment containing the rearranged light-chain variable region (V) (not shown). To ascertain whether more localized mutations in the V gene could account for the loss of slgM expression, we cloned and sequenced the rearranged V segments in populations of slgM cells that were sorted from wild-type, ΔXRCC2- and ΔXRCC3-DT40 subclones after 1 month of clonal expansion.

Sixty-seven per cent of V sequences cloned from the slgM subpopulations that were sorted from multiple wild-type DT40 clones carried mutations: in most cases (73%), these mutations rendered the V sequences obviously non-functional (Fig. 2). (Presumably, most of the remaining slgM cells carried inactivating mutations either in the heavy-chain variable region (VH) or outside the sequenced region of V.) As detailed below (see Methods) the mutations could be classified as being attributable to gene conversion template by an upstream V pseudogene, to non-templated point mutation, or as falling into an ambiguous category. Most (67%) of the inactivating mutations were due to gene conversion although some (15%) were stop codons generated by non-templated point mutations, demonstrating that the low frequency of...
point mutations seen in this study (and in others\textsuperscript{45}) in DT40 cells is not an artefact of polymerase chain reaction (PCR) but rather reveals that a low frequency of point mutation does indeed accompany gene conversion in wild-type DT40.

A markedly different pattern of mutation was seen in \( V_\lambda \) sequences of the slgM loss variants from \( \DeltaXRCC2\)-DT40. Nearly all the sequences carried point mutations, typically with multiple point mutations per sequence (Fig. 2). A substantial shift towards point mutations was also seen in the sequences from the slgM- \( \DeltaXRCC3\)-DT40 cells. Thus, whereas a \( V_\lambda \)-inactivating mutation in wild-type DT40 is most likely to reflect an out-of-frame gene conversion tract, in \( \DeltaXRCC2/3 \) it is likely to be a point mutation (Fig. 2c). Furthermore, whereas most of the non-functional \( V_\lambda \) sequences obtained from sorted slgM loss variants of \( \DeltaXRCC2\)-DT40 (53%) or \( \DeltaXRCC3\)-DT40 (64%) carried additional point mutations as well as the \( V_\lambda \)-inactivating mutation, such hitchhiking was only rarely observed in the non-functional \( V_\lambda \) sequences from the parental DT40 line (7%; Fig. 2d).

All of these observations point to the hypothesis that the high prevalence of slgM-loss variants in \( \DeltaXRCC2\)-DT40 cells might simply reflect a very high frequency of spontaneous immunoglobulin \( V \) gene hypermutation in these cells. We therefore analysed the spontaneous \( V_\lambda \) mutation frequency in wild-type and \( \DeltaXRCC2\)-/3- DT40 cells by amplifying with PCR the rearranged \( V_\lambda \) segments from total (unsorted) DT40 populations that had been expanded for 1 month after subcloning. The result shows that there is indeed a much higher spontaneous accumulation of mutations in the \( \DeltaXRCC2 \) and \( \DeltaXRCC3 \) cells than in the parental DT40 line (Fig. 3a, b). (In addition to the mutations shown in Fig. 3b, one \( \DeltaXRCC2\)-DT40 sequence contained a 2-basepair (bp) insertion in the leader intron, which was not obviously templated from a donor pseudogene, and one \( \DeltaXRCC3\)-DT40 sequence carried a single-base-pair deletion in the leader intron.) In \( \DeltaXRCC2\)-DT40 cells, mutations accumulate in \( V_\lambda \) at a rate of about \( 0.4 \times 10^{-4} \) per bp per generation (given a division time of about 12 h), a value similar to that seen in the constitutively mutating, human Burkitt lymphoma line Ramos\textsuperscript{46}. Somatic hypermutation in germinal-centre B cells in man and mouse is preferentially targeted to the rearranged immunoglobulin \( \text{VH} \) and variable light-chain (\( \text{VL} \)) segments. A similar situation applies to the point mutations in \( \DeltaXRCC2\)-DT40 cells. Thus, a significant level of apparent point mutation is also seen in the productively rearranged \( \text{VH} \) gene (Fig. 3c). However, this does not reflect a general mutator phenotype, as accumulation of mutations is much lower in the light-chain constant region (\( C_\lambda \)) than in the rearranged \( \text{VL} \), and is also low in the unrearranged \( V_\lambda \) on the excluded allele where the apparent mutation rate does not rise above the background level ascribable to the PCR amplification itself (Fig. 3c; see Methods).

The distribution of mutations over the \( V_\lambda \) domain in \( \DeltaXRCC2\)-DT40 cells is markedly non-random. The mutations, which are predominantly single-nucleotide substitutions, show preferential accumulation at hotspots that conform to an AGY (where Y is pyrimidine) consensus on one of the two DNA strands (Fig. 3d). Mutations also show overwhelming accumulation (96%) at G/C (Fig. 3e). This G/C-biased, hotspot-focused hypermutation in \( \DeltaXRCC2\)-DT40 cells, although not exhibiting an obvious bias in favour of nucleotide transitions, is otherwise markedly similar to the pattern of \( V \) gene hypermutation described in cultured human Burkitt lymphoma cells as well as that occurring \textit{in vivo} in frog, shark and MSH2-deficient mice\textsuperscript{12,18}. The immunoglobulin \( V \) gene hypermutation that occurs \textit{in vivo} in man and in normal mice seems to be achieved by this hotspot-focused, G/C-biased component acting in concert with a mechanism that targets A/T. The mechanism of immunoglobulin \( V \) gene hypermutation in \( \DeltaXRCC2\)-/3- DT40 cells is likely to be similar to that responsible for the G/C-biased phase of mutation in man and mouse. This accords well with the fact that bursal lymphoma cell lines show many parallels to the human Burkitt lymphoma lines in which immunoglobulin \( V \) gene hypermutation takes place. Both Romos and DT40 are p53\textsuperscript{47},

---

**Figure 1** Generation of slgM-loss variants by wild-type and repair-deficient DT40 cells. **a**. Representative flow cytometric analyses of the heterogeneity of slgM expression in cultures derived by 1 month of clonal expansion of single slgM\textsuperscript{+} normal (WT) or repair-deficient DT40 cells. An analysis of cultures derived from three representative slgM\textsuperscript{+} precursor clones is shown for each type of repair-deficient DT40. **b**. Fluctuation analysis of the generation frequency of slgM-loss variants. The abundance of slgM-loss variants was determined in several parallel cultures derived from slgM\textsuperscript{+} single cells after clonal expansion (1 month); median percentages are noted above each data set and are indicated by the dashed line. The (pPG-tXRCC2)\(\DeltaXRCC2\) cells are transfected with a human XRCC2 cDNA expression construct. The whole analysis was performed on multiple independent slgM\textsuperscript{+} clones (with distinct, although similar, ancestral \( V_\lambda \) sequences) giving, for each repair-deficient line, average median frequencies at which slgM-loss variants were generated after 1 month. WT, 0.4%; \( \DeltaRAD54, 0.07\% \); \( \DeltaRAD52, 0.4\% \); \( \DeltaXRCC2, 6\% \); and \( \DeltaXRCC3, 2\% \).
surface IgM+ B cells that express activation-induced deaminase (R. S. Harris and M. S. N., unpublished observations) and carry a c-MYC that has been activated by chromosomal translocation or retroviral integration. The main difference between immunoglobulin V gene mutation in human Burkitt lines and chicken DT40 cells is the higher proportion of transversion mutations observed with DT40 (Fig. 3e); this might reflect a difference in the polymerases involved in the process.

These results reveal that the DT40 chicken bursal lymphoma line, which normally exhibits a low frequency of immunoglobulin V gene diversification by gene conversion, can be induced to perform a high frequency of constitutive immunoglobulin V gene somatic mutation by ablating XRCC2 or XRCC3. This provides strong support to the earlier proposal that immunoglobulin V gene conversion and hypermutation might constitute different ways of resolving a common DNA lesion. This does not mean that modulation of XRCC2/3 levels will be the only way of shifting the balance between gene conversion and somatic hypermutation. Indeed, as with a previous proposal and experiment, we suspect a major reason that gene conversion is rarely observed in man and mouse may relate to the lack of closely located V gene donors present in accessible chromatin.

The initiating lesion for somatic hypermutation may well be a DNA break. It therefore seems significant that both XRCC2 and XRCC3 have been implicated in a recombination-dependent pathway of DNA break repair. Indeed, a similar induction of immunoglobulin V gene hypermutation in DT40 cells is achieved by ablating another gene (RAD51B) whose product is implicated in recombination-dependent repair of breaks, but not by ablating genes for Ku70 and DNA-PKcs (the catalytic subunit of DNA-

**Figure 2** Analysis of Vl sequences cloned from sIgM-loss variants. a, Comparison of Vl sequences from sIgM-loss cells sorted from parental sIgM+ clones of normal or XRCC2-deficient DT40 cells after clonal expansion (1 month). Each horizontal line represents the rearranged Vl/Jl (427 bp) with mutations classified as described in Methods as point mutations (lollipop shape), gene conversion tracts (horizontal bar above line) or single-nucleotide substitutions, which could be a result of point mutation or gene conversion (ambiguous, vertical bar). Hollow boxes, deletions; triangles, duplications. b, Proportion of Vl sequences carrying different numbers of point mutations (PM), gene conversions (GC) or mutations of ambiguous origin (Amb) among sorted sIgM-loss populations. Segment sizes are proportional to the number of sequences carrying the number of mutations indicated around the periphery of the pie charts. The total number of Vl sequences analysed is indicated in the centre of each chart with the data compiled from analysis of four subclones of wild-type DT40, two of ΔXRCC2-DT40 and three of ΔXRCC3-DT40. Deletions, duplications and insertions are excluded from this analysis. In wild-type cells, there were additionally 6 deletions, 1 duplication and 1 insertion. There were no other events in ΔXRCC2-DT40, and a single example each of a 1-bp deletion and a 1-bp insertion in the ΔXRCC3-DT40 database. c, Causes of Vl gene inactivation in wild-type, ΔXRCC2 (ΔX2) and ΔXRCC3 (ΔX3) DT40 cells expressed as a percentage of the total sequences containing an identified inactivating mutation. Missense mutation, black; gene conversion-associated frameshift, white; deletions, insertions or duplication-associated frameshift, grey. d, Additional mutational events associated with each inactivating mutation. Data are expressed as the mean number of additional mutations associated with each inactivating mutation with the type of additional mutation indicated as in c. Thus, ΔXRCC2-DT40 has a mean of 1.2 additional point mutations in addition to the index inactivating mutation, whereas wild-type DT40 has only 0.07.

© 2001 Macmillan Magazines Ltd
Figure 3 Analysis of immunoglobulin sequences cloned from sorted DT40 populations after 1 month of clonal expansion. a, V_\text{\lambda} \text{-} \text{S} \text{IgM} \text{+} \text{sequences} \text{obtained} \text{from} \text{representative} \text{wild-type} \text{and} \text{\DeltaXRCC2} \text{-} \text{DT40 clones;} \text{symbols} \text{as in Fig.} \text{2.} \text{b,} \text{Proportion} \text{of} \text{V_\text{\lambda} \text{-} \text{S} \text{IgM} \text{+} \text{sequences} \text{carrying} \text{different} \text{numbers} \text{of} \text{the} \text{various} \text{types} \text{of} \text{mutation.} \text{Data} \text{are} \text{pooled} \text{from} \text{analysis} \text{of} \text{independent} \text{clones:} \text{wild} \text{type,} \text{two} \text{clones;} \text{\DeltaXRCC2,} \text{two} \text{clones.} \text{c,} \text{Mutation} \text{at} \text{other} \text{loci} \text{of} \text{\DeltaXRCC2} \text{-} \text{DT40.} \text{Proportion} \text{of} \text{sequences} \text{derived} \text{from} \text{\DeltaXRCC2} \text{-} \text{DT40 in black} \text{and} \text{those} \text{from} \text{\DeltaXRCC3} \text{-} \text{DT40 in mid-grey.} \text{All} \text{mutations} \text{falling} \text{into} \text{the} \text{point} \text{mutation} \text{and} \text{ambiguous} \text{categories} \text{are} \text{included.} \text{Correction} \text{has} \text{been} \text{made} \text{for} \text{clonal} \text{expansion} \text{as} \text{described}^{16}, \text{so} \text{each} \text{lower-case} \text{letter} \text{represents} \text{an} \text{independent} \text{mutational} \text{event.} \text{Most of the} \text{27 mutations} \text{thereby removed} \text{from} \text{the} \text{original} \text{database} \text{of} \text{158 were at one of the seven principal hotspots; the correction for clonality will, if it gives rise to any distortion, lead to an underestimate of hotspot dominance.) Of the seven principal hotspots (identified by an accumulation of \text{\geq}5 mutations), five conform to the AGY consensus sequence on one of the two strands (black boxes).} \text{e, Nucleotide substitution preferences (as a percentage of the database of 131 independent events) deduced from point mutations in sequences from unselected \DeltaXRCC2} \text{and \DeltaXRCC3} \text{-} \text{DT40.} \text{A similar pattern of preferences is evident if the \DeltaXRCC2/\DeltaXRCC3 databases are analysed individually.}
DT40 cells are included for comparison. D-globulin V gene conversion, albeit at a reduced level (Fig. 3b).

Methods

dependent protein kinase), which are involved in non-homologous end-joining (Fig. 4).

The results, however, do not simply suggest that in the absence of XRCC2 a lesion that would normally be resolved by gene conversion is instead resolved by a process leading to somatic hypermutation. First, ΔXRCC2-DT40 cells retain the ability to perform immunoglobulin V gene conversion, albeit at a reduced level (Fig. 3b). Second, the frequency of hypermutation in ΔXRCC2-DT40 cells is about an order of magnitude greater than the frequency of gene conversion in the parental DT40 line.

We therefore propose that, in normal DT40 cells, only a small proportion of the lesions in the immunoglobulin V gene are subjected to templated repair from an upstream pseudogene, thereby leading to the observed gene conversion events. A large proportion of the lesions in the immunoglobulin V gene are subjected to a recombinational repair using the identical V gene located on the sister chromatid as a template, and which is therefore ‘invisible’. This would be consistent with the observation that detectable immunoglobulin V gene breaks in hypermutating mammalian B cells are restricted to the G2/S phase. In the absence of XRCC2, XRCC3 or RAD51B, we suggest that the ‘invisible’ sister-chromatid-dependent recombinational repair is perverted, thereby resulting in hypermutation. This hypermutation observed in the absence of XRCC2/XRCC3 could reflect either that sister-chromatid-dependent recombinational repair has become error prone, or that it has been inhibited, thereby revealing an alternative, non-templated mechanism of break resolution. Resolving which one of these two mechanisms applies will be important not only for an understanding of the mechanism of hypermutation, but also because it may provide an insight into the physiological function of RAD51 paralogues. Our identification of cell lines that perform immunoglobulin V gene hypermutation at high frequency and that they may be due to point mutation. The mutation prevalences in these data sets are: $\Delta$VH minus GACGATGACTTCGG) and the C

Supplemental data

All sequence changes were assigned to one of three categories: gene conversion, point mutation or an ambiguous category. This discrimination rests on the published sequences of changes from the consensus sequence of each clone.

Analysis of sIgM-loss variants generated in ΔXRCC2-DT40 clones

The frequency of sIgM-loss variants was monitored by flow cytometric analysis of cells that had been expanded for 1 month after cloning, and then stained with FITC-conjugated goat anti-chicken IgM. To each analysis the abundance of sIgM+ subclones was determined as the percentage of live cells (as judged by propidium iodide uptake and scatter gating) whose FITC fluorescence fell at least eightfold below that of the sIgM+ peak (see Fig. 1a).

Received 25 April; accepted 24 July 2001.


Figure 4 Analysis of sIgM-loss variants in DT40 cells deficient in DNA-PK, Ku70 and RAD51B. a, Fluctuation analysis of the frequency of generation of sIgM-loss variants after 1 month of clonal expansion. The median values obtained with wild-type and ΔXRCC2-DT40 cells are included for comparison. b, Proportion of Vý sequences amplified from the sIgM-loss variants derived from two sIgM- RAD51B-deficient DT40 clones carry various types of mutation. In addition, one sequence carried a 9-bp deletion, one carried a 4-bp duplication and one carried a single-base-pair insertion.

Methods

Cell culture, transfection and analysis

DT40 subclone CL18 and its mutants were propagated in RPMI 1640 medium supplemented with 7% fetal calf serum, 3% chicken serum (Life Technologies), 50 μM 2-mercaptoethanol, penicillin and streptomycin, at 37°C in 10% CO2. Cell density was maintained at 0.2–1.0 × 10^6 ml^(-1) by splitting the cultures daily. The generation of the DT40 derivatives carrying targeted gene disruptions has been described. Transfectants of two sIgM- ΔXRCC2-DT40 clones (that had given rise to sIgM- cells at a prevalence of 6.4 and 10.2% in the fluctuation analysis) were established by electroporation of a pSV2-neo-based plasmid containing the ΔXRCC2 open reading frame (cloned from HeLa cDNA) under control of the β-globin promoter.

CL18 is an sIgM- subclone of DT40 and is the parental clone for the DNA repair mutants described here. Multiple sIgM- subclones were obtained from both wild-type and repair-deficient mutants using a Mo-Flu (Cytomation) sorter after staining with FITC-conjugated goat anti-chicken IgM (Bethyl Laboratories). Reverseion of the single-base-pair frameshift in both the CDR1 region of CL18 Vý and in that of its ΔXRCC2 and ΔXRCC3 derivatives (which occurs at a lower rate), is most commonly achieved by a gene conversion using the υ8 donor. This is the reversion found in all subclones used in this study, thereby minimizing any possible effect of consensus-sequence variation on the conversion–mutation spectrum. However, the individual ΔXRCC2 and ΔXRCC3 sIgM- revertants used carried a small number of nucleotide differences from each other within Vý1 (see Fig. 3c), and these were taken into account when analysing mutation.

The frequency of generation of sIgM-loss variants was monitored by flow cytometric analysis of cells that had been expanded for 1 month after cloning, and then stained with FITC-conjugated goat anti-chicken IgM. To each analysis the abundance of sIgM+ subclones was determined as the percentage of live cells (as judged by propidium iodide uptake and scatter gating) whose FITC fluorescence fell at least eightfold below that of the sIgM+ peak (see Fig. 1a).

Mutation analysis

Genomic DNA was amplified by PCR from 5,000 cell equivalents using Pfu Turbo (Stratagene) polymerase and a hotstart touchdown PCR (8 cycles at 95°C (1 min), 68–60°C (1 min) per cycle; 1 min; 72°C (1.5 min); 22 cycles at 94°C (50 s), 60°C (1 min), 72°C (1.5 min)). The rearranged Vý was amplified using CVLF6 (5’-CAGAGGCTCGC GGGGGCGTCACTTGGCCG; priming in the leader-Vý intron) and CVLR3 (5’- GGGCAAGGTTTCCCAGCTGGCCGAACTTCAAG; priming back from 3’ of the light-chain joining segment (IýJý); the unarranged Vý1 using CVLF6 with CVLRBR1 (5’- GGATTTCTAGTGGAGAGGACGAGAGC; the rearranged Vý6 gene using CVHIF1 (5’- CGGGAGGCTCGCCTGAGTCCTGCTCC) with CJHRI1 (5’- GGGTATGCCCGGAGGACGAGATGCTCGG) and the Cý region using CJCIRJF (5’- CGGCAATCCGAATTTCCG CTACG; priming from within the Jý-Cý intron) and CMCMLAR (5’- GGGAGCATC GATCACCCTATGAGC; priming back from within Cý). After purification on QIAquick PCR purification kit (Qiagen), PCR products were cut with the appropriate restriction enzymes, cloned into pbU scriptSK and sequenced using the T3 or T7 primers and an ABI377 sequencer (Applied Biosystems). Sequence alignment (with GAP4 allowed identification of changes from the consensi sequence of each clone.

All sequence changes were assigned to one of three categories: gene conversion, point mutation or an ambiguous category. This discrimination rests on the published sequences of the Vý pseudogenes that could act as donors for gene conversion. DT40 is derived from an Fý hybrid bird with two immunoglobulin light-chain (A) alleles, Gý and Sý; the rearranged alleleý being Gý. The database of donor sequences was taken from ref. 28, but we implemented the modifications pertaining to the immunoglobulin A Gý allele identified by ref. 29. (The sequences/gene conversions identified in this work supported the validity of this Vý sequence database.) For each mutation the database of Vý pseudogenes was searched for potential donors. If no pseudogene donor containing a string $\geq 9$ bp could be found then it was categorized as an untemplated point mutation. If such a string was identified and there were further mutations that could be explained by the same and all then these mutations were assigned to the donor event. If there were no further mutations then the isolated mutation could have arisen through a conversion mechanism or could have been untemplated, and was therefore categorized as ambiguous.

Analysis of known Vý pseudogene sequences did not indicate that any of the mutations observed in the rearranged Vý2 (Fig. 3c) were due to gene conversion, suggesting that they may be due to point mutation. The mutation prevalences in these data sets are: 1.6 × 10^(-5) mutations per base pair for Vý6: 0.03 × 10^(-5) for the unarranged Vý1 and 0.13 × 10^(-5) for Cý as compared to 2.0 × 10^(-5) for point mutations in the rearranged Vý6 in ΔXRCC2-DT40, 0.13 × 10^(-5) for point mutations in rearranged Vý1 in wild-type DT40 and 0.89 × 10^(-5) for background PCR error.

Acknowledgements

We thank A. Johnson for cell sorting and C. Milstein for helpful discussions.

Correspondence should be addressed to J.E.S. (e-mail: jes@mrc-lmb.cam.ac.uk) or M.S.N. (e-mail: msn@mrc-lmb.cam.ac.uk).
Correspondence and requests for materials should be addressed to J.W.B.

Enos Lake. This work was supported by NSF-NATO and NSF-International Research fellowships to J.W.B. and NSERC operating grants to D. Schluter.

(1) M. atropurpureum, which was thought to be nodulated by Burkholderia genus. However, we performed phylogenetic analysis of gene sequences of the small subunit of ribosomal RNA (16S rRNA), and found that strain STM678 does not belong to any of the four branches of rhizobia described so far, nor even to the α-subclass of Proteobacteria, but instead belongs to the β-subclass of Proteobacteria (Fig. 1). From this analysis, we found the most closely-related sequences to that of strain STM678 (AJ302311) to be those of Burkholderia kururienesis (96.9% identity), B. brasilense (96.6% identity) and B. gladioli (96.8% identity). Phylorgenetic analyses of partial sequences of the 23S rRNA gene (AJ302313) and the dnaK gene encoding the chaperon heat shock protein (AJ302314) were consistent with the 16S rRNA analysis, thus unambiguously positioning strain STM678 in the Burkholderia genus within the β-subdivision of Proteobacteria.

To ensure that the Burkholderia strain STM678 was indeed a rhizobium, we checked its ability to re-nodulate a leguminous plant. Because seeds of the original host plant, A. carnea, were not available, we selected as test plant Macroptilium atropurpureum, a tropical legume capable of establishing a symbiosis with diverse rhizobia. Over a three-week period, strain STM678 formed 5 to 20 nodules per plant on the roots of M. atropurpureum (Fig. 2). The nodules exhibited the classical determinate nodule structure, with a central ‘infected’ tissue containing cells with intracellular bacteria and a peripheral tissue with vascular bundles (Fig. 2). Single colonies re-isolated from surface-sterilized nodules exhibited the characteristics of strain STM678, as assessed by 16S rDNA sequencing and nodA analysis using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP; see below). Hence, Koch’s postulates were verified. To eliminate the possibility that strain STM678 is a mixture of two different bacteria, a Burkholderia and a rhizobium, we isolated spontaneous mutants resistant to chloramphenicol, rifampicin and streptomycin, and showed by 16S rDNA and nodA PCR-RFLP analyses that the individual mutants retained the characteristics of both Burkholderia and rhizobia. Nodules induced on M. atropurpureum by strain STM678 were ineffective in terms of nitrogen fixation, probably because M. atropurpureum is not the original symbiotic partner of strain STM678. Supporting this conjecture, strain STM678 was indeed found to contain the nifH gene encoding dinitrogenase reductase, a key enzyme in nitrogen fixation (AJ302315). The highest identity values with other nifH genes were 81.2% (the α-Proteobacterium Azorhizobium caulinodans) and 81.1% (the β-Proteobacterium Herbaspirillum seropedicae).

Figure 1 Unrooted 16S rDNA tree of Proteobacteria (purple bacteria). The figure shows the phylogenetic relationships between the different rhizobial genera—as represented by type species in bold—including the new rhizobial Burkholderia sp. strains. α, β, δ, γ and ε represent the different subclasses of the Proteobacteria. The tree was constructed by using the neighbour-joining method and adapted from ref. 5. 16S rDNA sequences of published bacteria are available in GenBank. 16S rDNA from Burkholderia sp. STM 678 and Burkholderia sp. STM 815 are given in the text (AJ 302311 and AJ 302312).
Nodulation of legumes by rhizobia is controlled by a set of bacterial nodulation \( (\text{nod}) \) genes involved in the production of lipo-chitooligosaccharides (Nod factors) that act as signalling molecules for nodulating specific legume hosts. The \( \text{nodABC} \) genes are responsible for the synthesis of the core structure of the Nod factor, and as such are present in all rhizobia. We thus looked for the presence of \( \text{nodABC} \) genes in the nodulating \( \text{Burkholderia} \) strain STM678 by PCR amplification (see Methods). Sequencing of the amplified DNA revealed a genetic organisation of \( \text{nodAB} \) genes similar to that found in other rhizobia; that is, with \( \text{nodAB} \) in the same orientation and overlapping and preceded by a \( \text{NodD-depen-} \) dent regulatory sequence (\( \text{nod box} \)). A \( \text{nodC-like} \) sequence was found immediately downstream of \( \text{nodB} \). However this sequence is unlikely to correspond to a functional \( \text{nodC} \) gene as it lacks the \( \sim 600\)-base-pair \( 5' \) end of known \( \text{nodC} \) genes. We obtain evidence for the presence elsewhere in the STM678 genome of a longer \( \text{nodC} \) sequence that probably corresponds to the functional \( \text{nodC} \) gene (AJ306730). Such genetic unlinkage of \( \text{nodABC} \) genes is not unprecedented. The sequences of strain STM678 \( \text{nodAB} \) genes (AJ302321) revealed very high similarities with rhizobial \( \text{NodA} \) and \( \text{NodB} \) protein sequences available in databases, with values ranging from 62.8\% (\( \text{Sinorhizobium meliloti} \)) to 77.6\% (\( \text{Methylbacterium nodulans} \)) for \( \text{NodA} \) and from 55.6\% (\( \text{Rhizobium galegae} \)) to 70.9\% (\( \text{Mesorhizobium sp. N33} \)) for \( \text{NodB} \). To examine whether these genes were functional, we constructed a \( \text{nodA} \) mutant by introducing a \( \text{lacZ}-\text{kanamycin-resistance cassette} \) into the \( \text{nodA} \) gene of strain STM678. The \( \text{nodA} \) mutant did not form any nodules after inoculation on \( \text{M. atropurpureum} \), even after 30 days, indicating that the \( \text{nod} \) genes that we disrupted are required for nodulation of the \( \text{Burkholderia sp.} \) strain STM678.

By screening among bacteria isolated from root nodules collected from various legumes in French Guiana, we found a second \( \text{Burkholderia} \) strain, strain STM815, isolated from the legume \( \text{Machaerium lunatum} \). 16S rDNA sequencing (AJ302312) of this strain revealed the following sequence identities with its closest phylogenetic neighbours: 96.9\% (\( \text{Burkholderia kururiensis} \)), 96.8\% (\( \text{B. brasilense} \)), 96.6\% (\( \text{B. graminis} \)) and 96.9\% (strain STM678). These data clearly show that strain STM815 belongs to the \( \text{Burkholderia} \) genus, and most probably to a different species to strain STM678. The nodulation ability of STM815 was confirmed, as described for strain STM678, by inoculation of \( \text{M. atropurpureum} \) in axenic conditions and by re-isolation and characterization of the bacteria isolated from the induced nodules.

The \( \beta \)-subdivision of Proteobacteria contains many bacteria that interact with eukaryotes, including human pathogens, such as \( \text{Neisseria} \) and \( \text{Bordetella} \), and plant-associated bacteria. These latter bacteria include pathogenic \( \text{Ralstonia solanacearum} \), rhizospheric \( \text{Burkholderia} \) and endophytic \( \text{Azorhizobium} \). However the \( \beta \)-Proteobacteria had not been reported to include rhizobia, bacteria capable of nodulating leguminous plants. Here, we have identified two rhizobia belonging to the \( \text{Burkholderia} \) genus. These bacteria were isolated in different continents, from legumes belonging to different Papilionoideae tribes, and probably correspond to two distinct species. We have shown that the genetic control of nodulation by the \( \text{Burkholderia sp.} \) strain STM678 involves \( \text{nod} \) genes. Moreover, this strain has been shown to produce Nod factors. Hence rhizobia from both the \( \alpha \)- and \( \beta \)-Proteobacteria (now termed \( \alpha \)- and \( \beta \)-rhizobia) use the same strategy for establishing symbioses with legumes.

Furthermore, we have performed phylogenetic analyses that indicate a much smaller phylogenetic distance between the \( \text{nodAB} \) genes of strain STM678 and other rhizobia (Fig. 3) than between the 16S rRNA genes of \( \alpha \)- and \( \beta \)-Proteobacteria (Fig. 1). This suggests that the presence of \( \text{nod} \) genes in both \( \alpha \)- and \( \beta \)-rhizobia probably occurred through horizontal gene transfer. This transfer may have occurred after the appearance of legumes on Earth, about 70 million

---

**Figure 2** Nodules of \( \text{Macroptilium atropurpureum} \), three weeks after root inoculation with \( \text{Burkholderia sp.} \) strain STM678. **a**, Root segments with nodules; **b**, longitudinal section showing the typical structure of a determinate nodule with a central zone containing infected cells (Iz), and a peripheral region with vascular bundles (Vb).

**Figure 3** Unrooted \( \text{NodA} \) tree showing the close phylogenetic relationship between the \( \text{NodA} \) of strain STM678 and those of \( \alpha \)-rhizobia. The tree is based on full-length sequences, and constructed by using the neighbour-joining method. Bootstrap values (\% from 1,000 replications) are indicated. \( \text{NodA} \) sequences of published rhizobia are available in GenBank. \( \text{NodA} \) from \( \text{Burkholderia sp.} \) strain STM678 is given in the text (AJ302321). A, \( \text{Azorhizobium} \), B, \( \text{Bradyrhizobium} \), M, \( \text{Mesorhizobium} \). Me, \( \text{Methylbacterium} \). R, \( \text{Rhizobium} \). S, \( \text{Sinorhizobium} \).
Our current understanding of spatial behaviour and parietal lobe function is largely based on the belief that spatial neglect in humans (a lack of awareness of space on the side of the body contralateral to a brain injury) is typically associated with lesions of the posterior parietal lobe. However, in monkeys, this disorder is observed after lesions of the superior temporal cortex, a location topographically reminiscent of that for language in humans. Unlike the monkey brain, spatial awareness in humans is a function largely based on the belief that spatial neglect in humans, lesions located predominantly in the posterior parietal lobe are critical for this disorder. Analyses of computerized tomography scans of right-hemispheric stroke patients with neglect found that superimposed lateral projections of these lesions centred on the cortex, a location topographically reminiscent of that for language in humans. We now show that, contrary to the widely accepted view, the superior temporal cortex is the neural substrate of spatial neglect in humans, as it is in monkeys. Unlike the monkey brain, spatial awareness in humans is a function largely confined to the right side, rather than a shift from the temporal to the parietal lobe. One may speculate that this lateralization of spatial awareness parallels the emergence of an elaborated representation for language on the left side.

Spatial neglect is a characteristic failure to explore the side of space contralateral to a brain lesion. Patients with this disorder behave as if one side of the surrounding space had ceased to exist. Since the early post-mortem studies, we have believed that, in humans, lesions located predominantly in the posterior parietal lobe are critical for this disorder. Analyses of computerized tomography scans of right-hemispheric stroke patients with neglect found that superimposed lateral projections of these lesions centred on the inferior parietal lobule (IPL) and the temporoparieto-occipital (TPO) junction. More recent studies have confirmed the validity of this conclusion although evidence for additional pathology leading to this lateralization is lacking.

Spatial awareness is a function of the temporal not the posterior parietal lobe

Hans-Otto Karnath, Susanne Ferber & Marc Himmelbach

Department of Cognitive Neurology, University of Tübingen, Hoppe-Seyler-Strasse 3, 72076 Tübingen, Germany

Received 22 January, accepted 2 May 2001.

Handee devices
Pierce
Small helpings
www.piercenet.com

Designed for manipulating small volumes of chromatography supports, these devices are available in three formats: spin cup/mini-spin columns, resin separators and microcentrifuge tubes. Chromatography resin and samples are applied to the cup and a microcentrifuge is used to wash and elute the sample. Alternatively, Luer-lok adaptors can be purchased to do washes and elutions with a syringe, using a microcentrifuge tube. The resin separators are suitable for larger volumes (250 µl–1 ml) of resin, with no need to use a microcentrifuge or a syringe. The separator slides into a 16 mm \( /c_{525} /c_{450} \) 100 mm test tube containing the support and sample. The separator compresses the resin away from the sample, which is then decanted out of the tube. An instruction book can be downloaded from the company’s website.

CIS 3 S
Gerstel
www.gerstelus.com

Thermal desorption system

A new cooled injection system (CIS) enables Gerstel thermal desorption systems to be used with gas chromatography systems such as the Varian 3800GC, the Saturn GC/MS or the Perkin Elmer GC. Using the Gerstel Twister, it provides complete sample recovery and lowers detection capability to the ppt and ppq ranges. The cooled injection port serves as a re-focusing module for thermal desorption, providing headspace, large-volume liquid and SPME injections. The port functions in seven different modes and is optimised for capillary column GC columns with septumless head for ultra-low bleed and no sampled discrimination.

EZ-GRIP Chrompack capillary columns
Varian
www.varianinc.com

Gripping installment

EZ-GRIP Chrompack capillary columns have a toolbar that uses a series of tabs to hold column elements during installation. These tabs simplify pre-column/retention gap and dual-column installation, ensuring secure connection of single ferrule connectors. The tabs also help to identify column inlet and outlet ends for proper installation. They can be used with single and dual gas chromatography columns and the manufacturer claims that they can reduce installation time by as much as 90% over standard columns.

γ-RAM
IN/US Systems
www.inus.com

Radio-H PLC gamma detectors for online measurement

The γ-RAM is a research-grade, online radioactivity quantification system of \( ^{153} \), \( ^{131} \), \( ^{99m} \)Tc and other gamma emitters for H PLC techniques. Features include thallium-activated Na well-typed crystal for counting geometry, flexible GLP software with half-life correction, two independent radio channels, three analogue inputs for mass detectors, four analogue outputs for chart recording and chromatography systems, external control for any H PLC system and remote computer, and detector shielding for low backgrounds. Gamma and positron cells meet all applications needs. Liquid connections are located at the front panel for convenience and control of possible leaks. The system is recommended by the manufacturer for label verification and dosing in PET laboratories.

Switchos II
Dionex
www.dionex.com

Advanced microcolumn switching

Switchos II is designed for advanced microcolumn switching and is equipped with two 10-port valves, allowing more complex sample handling. It consists of a high-precision loading pump, two 10-port low dispersion switching valves and a four-channel solvent selection valve. The switching valves ensure dead volume-free connection of any capillary- or nano-LC column, such as the Fusica, NanoSeries, Micro- and Nano-Precolumns. The solvent selection valve permits automatic or manual selection of up to four different solvents without pump shutdown or purging. Applications include affinity extractions, isolation of phosphorylated peptides from complex protein digest, automated removal of detergents, extraction of drug compounds from biological fluids and high-throughput analysis.
Injection port septa
J G Finneran  www.jgfinneran.com
Optimized for specific analyses

This range offers septa specifically designed to accommodate different analyses in gas chromatography. Bleed and temperature optimized septa are suitable for injection-port temperatures up to 400 °C and have an injection life of up to 150 punctures. Advanced Green general-purpose septum provides low bleed conditions, an injection life of up to 100 injections and heat stability to 300 °C. LongLife Yellow 11 mm septum is suitable for prolonged and unattended operation, with heat stability to 350 °C and very low bleed at 300 °C. Low Bleed Gray 11 mm septum offers ease of sample injection and an injection life of up to 100 injections. The most economical choice is the 11 mm Low Bleed Red septum.

788IC
Metrohm  www.metrohm.co.uk
Integrated sample filtration for IC

In modern ion chromatography (IC) it is recommended that samples be filtered before injection, using membrane filters with pore sizes of 0.45 mm and lower. But the relatively high price of such disposable filters limits their use. The 788 IC filtration sample processor integrates membrane filtration as a sample preparation step directly in the IC system. The system is particularly suitable for light- to medium-loaded samples such as drinking water, surface water and wastewater, digestion solutions or extracts.

1100 Series
Agilent Technologies  www.agilent.com
Scalable sample purification

The 1100 Series purification system features scalable choices that match the purification process to sample size, ranging from microgram to gram quantities. A choice of pumps with flow rates up to 100 ml/min, flow cells and a high-capacity fraction collector optimize each analysis. Fraction collection is based on time, peak and mass, allowing users to collect and automatically reanalyse fraction. Both purification and purity checks are completed in an automated system.

Application note CD
Jones Chromatography  www.joneschrom.com
HPLC in the round

This CD contains 88 new applications for Genesis HPLC columns. It includes general applications and application notes pertaining specifically to pharmaceutical, environmental, biomedical and petrochemical work, as well as applications in food and beverage, quality control, vitamins, and peptides and proteins. The CD is free and can be ordered online at the company’s website.

LumaPlate-96
Packard Instrument  www.packardinstrument.com
Bar-coded microplates

LumaPlate-96 and DeepWell LumaPlate-96 are scintillator-coated microplates constructed from polystyrene, which provides greater chemical resistance to acetonitrile for counting HPLC fractions. The 3H background count rate is lowered by about a factor of two with the LumaPlate-96 and by about 30–40% with the DeepWell LumaPlate-96. Both microplates come with pre-applied bar code labels that can be read by the TopCount NXT microplate scintillation and luminescence counter. The standard bar code symbologies are code 128, code 39 and Interleaf 2 of 5. Other symbologies are available by special order.

High-performance chromatography columns
OmniFit  www.onmifit.co.uk
Smooth operators

These chromatography columns are precision-bored from borosilicate glass and have fully inerted wetted parts, resulting in minimal dead volume, elimination of sample contamination problems and smooth solvent flow. The endpieces are fully adjustable, allowing bed length to be varied, and the PTFE primary seal have a choice of viton, silicone or chemraz backup seals. The frit design allows rapid replacement in the event of a blockage. The columns are available in diheteros of 5, 10, 16 and 27 mm and in lengths between 50 mm and one metre. Accessories include column injectors, support rods and filling funnels. Applications include bioseparations and other analytical and semi-preparative laboratory work as well as general research and production.

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENT

Polyclonal Antibodies & Custom Peptides

Please contact us about complete services for your custom peptide and polyclonal antibody needs.

ResGen
Innogen Corporation

U.S. or Canada 800-533-4363
FAX 256-536-9016

Check out our web site at www.resgen.com

© 2001 Macmillan Magazines Ltd