EMBL

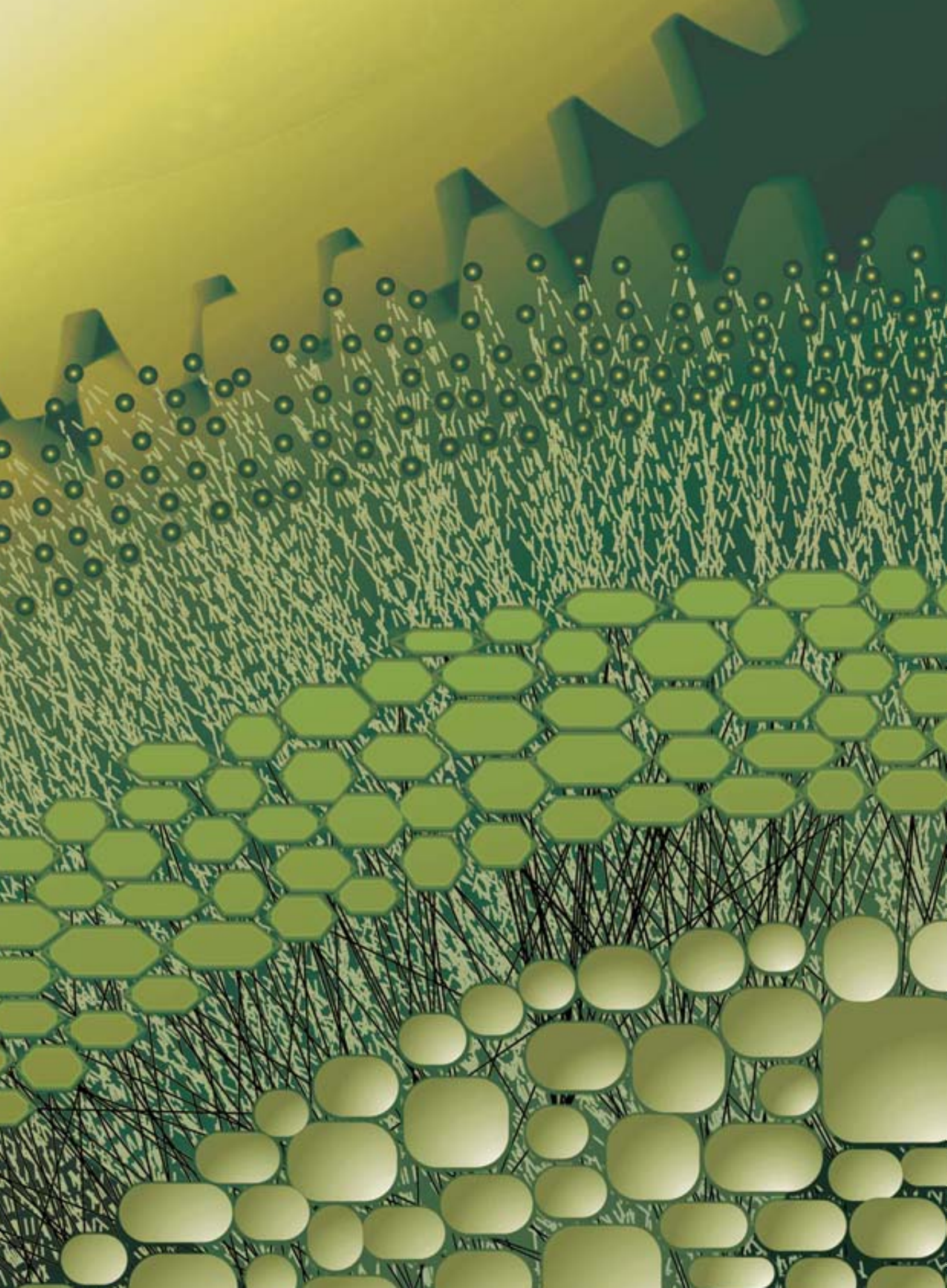
Annual Report

Annual Report

2005-2006

2005-2006

EMBL



Annual Report 2005-2006

Contents

Lab Notebook

The Director General's Report

Contents of The Director General's Report:

First year in office

This is my first DG Report at the end of one year in office as EMBL's fourth Director General. The year has been dominated by the development of a programme of activity for the next five-year period starting in 2007. This has been a unique opportunity for me to revisit all parts of EMBL, to discuss new ideas and to develop strategies with the larger EMBL community. The outcome of this process is the draft EMBL Programme 2007-2011, which has been received very positively by EMBL Council delegates and by a large number of other scientists and ministry representatives in the many member states that I have visited in the past months to introduce myself and to discuss EMBL's future.

It will be no surprise to anyone who followed my early presentations to EMBL Council that the most pressing issue on my agenda has been to discuss ways and means to obtain more support from the member states for the European Bioinformatics Institute (EBI) in Hinxton. EMBL-EBI represents an essential European infrastructure that provides support for an enormous user community and that must in future be maintained and supported in a more stable manner than is currently the case. The EMBL-EBI Outstation was established in 1993 when the data libraries moved from Heidelberg with the assurance that it would receive significant funding from the European Union (EU). This has turned out to be correct, and currently about 25% of EBI funding comes from the EU. However, this funding is provided in the form of many individual grants which are subject to instability and fluctuation. This instability stands in contrast to the very large, complex and rapidly growing core biomolecular databases that the EBI builds, maintains and serves to the scientific community in Europe and the world, and makes the EBI extremely vulnerable. Despite the efforts of EMBL and the European Commission's Directorate-General for Research to establish mechanisms by which more stable funding from the EU to the EBI would be available in Framework Programme 7 (FP7), I must now say that the reduction in the FP7 budget has led to cuts specifically in the budget for research infrastructures and that it is unlikely that the EBI can be supported by the EU in a more stable manner in FP7 than it has been in FP6.

In preparation for this eventuality, I had already asked the EMBL member states to consider supporting the EBI at a much higher level. I consider this essential if the high-quality service provided to the scientific community is to be maintained. I am grateful for the unanimous support that I have received from the EMBL member states for the concept that there should be an increase in EMBL's contribution to EBI funding from the present 50% to 60% or even 65% over the course of the next Programme period, while at the same time funding an expansion of the EBI from almost 300 staff to 400. The additional staff and computing infrastructure will be used mainly to cope with the exponential production, collection and distribution of data by the EBI, including nucleotide and protein sequences, macromolecular structures and gene-expression profiles. Research at the EBI will include finding ways to improve our services, data integration and new exploratory areas such as text mining and modelling. A building extension that will enable the necessary growth has been generously funded by the UK Medical Research Council (MRC), Biotechnology and Biological Sciences Research Council (BBSRC) and the Wellcome Trust, and will be completed in the middle of 2007.

Summary of EMBL Programme 2007-2011

Over the next five years, EMBL will continue to build on its strengths in research, service provision and advanced training. Research will focus on systems biology, which will be based on a horizontal integration of projects and technologies across the existing Units into four EMBL Centres. Particularly in the areas of Computational Biology, Molecular and Cellular Imaging, and High-Throughput Functional Genomics, the purpose of the Centres is to promote ambitious, interdisciplinary projects. A fourth Centre, for Disease Mechanisms, has been established to encourage and facilitate collaborations with clinical researchers. We are planning to increase the number of groups in our newest Outstation in Monterotondo to strengthen EMBL's Mouse Biology programme. Central internal scientific support services, in particular our information-technology (IT) infrastructure and the Core Facilities, will have to be maintained or upgraded. EMBL's leading role in service provision to external scientists in bioinformatics and structural biology can only be maintained if we invest in these important infrastructures. EMBL-EBI will have to grow to 400 staff during the next Programme, and, to an even greater proportional extent, increase its computing infrastructure. Access to synchrotron radiation and structural-biology technology platforms such as high-throughput crystallisation will have to be maintained. In Hamburg there will be a major effort in new beamline construction so that access to world-class facilities for structural biology will be available when the new synchrotron radiation source, Petra-III, becomes available. Our training activities have been combined into the EMBL International Centre for Advanced Training (EICAT) to use synergies better and to help in raising external funds. In March 2006, EMBL Council approved the construction of the Advanced Training Centre (ATC) in Heidelberg and building will start in autumn 2006, to be completed two years later. This building will allow us to bring together many facets of our training activities under one roof, and will also provide greatly improved facilities for courses and conferences. The administration needs to introduce modern personnel management software and will expand our vocational training programme for non-scientific staff. EMBL is in the process of introducing a long-term care insurance scheme for all staff members and their families.

Not all of the activities described above will require an increase in EMBL's annual budget. Over the next five years, large investments will be made by two of our host countries, Germany and the UK, to expand and upgrade our facilities in Heidelberg, Hamburg and Hinxton:

- construction of new beamlines at the Petra-III ring in Hamburg (funded mostly by Germany)
- construction of the Advanced Training Centre in Heidelberg (funded mostly by the German government and the Klaus Tschira Foundation, approved by EMBL Council in March 2006)
- construction of a building extension at EMBL-EBI (funded mostly by the UK MRC, BBSRC and Wellcome Trust).

Other activities will be made possible by reallocating existing resources. We have, for example, replaced the Biochemical Instrumentation Programme by using some of the resources for chemical biology and Core Facilities.

The largest requested increase is for the EBI. We are asking the member states to increase the EMBL contribution to the EBI budget from 50% to 60% or even 65% and to expand the EBI staff and infrastructure to the size required to enable it to serve its large and growing user community. The first phase of the refurbishment of the main Laboratory was approved by EMBL Council in April 2005. The second phase, also approved by Council, has been reduced significantly in cost due to the construction of the ATC. The availability of the ATC will alleviate the overcrowding in the Heidelberg main building and will make some of the work that was originally planned in refurbishment phase II unnecessary or less costly.

EMBL Scientific publications and collaborations 2005

- Total number of peer-reviewed publications: 517
- Internal collaborations: publications co-authored by more than one EMBL Group Leader: 47
- External collaborations: 649 in total of which 122 resulted in a publication

The Carl-Ivar Brändén Building

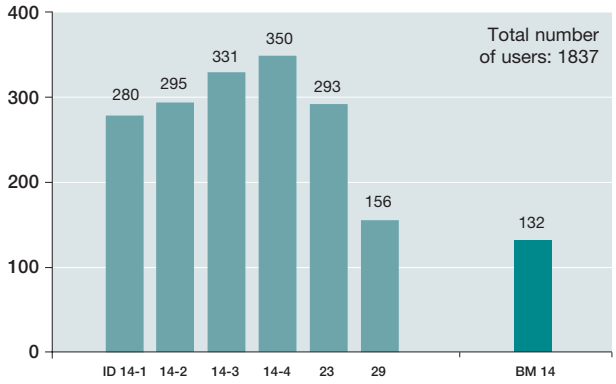
State of the Laboratory

EMBL Grenoble

We and our Grenoble partner organisations opened the Carl-Ivar Brändén Building in January 2006. It is shared by the Partnership for Structural Biology (PSB) [which is formed by EMBL, the European Synchrotron Radiation Facility (ESRF), Institut Laue-Langevin (ILL) and the Institut de Biologie Structurale (IBS)] and the Institut de Virologie Moléculaire et Structurale (IVMS). Building on the strengths of its partners, the PSB is developing into the major focused centre for structural biology in Europe. The new building combines groups and equipment from all these institutions, and houses activities such as high-throughput protein expression and purification, high-throughput crystallography, electron microscopy, and deuterium labelling. PSB Science Days have been established to bring scientists from all partner institutes together on a regular basis to exchange information and discuss new ideas.

More than 1,800 scientists used the Grenoble facilities in 2005: 373 peer-reviewed papers were published that acknowledged use of beamlines operated by EMBL together with ESRF in the Joint Structural Biology Group (ID14, ID23, ID29) and 51 publications acknowledged use of BM14, which is a UK MRC-EMBL Collaborate Research Group beamline. Services provided with EMBL participation in Hamburg and Grenoble together accounted for approximately 80% of the macromolecular structures solved in Europe in 2004 and the first half of 2005. The use of the neutron source at ILL for life-sciences applications is also supported by EMBL, albeit to a smaller user community.

Grenoble Beamline Users 2005



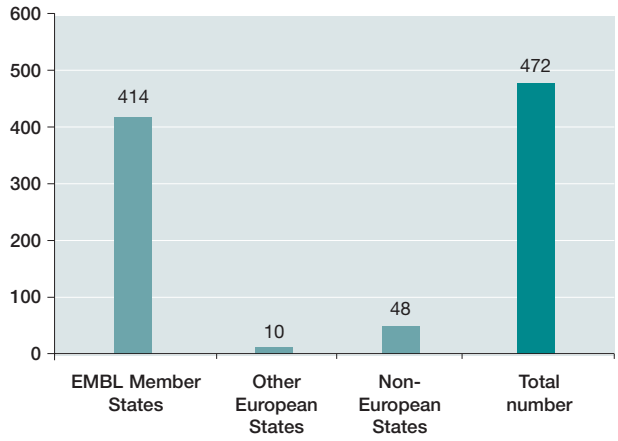
JSBG beamlines (ID14-1 to 4, ID23, ID29) and CRG with MRC (BM14)

Publications quoting use of beamlines ID14-1 to 4, ID23, ID 29 and BM14	
BM 14	51
ID 14-1	100
ID 14-2	71
ID 14-3	25
ID 14-4	68
ID 23	27
ID 29	81

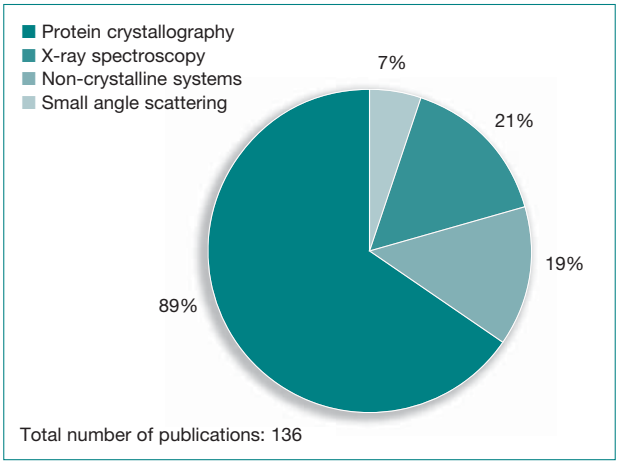
EMBL Hamburg

In November 2005 we opened a new high-throughput crystallisation facility in the Hamburg Outstation. The new facility was made possible by major funding from the German Ministry for Science and Education and is the first such facility to be made available to a broad user community in Europe. It will help to increase the success rate of many structural biology projects and will shorten the time it takes to solve a structure. This is a further step towards building an integrated centre for structural biology in Hamburg.

Hamburg Facility Users in 2005 by nationality



Publications in 2005 acknowledging EMBL Hamburg beamlines



We operate seven beamlines that depend on synchrotron radiation from the DORIS-III storage ring. The beamlines are used for protein crystallography, small angle X-ray scattering and X-ray spectroscopy.

EMBL Monterotondo

Alliances with other European academic research and clinical centres have established EMBL Monterotondo as a hub for the international mouse research community. The Outstation's participation in several EU-wide mouse research and information initiatives serves to link genetics/genomics, phenotyping, physiology and translational activities. Naturally converging interests between the Mouse Biology groups include new collaborative explorations in developmental neurobiology, genetic and pharmacologic manipulations of adult behaviour, inflammation and regenerative processes as well as stem-cell biology. The new, state-of-the-art animal facility provides a full range of mouse transgenic and gene-knockout production, rederivation and cryo-preservation services both to EMBL groups and to external research groups. Other centralised core facilities include a behaviour phenotyping suite.

A newly formed collaboration with the Magdi Yacoub Institute in Harefield, UK, and Imperial College London provides exciting opportunities for the application of basic research in the mouse to advanced diagnosis and treatment of certain types of human heart disease. The collaboration will initially run for four years and the scientific focus will be to study the molecular mechanisms that lead to heart failure and to investigate the molecular and cellular basis of new therapies that have been developed at the Yacoub Institute's Heart Science Centre which have been effective in reversing damage after heart failure.

EMBL-EBI

The EBI continues to host Europe's major core biomolecular resources – collecting, archiving, annotating and distributing data throughout Europe and beyond. In 2005 we received and processed 12 million EMBL-Bank entries, including more than 70 new genomes, as well as data from environmental sequencing projects; 1.5 million UniParc entries; 17,000 microarray hybridisations; 6,000 macromolecular structures; and 8 new eukaryotic genome sequences in Ensembl.

The EBI's services are used more and more: in 2005, on average, 157,000 unique hosts were served per month. By the end of 2005 there were, on average, 2.2 million requests per day compared with 1.4 million at the end of 2004. The EBI continues to provide extensive training for users of our services, organising 70 workshops in 2005 and many software demonstrations and tutorials at conferences.

Funding from EMBL, the Wellcome Trust and two UK Research Councils (MRC and BBSRC) has been secured for a 2000-m^2 extension. Construction started in January 2006 and is due to be completed in mid-2007. The extension will provide space for additional staff and includes a new IT training room.

EMBL Heidelberg

Research

A major characteristic of the new activities that developed in Heidelberg this year is the size and ambition of several projects. Examples include work that is coordinated in the new Centre for Computational Biology, which involves groups from the Developmental Biology and Structural and Computational Biology Units and the EBI. The goal is to develop a system for describing and annotating gene expression patterns in such a way that they can be digitised for comparison and searched across evolutionary time. A second example involves many groups from the Structural and Computational Biology Unit and both the Hamburg and Grenoble Outstations. The goal of this project is to purify and characterise protein complexes, with a view to moving towards a high-resolution description of a cell. Groups from the Cell Biology and Biophysics Unit and the Gene Expression Unit, with help from the Genomics Core Facility, have set up a high-throughput, high-content light microscopy-based screening system and are using it to analyse the contributions of genes to the human cell cycle on a genomic scale. Both this project and the protein-complex analysis described above depend heavily on external, EU support. This highlights a challenge for EMBL, which is not organised in such a way that expensive, labour-intensive projects are easy to support or to maintain using internal resources. One response to this will be the EMBL Centres, but we will need to consider the general question of how large-scale projects, which are often transitory, can be housed and staffed in EMBL.

Refurbishment update

The refurbishment of the main Laboratory in Heidelberg is proceeding as planned. As part of phase I, we have completed a major renovation of the sanitation facilities, renewed a large part of measurement and control technology, reinforced electrical substations, replaced cooling systems, replaced and expanded ventilation systems, and installed fire detection and video surveillance systems. Building refurbishment that has been completed includes: upgrading of fire protection, renovation of heating substations and replacement of elevator machinery.

In 2006 we are planning to complete, also as part of phase I of the refurbishment, the expansion of the steam boiler system, replacement of waste autoclave and renovation of level 6 in the main building.

EMBL Heidelberg

Core Facilities survey

The Core Facilities were established during the current Scientific Programme and have been extremely popular with EMBL scientists. Due to the limited size of research groups at EMBL, a mechanism for providing access to large specialised equipment, such as advanced light microscopes or robotics, is essential. As of last year we had not formally analysed the performance of the Core Facilities and therefore it was decided to organise a formal Scientific Advisory Committee review (see below for my response to the review report). Also, in preparation for the next EMBL Programme, we decided to survey the users of the facilities to evaluate their level of satisfaction and ask them about future needs. The online survey was carried out in mid-2005 and involved a large number of EMBL users.

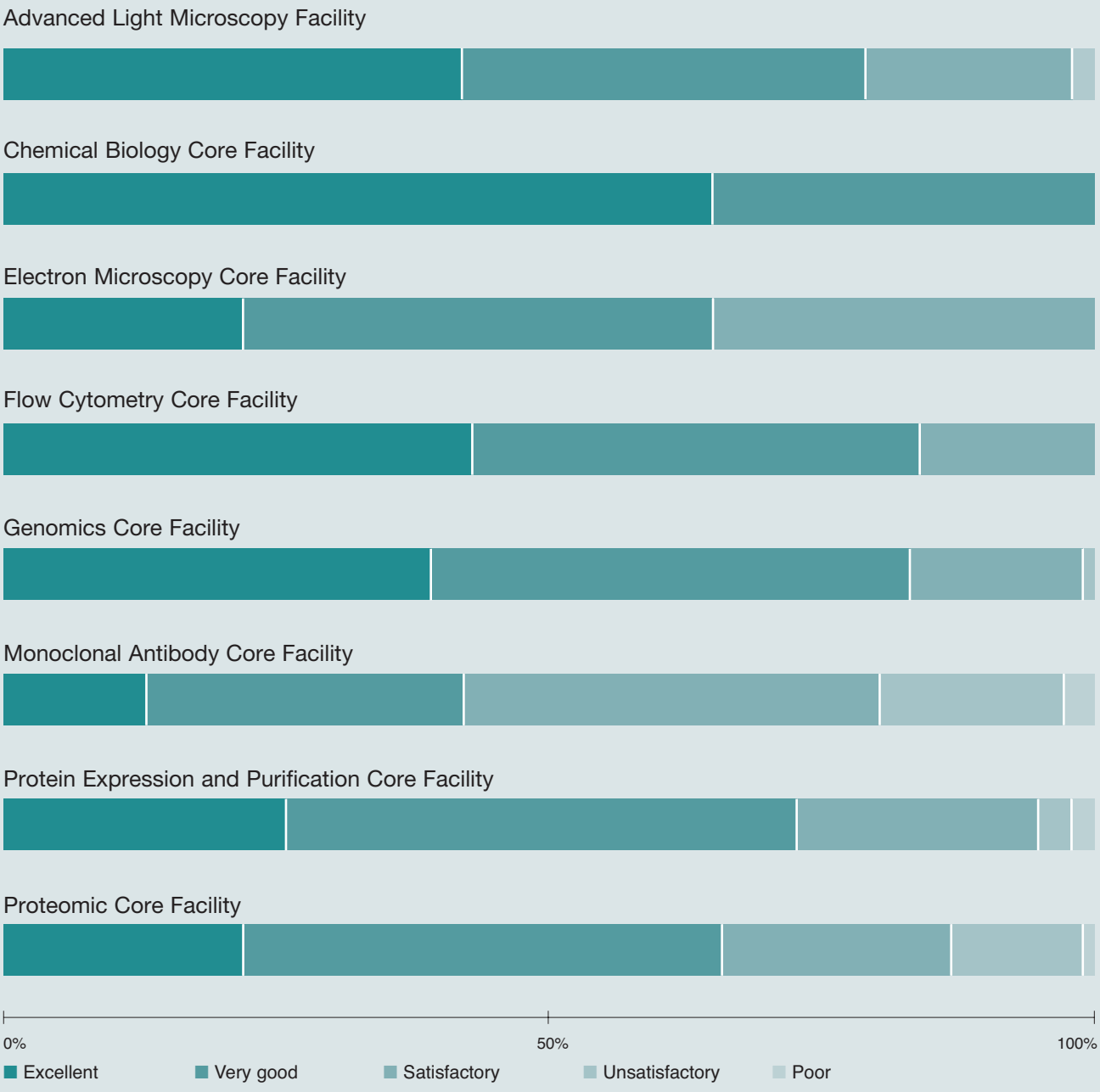
The goal of the survey was to answer the following questions:

- Do our facilities match the requirements of the research carried out at EMBL?
- Is usage of existing facilities appropriate?
- Are the users satisfied with the services provided?
- What is missing?

Participation in the survey was extremely high: 70% of the consulted personnel responded (470 responses to 678 requests). The answers came from all Units and Outstations and reflected the distribution of staff at the different sites.

An important part of the survey dealt with the question of overall satisfaction with each Core Facility. Questions covered accessibility of the facilities, results, equipment, proposed methods, advice to users from the staff of the facility, support during projects, friendliness of the staff, delivery speed and quality of results, monitoring of project status and quality of information available on websites. In total, the survey provided a considerable quantity of valuable information that could be compiled easily as the answers were collected in an electronic format. The overall measure of the satisfaction of users of the various Core Facilities is shown below.

User satisfaction with EMBL Core Facilities



Partnerships

The EMBL partnership scheme, which was introduced in 2001 to form better links between EMBL and specific research institutions in EMBL member states, has become very popular. We have extended the lifetime of the first partnership, the Molecular Medicine Partnership Unit, with the University of Heidelberg. The Partnership for Structural Biology in Grenoble moved into the new Carl-Ivar Brändén Building, which is shared with the Institut de Virologie Moléculaire et Structurale. We are supporting the development of new partnerships in molecular medicine in Finland, Norway and Sweden which will become nodes of a Nordic Centre for Molecular Medicine. A partnership in marine biology is being set up in France.

EMBL/CRG Partnership

We have recently signed a cooperation agreement with the Spanish Ministry of Education and Science and the Centre for Genomic Regulation (CRG) in Barcelona to establish the EMBL/CRG Systems Biology Research Unit there. The Unit will be modelled on EMBL in terms of scientific management, evaluation and staff turnover and will include four multidisciplinary groups led by young scientists who will be recruited internationally. Luis Serrano, who is currently Coordinator of the Structural and Computational Biology Unit at EMBL Heidelberg, but who will move to the CRG in 2006 will be in charge of the partnership Unit. The Unit has been established for nine years and is funded by the Spanish Ministry of Education and Science.

Molecular Medicine Partnership Unit (MMPU)

Almost a year ago, shortly after the summer Council meeting in 2005, EMBL signed a new agreement with the University of Heidelberg and its medical faculty to extend the Molecular Medicine Partnership Unit for a further ten years. The partnership started three years earlier as a small pilot project organised by Matthias Hentze and Andreas Kulozik. The MMPU was very positively reviewed by an independent committee in early 2005. It will now be expanded and we hope that the medical faculty will be able to provide common space and facilities in the mid-term. The number of collaborative projects has been expanded to four at present and includes EMBL groups from the Gene Expression and the Structural and Computational Biology Units. The MMPU groups carry out basic research related to disease in projects covering cystic fibrosis, iron metabolism disorders, identification of tumour markers and mRNA metabolism defects. The MMPU plans to grow by including additional groups and projects.

EMBL Partnership with Sars Centre, Bergen, Norway

The partnership with the Sars Centre has resulted in several collaborations involving groups in the Developmental Biology Unit centred on questions of evolution and development. EMBL members of the Sars scientific advisory board have helped and supported their successful efforts in pursuing international recruitment and scientific evaluation. When I visited Sars soon after taking office I was very impressed by the quality and variety of the research being carried out there.

New Member State: Croatia

We welcome Croatia, which joined EMBL this year and is our first member state from Eastern Europe. The initial contact between EMBL and Croatia came at the initiative of the EMBC, through a joint visit to the research ministry organised by the Croatian EMBC delegate, the EMBO Executive Director and myself. Croatia has strength in epidemiology and disease studies, and has ambitious plans for supporting its life-science research community.

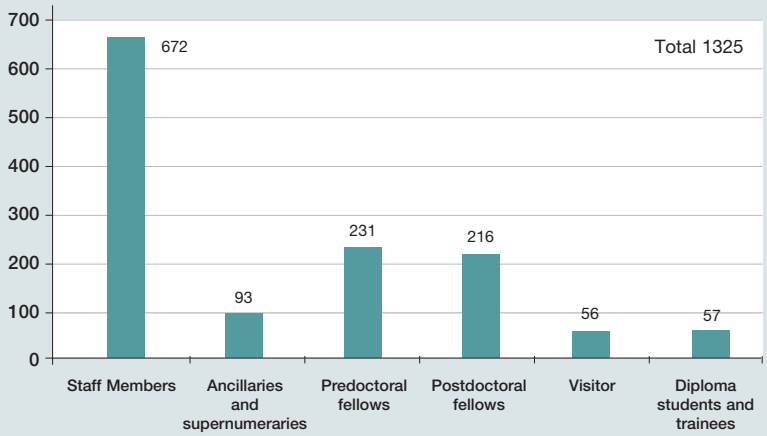
International collaborations

After an approach from the National Institute for Natural Sciences (NINS), EMBL agreed to participate in a collaboration with their National Institute for Basic Biology (NIBB) in Okazaki, Japan, to promote interactions between European and Japanese scientists. Activities to date have included organising a series of joint NIBB/EMBL meetings that alternated between EMBL and Japan. These meetings have been very successful and popular among scientists from both institutes, and will continue. The NINS has very generously provided travel grants for EMBL scientists to attend events in Japan and this has made possible several visits to NIBB of group leaders as well as PhD students and postdoctoral researchers. The Shimura Awards, named after the current president of the NINS, Yoshiro Shimura, will help to bring more EMBL scientists to Japan and to encourage collaborations and exchange of information.

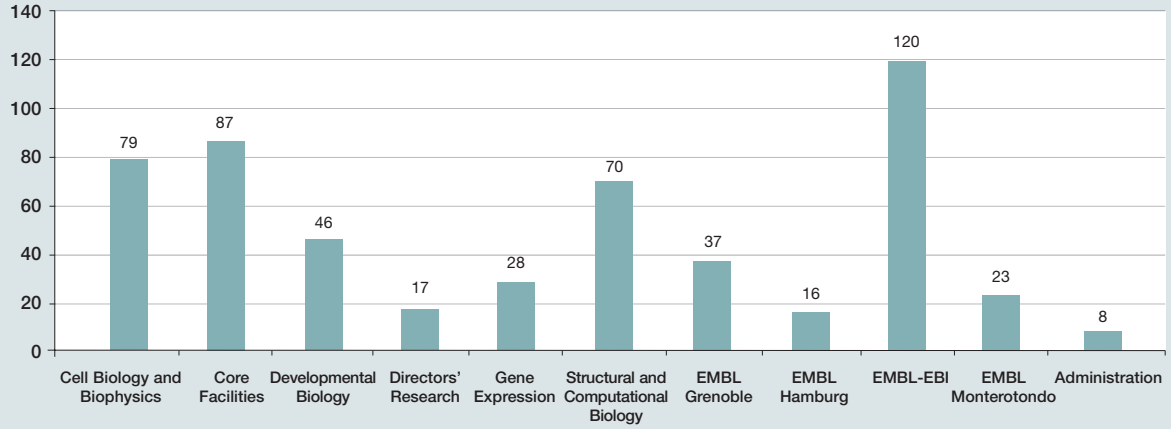
Personnel Statistics

On 31 December 2005, 1325 people from more than 60 nations were employed by EMBL. 82% were from EMBL Member States and 43% were female.

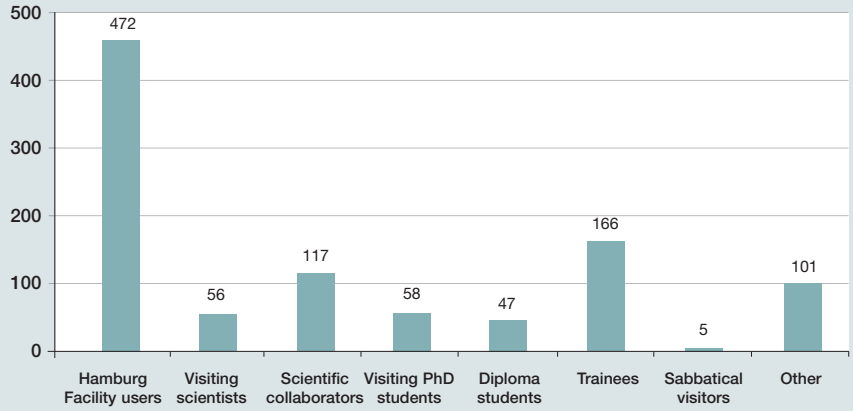
Personnel on 31 December 2005



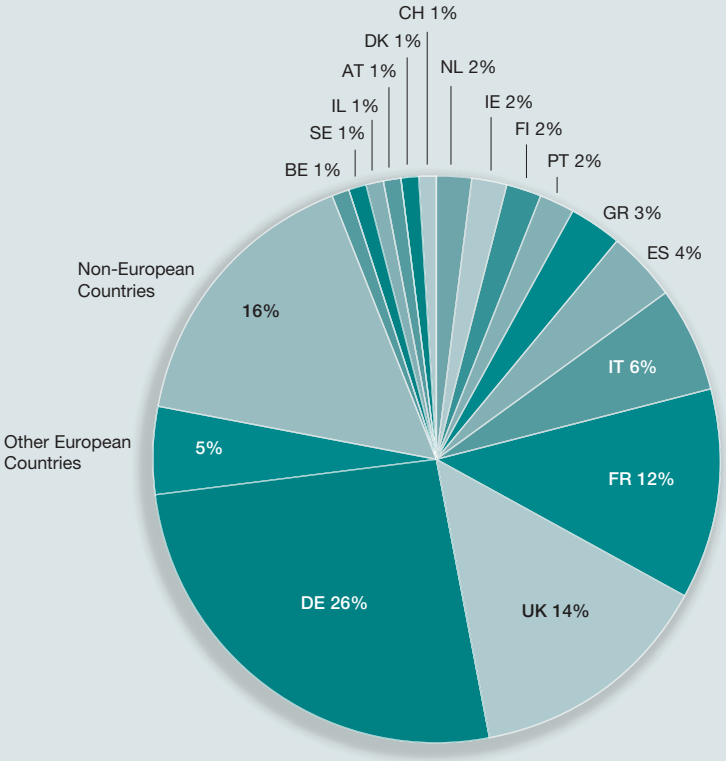
Visits to EMBL Research Units during 2005



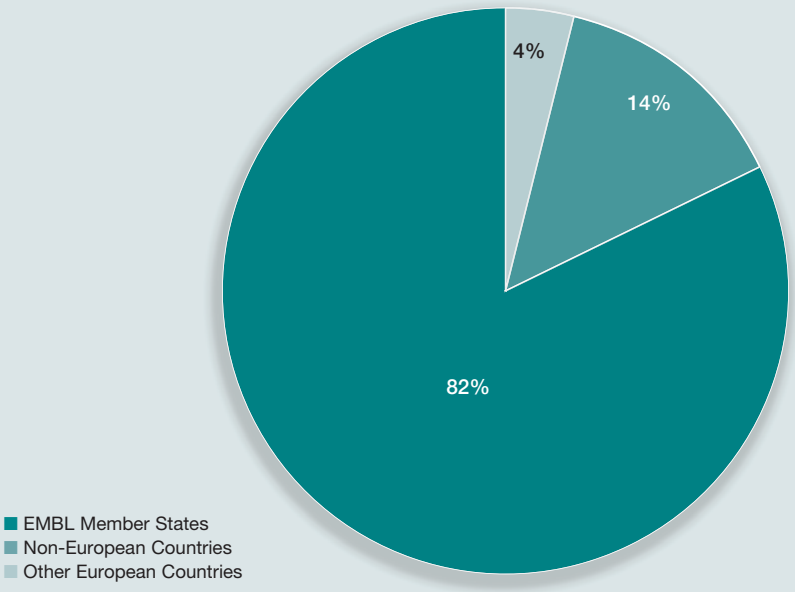
Visitor Types during 2005



Staff Nationalities – Research



All EMBL staff nationalities



Financial report

External Funding

	2005		2004	
	k Euro	*%*	*k Euro*	*%*
EU	12,801	47.0	12,844	47.1
NIH	3,773	13.9	2,808	10.3
DFG	1,947	7.2	1,453	5.3
Wellcome Trust	1,766	6.5	2,107	7.7
BMBF	1,496	5.5	2,162	7.9
BBSRC	913	3.4	316	1.2
Swissprot	867	3.2	638	2.3
VW Foundation	393	1.4	212	0.8
MRC	371	1.4	513	1.9
HFSPO	362	1.3	691	2.5
Other	2,523	9.3	3,503	12.9
TOTAL	**27,212**	**100**	**27,247**	**100**

Member States Contributions

	Ordinary and one-off contributions				Pension contribution	
	2005		2004		2005	2004
	Euro	*%*	*Euro*	*%*	*Euro*	*Euro*
Austria	1,408,712	2.22	1,407,460	2.31	25,463	24,879
Belgium	1,725,990	2.72	1,699,919	2.79	31,198	30,048
Denmark	1,097,780	1.73	1,096,722	1.80	19,843	19,386
Finland	869,341	1.37	785,984	1.29	15,714	13,893
France	10,108,463	15.93	9,974,076	16.37	182,717	176,305
Germany	13,922,139	21.94	14,476,730	23.76	251,652	255,895
Greece	932,796	1.47	913,935	1.50	16,861	16,155
Israel	774,157	1.22	645,848	1.06	13,993	11,416
Italy	8,249,217	13.00	7,993,885	13.12	149,110	141,302
Netherlands	2,773,006	4.37	2,668,690	4.38	50,124	47,173
Norway	1,262,765	1.99	956,585	1.57	22,825	16,909
Portugal	736,084	1.16	700,683	1.15	13,305	12,386
Spain	4,429,195	6.98	4,088,335	6.71	80,061	72,267
Sweden	1,687,917	2.66	1,602,434	2.63	30,510	28,324
Switzerland	2,074,995	3.27	2,041,121	3.35	37,507	36,080
United Kingdom	11,402,956	17.97	9,876,590	16.21	206,116	174,582
SUB TOTAL	***63,455,512***	***100.00***	***60,928,997***	***100.00***	***1,147,000***	***1,077,000***
Ireland	455,488		376,844		8,258	6,688.00
Special Contribution Ireland	146,394		146,394			
Iceland	40,118				721	
Special Contribution Iceland	16,396					
One-off contribution Germany	1,194,564					
TOTAL CONTRIBUTIONS	**65,308,472**		**61,452,235**		**1,155,979**	**1,083,688**

Income/expenditure statement

INCOME	2005	2004
	k Euro	*k Euro*
Member State Contributions	65,309	61,452
Internal Tax	16,824	14,905
External Funding	27,212	27,247
Other Receipts	9,021	8,603
Total Income Excluding pension contributions	**118,366**	**112,207**

EXPENDITURE	2005	2004
	k Euro	*k Euro*
Staff Costs	71,031	65,798
Operating costs	29,826	30,827
Capital	11,326	10,044
Special Capital Investment	3,836	1,306
Total Expenditure	**116,019**	**107,975**
Surplus transferred to reserves	2,347	4,232

In 2005, EMBL scientists coordinated 14 Framework Programme projects funded by the EU:

Project (Instrument)	**Coordinator**	**EMBL participants**	**Description**
3D Repertoire (IP)	L. Serrano	B. Böttcher, P. Bork, E. Conti, R. Russell, M. Sattler, K. Scheffzek, D. Suck, D. Hart, C. Müller, M. Wilmanns	A multidisciplinary approach to determine the structures of protein complexes in a model organism
BioSapiens (NoE)	J. Thornton	A. Brazma, C. Ouzounis, E. Birney, K. Henrick, R. Apweiler, P. Bork	A European network for integrated genome annotation structural genomics
BIOSTAR (EST)	J. Thornton		Early-stage-training
BIOXHIT (IP)	V. Lamzin	M. Weiss, W. Meyer-Klaucke, P.Tucker, R. Ravelli, K. Henrick	Biocrystallography on a highly integrated technology platform for European structural genomics
CISB (IF)	S. Cusack		The Centre for Integrated Structural Biology
Combio (STREP)	L. Serrano	F. Nédélec, I. Vernos	An integrative approach to cellular signaling and control processes: bringing computational biology to the bench
EMBRACE (NoE)	G. Cameron	P. Rice, T. Gibson, A. Bleasby	A European model for bioinformatics research and community education
ENFIN (NoE)	E. Birney	J. Ellenberg, H. Hermjakob	European rat tools for functional genomics
E-Star (EST)	A. Ephrussi		Early-stage training in advanced life science research across Europe
FLPFLEX (STREP)	N. Rosenthal		A flexible toolkit for controlling gene expression in the mouse
NETSENSOR (STREP)	L. Serrano		Design and engineering of gene networks to respond to and correct alterations in signal transduction pathways
Saxier (SSA)	D. Svergun		Small angle X-ray scattering at high brilliance European synchrotron for bio- and nano-technology
SYMBIOmatics (SSA)	G. Cameron		Synergies in medical informatics and bioinformatics
TEMBLOR (QL)	G. Cameron	R. Apweiler	The European Molecular Biology linked original Resources

2005/2006 Reviews of EMBL Scientific Units

EMBL Units are reviewed in depth every four years by expert international panels organised by the Scientific Advisory Committee. To ensure openness, the review reports are submitted in confidence to EMBL Council and the Director General. The formal responses of the Director General to the reports are made public, to communicate the adjustments made by the Laboratory in response to the reviews, when needed.

Director General's Response to the EMBL Cell Biology and Biophysics Unit External Review Report, Heidelberg, 12-13 May 2005

1. This detailed and differentiated report from a panel of recognised experts will be extremely useful to us in planning the future of this important EMBL Unit. We are very grateful to the panel and its Chair for their efforts and advice.

2. The panel appreciated the creative and thoughtful scientific leadership of Eric Karsenti, and acknowledged his role in providing an original scientific direction for the Unit. It expressed enthusiasm for his decision to stay at EMBL following an attractive offer elsewhere, but noted that the period of uncertainty surrounding this decision had influenced organisational aspects of the Unit, and urged that this should be corrected. We concur with both the positive evaluation of the role of E. Karsenti and with the need for some tightening of the Unit's organisation.

3. The panel noted the major efforts of the Unit in technology development, particularly in the areas of light microscopy, modelling and simulation, and regarded the combination of technology development with basic research in the Unit as a strength. We agree with this evaluation and will be guided by it when deciding on future priorities.

4. Members of the Unit (Rainer Pepperkok, Claude Antony) were praised for their roles in providing core facilities for light (Advanced Light Microscopy Core Facility; ALMF) and electron (Electron Microscopy Core Facility) microscopy to EMBL scientists, and to a broader community. The panel recommended that the ALMF should investigate the incorporation of a fluorescence lifetime imaging microscope, developed by Philippe Bastiaens, into the ALMF, to facilitate access to this technology. It further recommended analysing the case for expanding the EM facility to incorporate medium-resolution EM tomography capability. We will investigate the demand for, and cost of, these recommendations in relation to the next Scientific Programme.

5. A significant number (6) of the groups in the Unit study the behaviour of the microtubule cytoskeleton and its associated factors. Although they do so using a great variety of approaches, the panel suggested that future recruitments should be made with a view to broadening the focus of the Unit. We agree with this recommendation and will act on it.

6. The panel noted that some of the younger group leaders in the Unit had important roles within the broader EMBL context and contrasted this favourably with the roles played by some of the more senior faculty members. We will take these comments into account in future decisions and planning for the Unit.

Director General's Response to the EMBL Core Facilities Review Report, Heidelberg, 23- 24 March 2006

1. The review panel carefully examined the structure, development and performance of the Core Facilities. Their opinions and advice are invaluable to EMBL and we are very grateful for their expert input.

2. The panel endorsed the principles adopted when EMBL set up its Core Facilities. They agreed that the Facilities should continue to be service-driven and that evaluation of the staff should be based on service provision and not research output.

3. This was the first external review of the Core Facilities, and the panel provided several valuable suggestions for future reviews. These include more direct interaction between future panels and representative users, e.g. the chairs of the user committees, and the collection and provision of information on the contribution of the facilities to EMBL's scientific output. This advice will help us to further improve the review process.

4. The panel pointed out the value to EMBL of the Core Facilities not only in providing support to existing scientific staff but also in making EMBL more attractive to the best potential new recruits.

5. The panel praised all the Heads of Core Facilities and, in particular, the Unit Coordinator Christian Boulin, for their commitment to setting up scientific services of the highest quality and their motivation to continue to improve the performance of the Core Facilities.

6. The panel recommended regular meetings between the individual Facility Heads, the Unit Coordinator and the user committees. These meetings should help the Core Facilities to set priorities and to develop strategies for selecting future activities and investments. The panel recommended that the outcome of such meetings should be recorded and might form part of future reviews. We will follow this advice.

7. The panel supported the interaction of the Core Facilities with external users but recognised and recommended that in-house users should continue to have priority of access. Interaction with external scientists, with a view to advising member state institutes who wish to set up their own Core Facilities was judged to be particularly valuable.

8. The panel discussed and provided advice on several issues related to Core Facility funding:

- They emphasised the requirement for ongoing investment in equipment for the Core Facilities if they are to remain on the cutting edge.
- They advised that the past and present levels of success in partnering with companies who provide equipment, which they felt was highly successful and deserved praise, may be difficult to sustain. This should not influence the necessary upgrading of equipment.
- They pointed out that the large quantity of data generated by some of the Core Facilities requires the provision of additional computational storage and handling capacity. Expertise in the interpretation of high-throughput or high-complexity data is also limiting.
- They emphasised the need for careful prioritisation of activities and equipment purchase both within and between Core Facilities.

All of these comments underline the panel's message that the Core Facilities are extremely valuable to EMBL and that we must think strategically about how best to support and maintain them.

Scientific Report

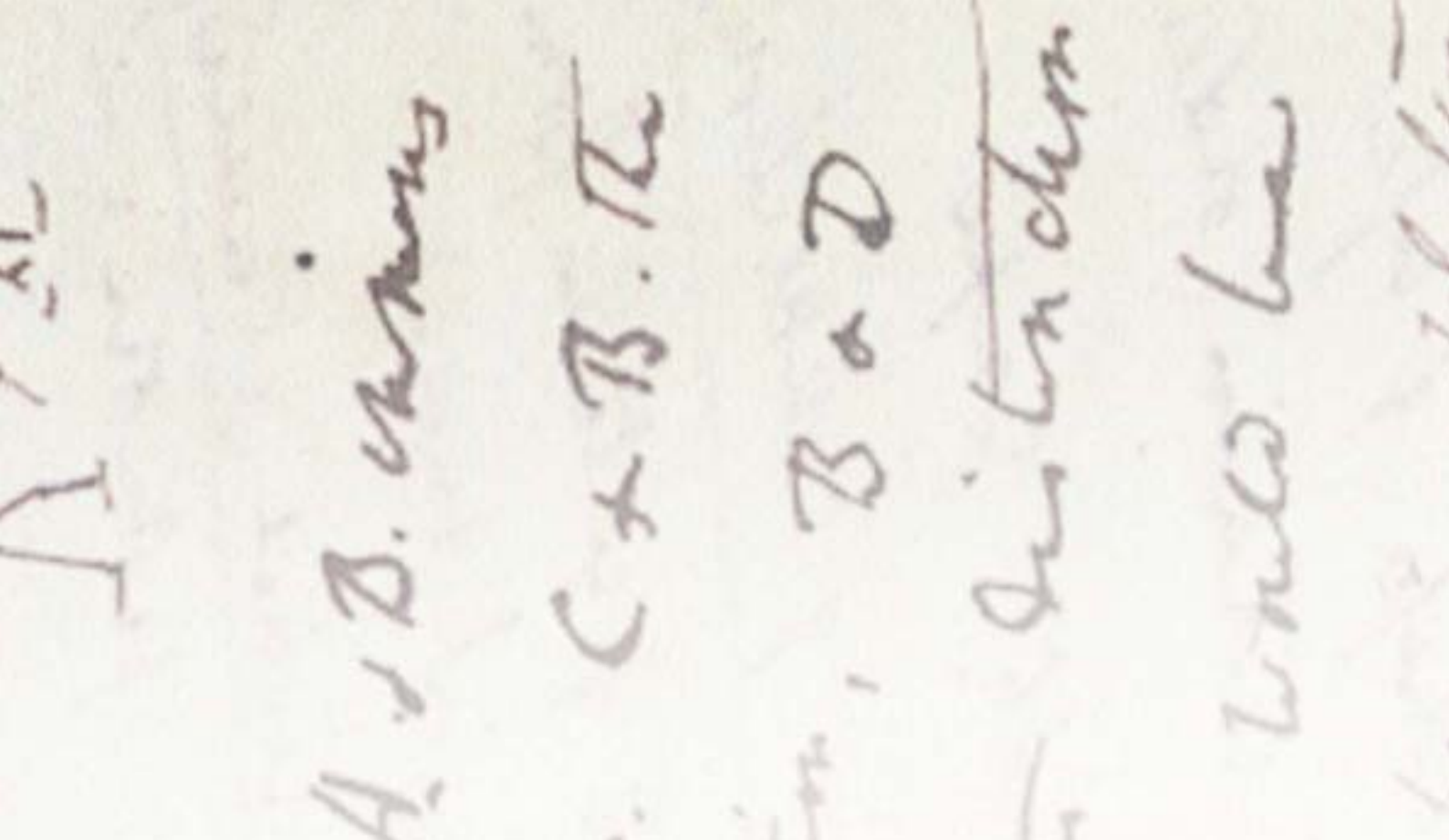
One long argument

"This whole volume is one long argument," wrote Charles Darwin at the end of *The Origin of Species*, summarizing over twenty years of work and 450 pages of facts, observations and well-founded speculations. Ideally, the book in your hands would also be one argument (although not quite as long!). But given the huge range of questions that today's molecular biologists are investigating, with the help of methods from many different disciplines, that is an enormous challenge. EMBL scientists are working on some of the smallest phenomena in biology (how the positions of single atoms influence the behavior of molecules) and some of the largest (the fates of species over vast stretches of evolutionary time), as well as most things in between. They would like to draw all of these levels of life into a single, unified system. This is still beyond the grasp of the life sciences, but it provides a good standpoint from which to understand what has happened at EMBL over the past year.

Annual reports can be thought of as milestones as research moves towards this goal of unification. Like markers that line a country road, they are equally spaced, but science is flying by at an ever-increasing pace. The rate of change and the accompanying shifts in our perspectives on life naturally raise concerns about where things are headed and whether society can keep up, particularly since this type of science is a recent historical development. Nearly everything that is happening today can be traced directly back to a series of discoveries made about 150 years ago, and a brief glance backwards can help put them into perspective.

In the summer of 1859, as Darwin finished his book, most of the world had its eyes turned upwards: massive solar storms illuminated the night sky with spectacular Northern Lights. They could be seen as far south as Rome, with a bizarre effect on telegraph lines. For hours at a time, operators were able to send messages without using any batteries or power; at other times the lines couldn't be used at all, or they burst into flames. These odd events snatched headlines everywhere, although the real news was happening much more quietly, in laboratories across the continent, where scientists were launching a revolution that would lay the cornerstones of modern biology.

Evolution was one of these, but there were important others. A tremendous leap forward in microscopy had just enabled two German scientists to prove that plants and animals were made up of cells – the birth of cell biology. This radically changed the way people understood the growth of embryos and would eventually lead to what we call developmental biology. In France, Louis Pasteur was creating modern medicine with his demonstration that microorganisms were responsible for disease and did not arise through "spontaneous generation." This meant that life might be understandable as the result of mechanistic processes. Chemists across the continent were learning to break down organic substances into more basic elements, opening the field of biochemistry. In 1858 Archibald Couper drew the first model of a molecule, demonstrating that it wasn't enough to know what substances were made of – one also had to understand how their atoms were arranged (the spirit of today's structural biology).

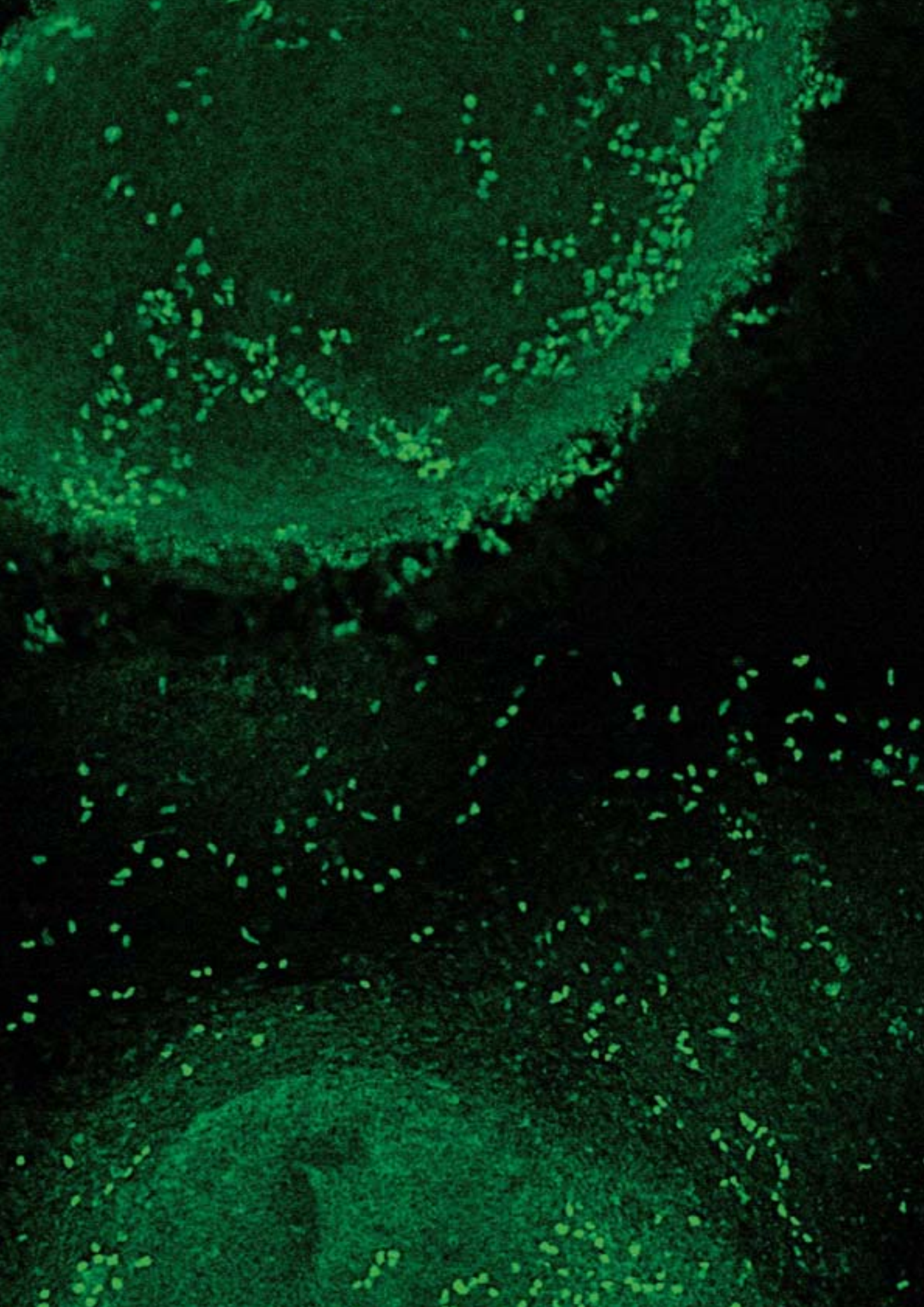


And farther to the east, Gregor Mendel was carefully breeding peas in a monastery garden. In doing so, he was not only laying the foundations of genetics, he was also taking the unorthodox step of applying mathematics and statistics to a biological question. Today that approach is echoed in bioinformatics and modeling.

The pioneers of these fields came from different backgrounds – sometimes they had never heard of each other – and were largely unable to foresee the impact that one discovery would have on the other areas. Over the past century, these strands of science have become intertwined, usually in twos or threes, into a long braid. Genes were identified as the basis of heredity and were linked to evolution. Chemistry and physics led to the discovery of the nature and functions of proteins and DNA, and that permitted dissecting the biochemistry of the cell. Studies of flies showed how genes determined the features of adult organisms, and later how they directed the development of embryos.

EMBL has units with names that reflect their origins and their predominant domains of activity. Yet today all of these strands of thought – evolution, biochemistry, genetics, cell biology, biophysics, structural biology, development and medicine – are finally being drawn together. It is still rare that all the strands are woven into a single, rope-like explanation of a biological problem, but when today's scientists speak of understanding life, this is what they mean. All the levels of biological organization are inherently linked in organisms' everyday lives and their lifelong development. The same is true of genetic diseases, where flawed information in the genetic code works its way up through RNAs and proteins, the complexes and structures they form in cells, tissues and organs, and finally the fates of organisms. Treating such diseases will require a thorough understanding not only of the rules of each level of organization, but also of how they are nested within each other, how much freedom can be tolerated at one scale without disrupting higher functions, and the dialogue between these systems and the environment.

"Systems biology" is the attempt to understand complex and multi-leveled biological problems through combinations of models and experiments. Such approaches are spreading through the Laboratory, which is reflected in EMBL's plan for the upcoming years. However, before unification becomes the standard method of operation across the life sciences, and before we understand the effects that the manipulation of molecules have on an organism's holistic well-being, a great deal remains to be done in all of the sub-disciplines. Whenever possible, that should be done in a highly-interdisciplinary way, with a view to the whole and with care for the societal implications of science. The goal of this report is to show how this works on a practical level at EMBL, and how the Laboratory's particular way of organizing and practicing science across its five units is moving us steadily towards this vision. ■



The powers of proteins

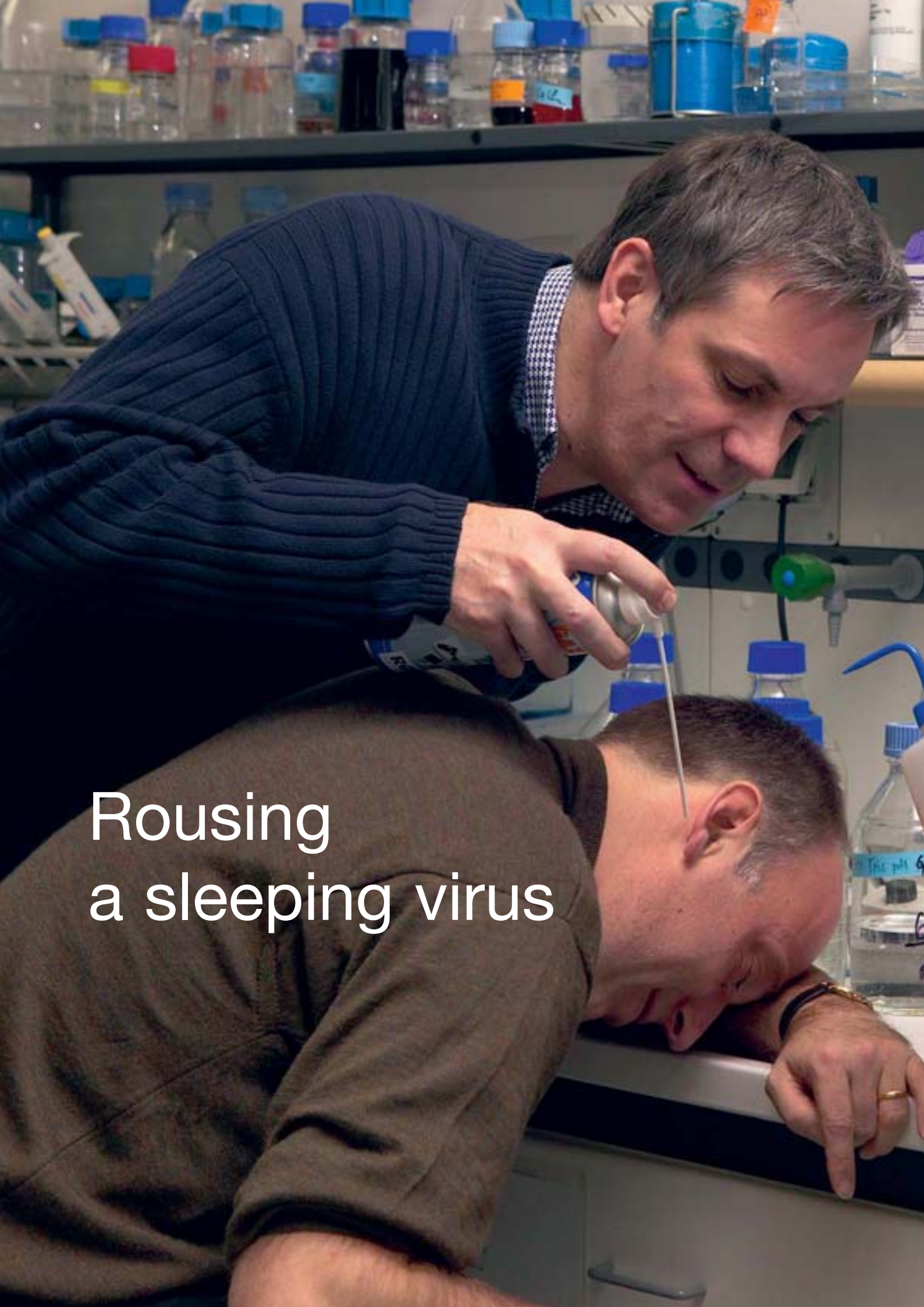
IT WASN'T VERY LONG AGO that many scientists could build their reputations on work with a few important proteins: exposing their roles in the biochemistry of the cell, the other molecules they worked with, or their structures. Limitations in technology usually required a molecule-by-molecule approach to the cell; small battles of discovery were hard-won; certain proteins acquired an inflated reputation because of the clear roles they play – and because their functions could be grasped with the methods at hand.

This underplays the fact that nearly all molecules accomplish their tasks by working in machines and networks. New technologies are giving biologists a look at this more complex level of organization, but interest in the individual protein hasn't waned. Complex phenomena are built on the physical and chemical characteristics of specific molecules, and step-by-step work is still necessary to uncover their functions in biochemical pathways, communication networks, the cell cycle, development and other processes.

All proteins aren't created equal. When certain genes (and thus the proteins they encode) are deleted from the cell, the result is sudden death; others can be removed with no effect on an organism at all. A single mutation may lead to cancer. A virus' ability to reproduce may hinge on one protein.

This chapter of the annual report deals with molecules that are significantly connected to large-scale events such as the formation of tissues or the onset of disease. Here is a view of the cell from the bottom up, from the perspective of single molecules. The later section on "Proteomics" looks at some of the same biological processes from the top down, from the angle of networking and more general principles of their interactions. In years past, linking these perspectives was often impossible. That is changing; most of the studies in this section could serve as an introduction to proteomics, or to a systems approach to biology, because each manages to build a link between single molecules and the global systems in which they operate. For example, two of the studies show how interacting molecules help generate self-organizing structures in the cell and what happens to cells when the molecules change. A systems view of life will have to reveal the links between molecules and entire organisms, and the cell is the gateway to understanding them. Although we frequently know how a mutation alters the design of an animal's body, we are seldom able to trace the defects to clear changes in the behavior of cells. That is starting to change.

Another theme of this section is methodology: two of EMBL's Outstations and several of our Core Facilities are chiefly devoted to exploring new ways of investigating the physical and chemical characteristics of molecules. EMBL has a particular recipe for marrying technology, services and science. It encourages groups to use new methods to expore the most challenging problems, often in conjunction with collaborators and industry; this pushes technology to become better, and each improvement leads to new insights into life. ■



Rousing a sleeping virus

It's not the Epstein-Barr Virus that Patrice Morand and Carlo Petosa are awakening – but Grenoble group leader Christoph Müller.

VERY LIKELY, millions of copies of the *Epstein-Barr Virus* (EBV) slumber in your cells; with luck, they will stay that way. As many as 90 percent of us are infected with the virus, which takes up residence in our immune system's *B cells*. Once there, it usually keeps a low profile, like a houseguest that doesn't want to overstay its welcome by disrupting the host's routine. Every once in awhile, however, EBV rouses itself for a brief spurt of reproduction – allowing the virus to infect new cells, as well as new individuals. Afterwards it usually becomes passive again. Most people are infected with the virus early in childhood. However, if the virus hits later in life, EBV can cause *infectious mononucleosis*, a disease characterized by a general sense of fatigue that can persist for months. More seriously, in individuals with weakened immune systems (like patients who receive organ transplants or have AIDS), EBV can cause the uncontrolled growth of the B cells it infects. This can lead to cancers such as Hodgkin's disease and aggressive lymphomas.

For years the virus has interested Patrice Morand, a physician at the Institute for Molecular and Structural Virology (IVMS), part of the Université Joseph Fourier in Grenoble. Patrice is a clinical doctor with a background in infectious diseases. "Until about 1999 I actively saw patients," he says. "Some of them, particularly those with a suppressed immune system, were having severe problems with the Epstein Barr virus. I was fascinated by its capacity to lie in wait for so long in the cells."

Fascinated but frustrated, he says, because this long latent phase contributes to the difficulty in finding an effective treatment for the disease. Another problem is that EBV specifically attacks cells that could help protect the body. The frustration has led Patrice, like many researchers, to dig below the clinical picture to more fundamental aspects of the virus, such as how it can hibernate for so long, and what wakes it up.

He's in the right place at the right time to pursue these questions. Several groups at EMBL's Grenoble Outstation have a keen interest in viruses, and over the past few years they have strengthened their bridges to the medical research community in Grenoble to try to get a handle on the molecular and cellular mechanisms that underlie disease. Rob Ruigrok, former EMBL group leader, has now become head of the IVMS, which has teamed up with EMBL and the European Synchrotron Radiation Facility (ESRF) to apply the tools of structural biology to investigate EBV and other viruses such as influenza, adenovirus, and Ebola. Together the institutes have formed the Partnership for Structural Biology (PSB), which shares common facilities in a new building on the Grenoble campus (see story on page 124). They are working together in one of the largest structural biology projects funded by the European Union, called SPINE, which has a strong medical focus.

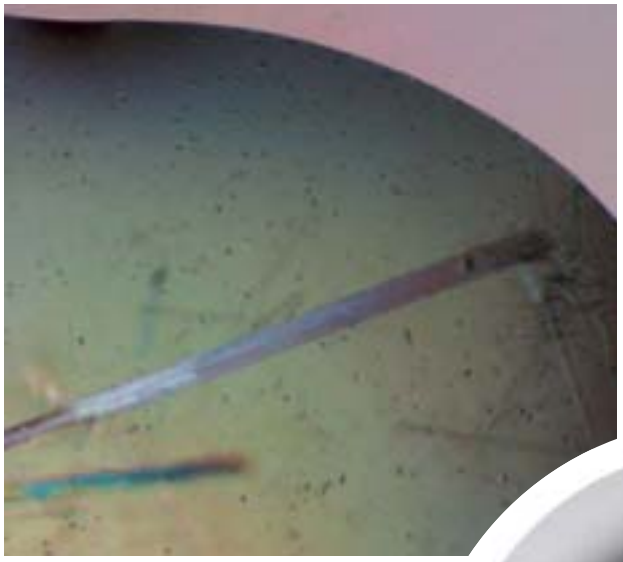
"Rob established the connection in my case," Patrice says. "I began to look at the structure of EBV enzymes with him and Wim Burmeister, who was a PhD student at EMBL and is now a professor at the university. Then we got in touch with Christoph Müller's group at EMBL, because of their expertise in proteins which bind to DNA."

Those proteins include *transcription factors*, which play a crucial role in the life cycle of EBV. Waking the virus involves activating its genes. Upon infecting a B cell, EBV is taken apart and its DNA enters the nucleus of the cell. Transcription factors rouse quiet genes; they often work by entering the nucleus, docking onto specific DNA codes, and calling up helper molecules that use the information in genes to create new molecules.

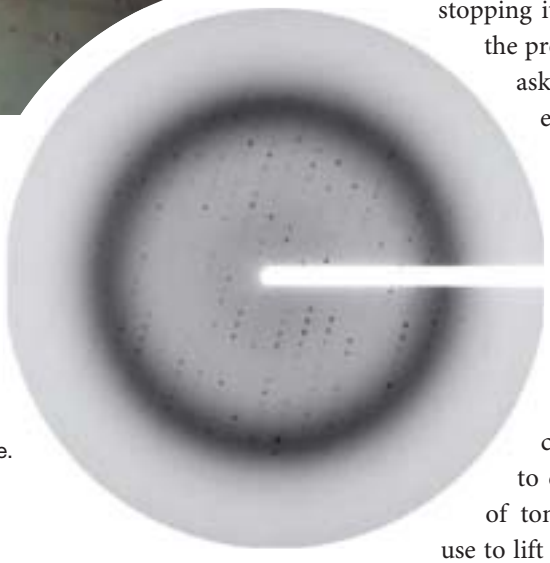
"EBV has a very powerful transcription factor called *ZEBRA* (also known as Zta, EB1 or Z)," Christoph says. "It's an unusual protein because it activates about 50 different EBV genes – over half the total number brought into the cell by the virus."

Those genes are necessary to switch EBV from its inactive to its reproductive mode. "If you could stop ZEBRA, you could prevent any of this information from being used," Patrice says. "That would keep the virus from reproducing and infecting new cells. Furthermore, ZEBRA probably also promotes tumour growth, reinforcing the idea that ZEBRA is an attractive therapeutic target."

* * *



Opposite: EMBL's Grenoble Outstation



Above: Crystals of ZEBRA protein were studied on a Grenoble synchrotron beamline, producing a diffraction pattern (right). This could be interpreted into the structures seen on the next page.

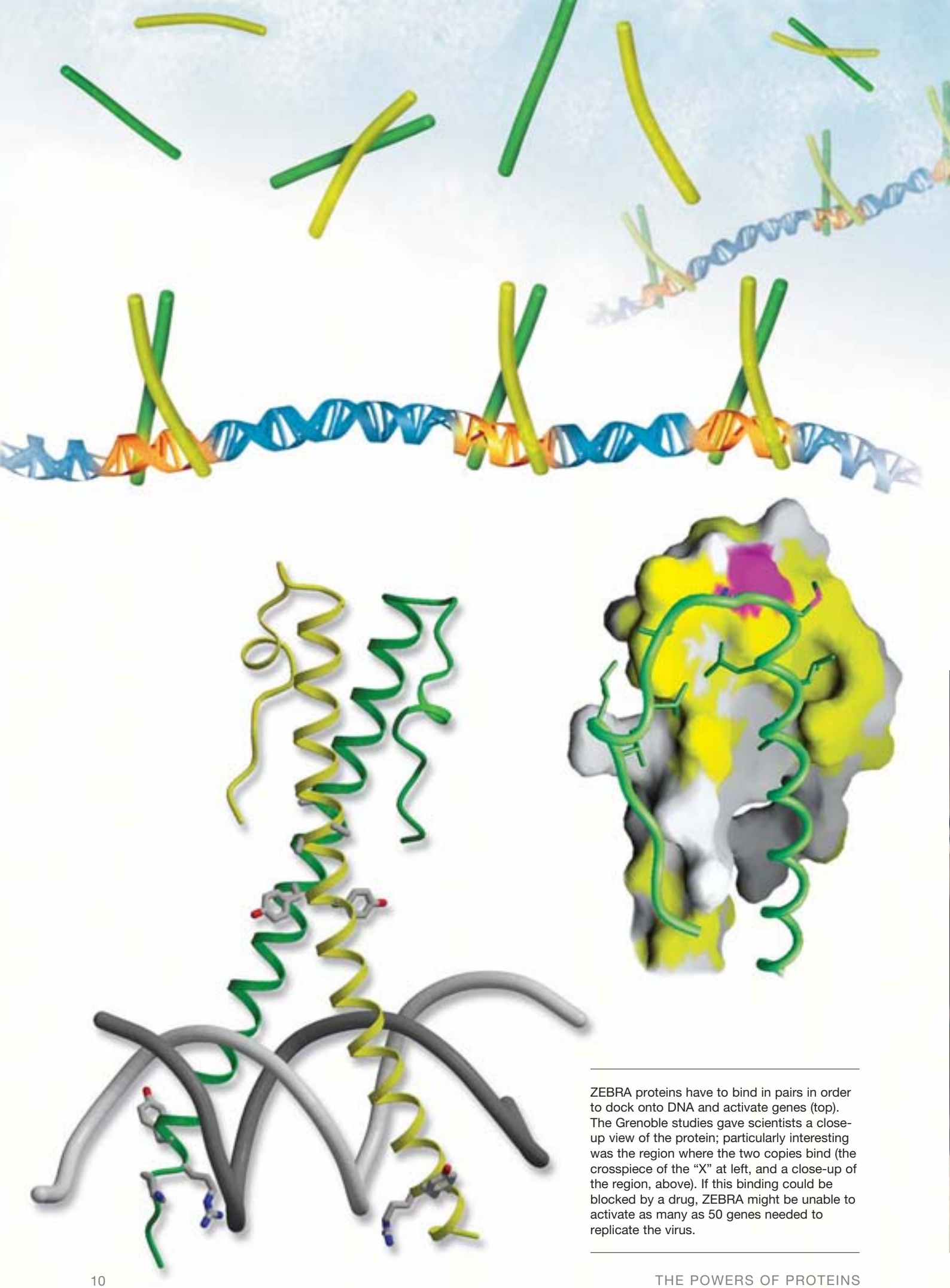
The best therapy currently available in the fight against EBV blocks about twenty of its genes. But those are molecules that are needed much later in the life cycle of the virus, the *lytic* phase, when its reproduction is already well underway. It would be much better to stop

"ZEBRA is unusual because it activates about 50 different EBV genes – over half the total number brought into the cell by the virus."

EBV earlier – and to affect more genes. One method might be to put a molecular roadblock between ZEBRA and DNA. You could, for example, try to design a drug that attached itself directly to the part of ZEBRA that binds to DNA, thereby disrupting ZEBRA's ability to turn on viral genes.

Christoph doesn't give that approach very good chances of success. ZEBRA's DNA-binding part is similar to that found in some of the cell's own transcription factors. That's interesting because it suggests that long ago, ZEBRA likely evolved out of one of these factors. But when it comes to designing a therapy, it causes a problem: any drug blocking the ability of ZEBRA to bind DNA would probably also block that of some similar cellular transcription factors, with potentially toxic effects on the host. "It's hard to see how you could interrupt only EBV genes by such a strategy without interfering with some of the vital functions of the cell," Christoph says.

The alternative would be to attack some other part of ZEBRA, and here a structural plan of the molecule might reveal weak points that could be useful in stopping it. Patrice managed to turn the protein into crystals. He then asked Carlo Petosa, a staff scientist in Christoph's group, to help him analyze these crystals using X-rays. Together, they succeeded in obtaining the first three-dimensional structure of ZEBRA on the ESRF beamlines. The picture revealed that two copies of the molecule bind to each other like a long pair of tongs, something you might use to lift a sausage off a hot grill. In this case what's being grasped is DNA.



ZEBRA proteins have to bind in pairs in order to dock onto DNA and activate genes (top). The Grenoble studies gave scientists a close-up view of the protein; particularly interesting was the region where the two copies bind (the crosspiece of the "X" at left, and a close-up of the region, above). If this binding could be blocked by a drug, ZEBRA might be unable to activate as many as 50 genes needed to replicate the virus.

"The part that directly grips genes resembles some of the cell's transcription factors," Carlo says. "That was no surprise to us. But when we looked at other parts of the structure, we found something very interesting. For tongs to work, the two arms have to be joined together by a small bolt, or a pin. Two copies of ZEBRA are needed to activate genes, and they're also joined together by a sort of pin."

He points out the area he's talking about. One ZEBRA molecule inserts a knob-like protrusion into a deep pocket in the other. "Such a pocket is the ideal type of structure to look for if you want an effective drug," Carlo says. "If you could seal it up, then the two copies of ZEBRA can't join, so they can't activate EBV genes."

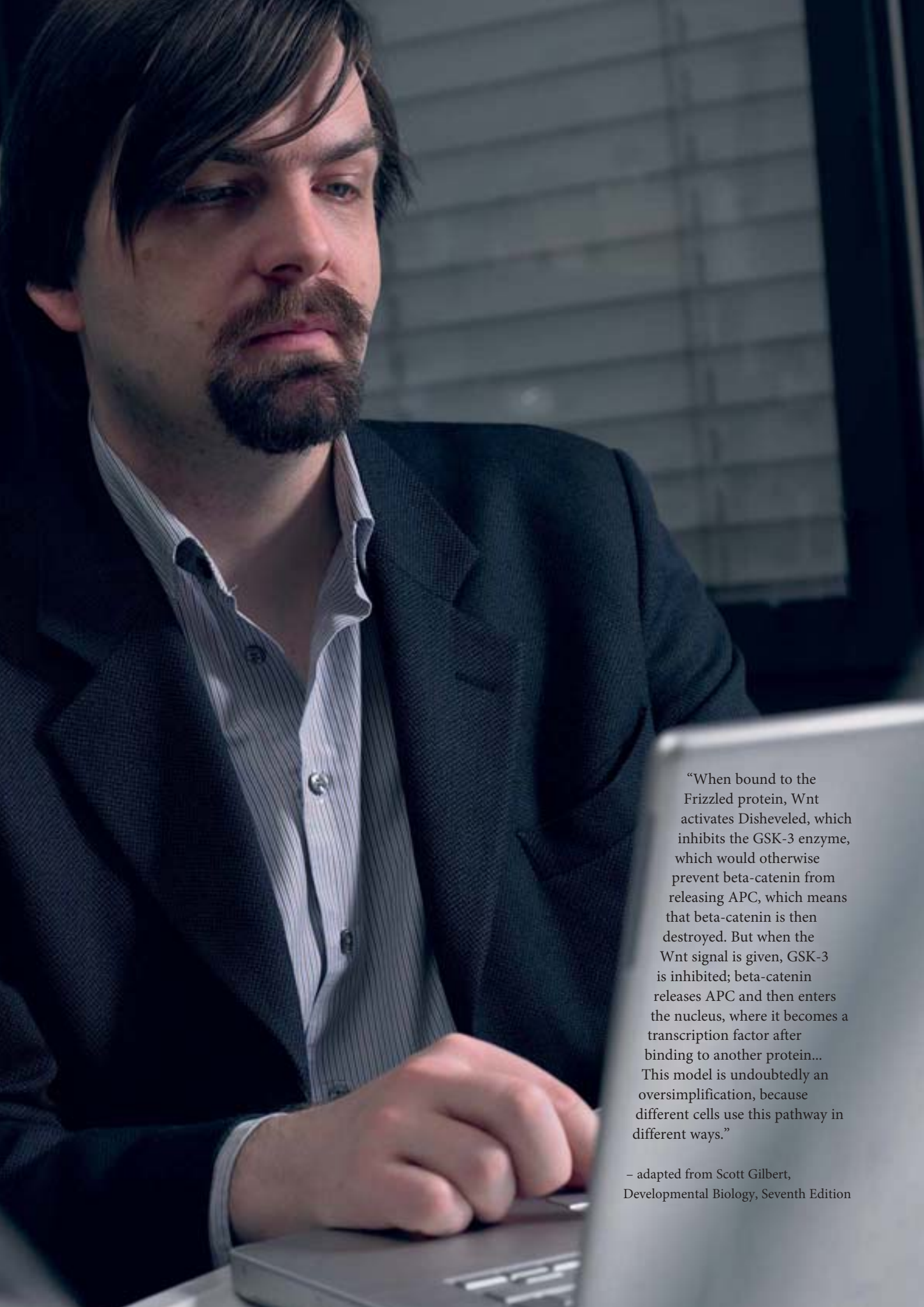
One ZEBRA molecule inserts a knob-like protrusion into a deep pocket of the other. Such a pocket is the ideal type of structure to look for if you want an effective drug.

The pocket is also unique, he says; scouring databases of known protein structures didn't turn up anything similar in human proteins. So you would be unlikely to do any damage to the host cell by blocking the pocket with a drug – but where to find one?

They decided to approach Joe Lewis, head of the Chemical Biology Core Facility at the EMBL Heidelberg, for help. "The Core Facility has an agreement with the pharmaceutical company Tripos, which provides a library of compounds that can be screened," Patrice says. "Something which already exists might fill the requirements, or it could be close enough to work if it is modified a bit, or the company might have to invent something new."

The partnership that has been established in Grenoble is well-equipped to stay involved. If a compound is discovered that stops ZEBRAs from forming pairs in the test tube, scientists will need a direct picture of how it works. The teams in Grenoble have the methods and equipment to investigate its structure, and someday in the future, when a therapy is ready to be tested, Patrice knows a clinic close at hand. ■

Mounting a crystal on the beamline.



"When bound to the Frizzled protein, Wnt activates Disheveled, which inhibits the GSK-3 enzyme, which would otherwise prevent beta-catenin from releasing APC, which means that beta-catenin is then destroyed. But when the Wnt signal is given, GSK-3 is inhibited; beta-catenin releases APC and then enters the nucleus, where it becomes a transcription factor after binding to another protein... This model is undoubtedly an oversimplification, because different cells use this pathway in different ways."

– adapted from Scott Gilbert, Developmental Biology, Seventh Edition

Labyrinths and identities

EVERY YEAR the village of Ladenburg, near Heidelberg, cuts a labyrinth into a vast field of corn; passing between stalks that stretch high above your head, you can wander for hours without coming to the centre, or finding the exit again. Still, these forking pathways are simple compared with some of the signaling that takes place in the cell. A signal activates molecules which may deactivate others, ultimately leading to an event which may activate hundreds of genes, and repress hundreds more. The whole system may be self-regulating – designed to shut itself down, or amplify itself. Pathways cross over each other, the way visitors in a labyrinth may converge after following different routes, and yet the cell manages to keep their destinations straight.

Underlying each branch in the pathway and each act of regulation is an ancient evolutionary choice, the way the streets of today's towns lay above older roads, built to accommodate ancient buildings and earlier forms of transportation. Signaling using the *Wnt* molecule probably dates back to the first animals. Like nearly every ancient genetic road, nature has put it to many uses in different species – even within a single species. The pathway plays a central role in building body plans and organs, and now Claus Nerlov has shown that it helps determine the identity of our blood cells.

Differentiated cells are the result of a series of decisions – genetic programs which are activated in a certain sequence – like choosing a specific route through a labyrinth. Very basic forms of stem cells in the bone marrow receive signals that prompt them to specialize into less generic types; after other signals and several more rounds of decision-making, they then become red blood cells, or B cells, or dozens of other types. Claus and his group in Monterotondo have been tracing the pathways that produce many of these kinds of cells.

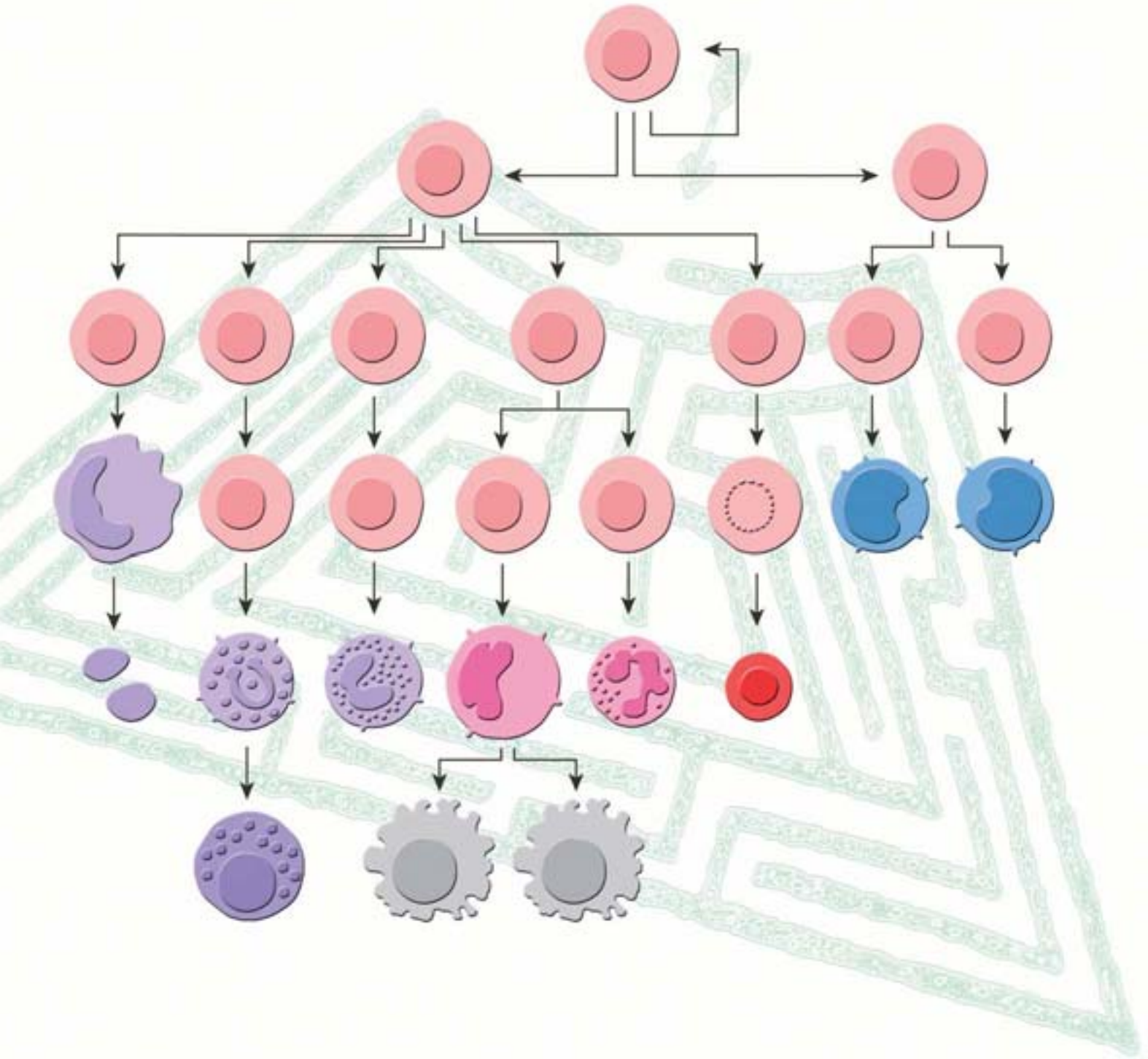
When signals fail, he says, cells may divide at the wrong time, or stem cells may not specialize properly. When cells lose their control over Wnt, the result may be cancers such as leukemia. There may be ways to repair these pathways, Claus says, but first we will have to understand them exhaustively.

* * *

Which signals guide the development of *hematopoetic stem cells* into blood, and along which pathways? Evidence from some studies suggested that the "classic" pathway, in which Wnt exerts control over beta-catenin, was involved (see the quote on the opposite page). Other research suggested that Wnt's effects on blood might be directed through other molecules, along a different route.

Peggy Kirstetter, a postdoc in Claus' group, decided to try to find out. The technique she used was to introduce into cells a version of beta-catenin that didn't break down – in essence, leaving the Wnt pathway switched on. She did this in a strain of mouse with a specially engineered version of the beta-catenin gene.

"The Wnt pathway is important in many different ways during development," Peggy says. "This means that we



Blood cells have a complex lineage, starting with hematopoetic stem cells and differentiating into many types. Intricate networks of signals and transcription factors guide cells through a labyrinth of specialization. Mutations or other problems can cause cells to take “wrong turns” – sometimes leading to cancer. If a signaling pathway called Wnt, operating through the beta-catenin protein, is too active, many types of blood cells fail to develop.

EMBL Monterotondo

can't use the most straightforward genetic techniques to study it, such as removing genes in the pathway entirely from an animal's genome. That would eliminate its function in all cells, stopping the embryo's development at such an early phase that we couldn't observe the protein's role in blood. So we used a strain of a mouse in which it can be activated *conditionally* – only in cells in certain tissues. In this case we introduced the altered form of beta-catenin only in the hematopoetic lineage."

Monterotondo groups have developed several strains of such animals, called *Cre* mice, permitting the control of genes in specific tissues at precise times. These strains have become important in the identification of connections between genes, animal development and diseases such as leukemia.

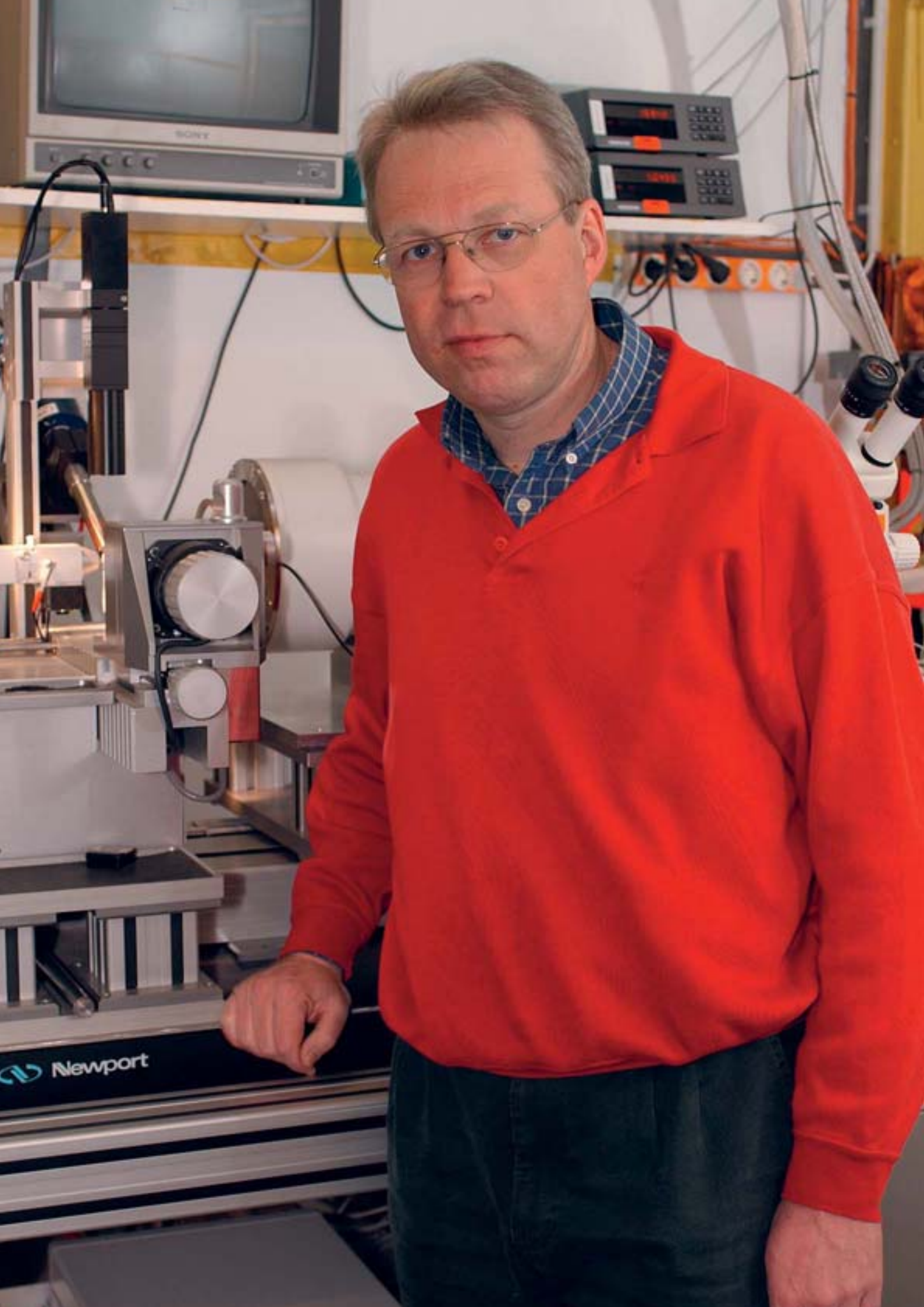
The modified beta-catenin protein had dramatic effects. Several types of blood cells vanished entirely; the same thing happened to more basic types higher up in the blood lineage hierarchy. "Blood cells typically have a relatively short lifespan and have to be replaced by stem cells that launch developmental programs," Peggy says. "This wasn't happening; our mice were unable to replenish the hematopoetic system, and there was an almost complete absence of some progenitor stem cells in the bone marrow."

Transplanting this marrow into other mice, which had defective immune systems and were unable to produce their own blood, led to the same effects. This confirms that the Wnt pathway, operating through beta-catenin, is able to block blood cell formation. On the other hand, it doesn't seem to damage cells that already exist. So it seems to be a decision-maker, a selector of paths, rather than somehow maintaining the vitality of existing cells.

Where does development break down? "In several places," Claus says, "for example, the progenitors of myeloid cells originate from bone marrow stem cells. The active form of beta-catenin causes that population to increase, but at the same time blocks their development into myeloid cell types. To figure out why we looked at other proteins that were becoming activated in LSK cells, and discovered some abnormalities. It looked like they took the next developmental step too early, like taking the wrong fork in a labyrinth, and coming to a dead end."

Other types of cells, B- and T-cells, were also blocked at early stages, but in a different way. This hints that they may be controlled at other points in the Wnt pathway. Claus and his group will now try to find the precise molecules involved.

"It's interesting to note that the failure to develop leads to unexpectedly high numbers of more basic cells – the stage where wrong decisions are made," Claus says. "You can think of it as a cluster of people standing at a crossroads in a labyrinth, hesitating before they go on. This is suggestive because we know that there are strong connections between cells' decisions to divide, develop, or die. If cells don't commit themselves to the right developmental path at the right time, they're very likely to die, or begin an inappropriate type of reproduction. Acute leukemias and other forms of cancer stem from defects such as this. To truly understand them, we're convinced we need to pinpoint the exact branches in the pathways where things go wrong." ■


SONY
Newport

Matthias Wilmanns,
Head of EMBL Hamburg

A seed of symmetry

IMAGINE USING ONE HAND to grasp two long snakes by their tails so that they won't wriggle away in opposite directions. Molecules carry out a similar act as muscles form in animal embryos. The snakes in this story are proteins, and Matthias Wilmanns and his group in Hamburg think that how their tails are caught by another protein plays a key role in the development of this complex tissue.

The long bands that make up muscle arise from an unusual fusion of several cells, and for many years Matthias and his lab have been trying to pin down the steps by which this happens. Their work has focused on a protein called *titin*, whose size and position suggest it could help structure the tissue.

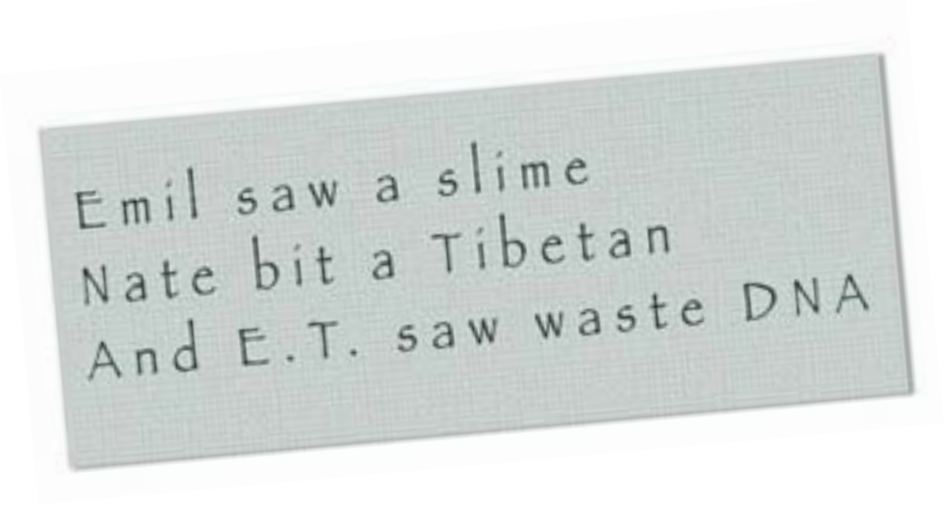
Under the microscope, fully-formed muscle contains millions of tiny compartments, stacked end-to-end in long strips. Each of the subunits in the strip, called a *sarcomere*, functions like a piston: an outer sheath of proteins slides in and out around a rod-like core. When they all act in concert, the result is large-scale contractions and relaxation that allow our bodies to move. Neighboring sarcomeres are connected by a thick band of proteins called a *Z-disc*, an anchor point for proteins on either side.

"The molecules in this region are so densely packed that it's been hard to understand their interactions," Matthias says. "Whatever their organization, symmetry is a crucial feature, because of the mirror-like structures on either side. The same types of molecules exit the Z-disc in opposite directions."

"Titin could have a scaffold-like function as the sarcomere is built: by linking to molecules along the long axis, it helps organize them."

One of these is titin, the largest protein produced by our cells. The molecule is so long that it extends from its anchoring point in the Z-disc fully half the length of the sarcomere. "That means it comes into contact with all the major components of the structure," Matthias says. "We've thought it could have a scaffold-like function when the sarcomere is being built: by linking to molecules along the long axis, it helps organize them."

Showing that this is the case has been a huge project – one that several EMBL groups and their collaborators have been working on for more than ten years. In the beginning, tackling the molecule was a crazy dream, Matthias says. Some scientists didn't even think titin really existed – how could the cell synthesize such a huge molecule? One of the first tasks was to sequence the gene, itself an enormous job back in the days before high-throughput DNA sequencing techniques. Titin has up to 38000 residues and consists of 300 modules, many of which are nearly identical, like beads threaded onto a string.



Palindromes, like the sentences to the left, also occur in the cell. Telethonin binds to two copies of the protein titin, essentially reading it like a palindrome.

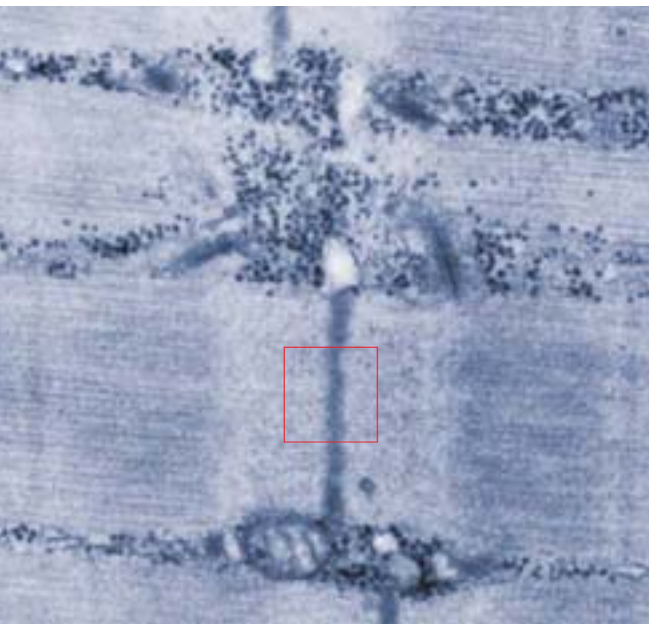
But step by step, the researchers have been exposing the details of both the structure and function of titin's modules. Understanding how these bind to other proteins can go far towards explaining the molecule's behavior and possibly its role in structuring the sarcomere. Recently the scientists achieved a breakthrough when they attained a close-up glimpse of titin's anchor point in the Z-disk. Matthias' group obtained crystals of the module bound to another protein called *telethonin*, which they analyzed on a synchrotron beamline at the Hamburg Outstation.

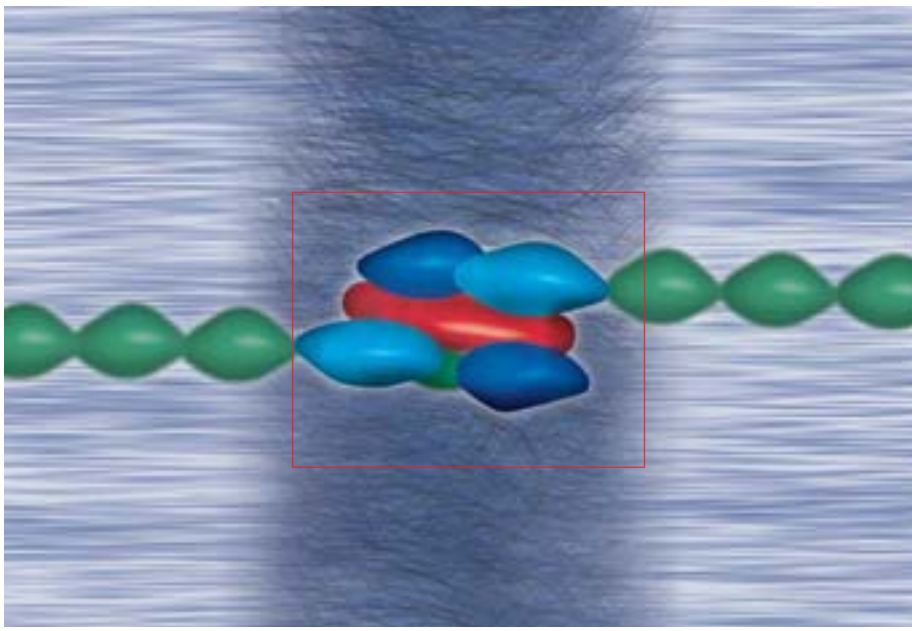
"It is interesting to discover a subunit that in itself is symmetrical. It's like finding a seed around which the whole structure might form."

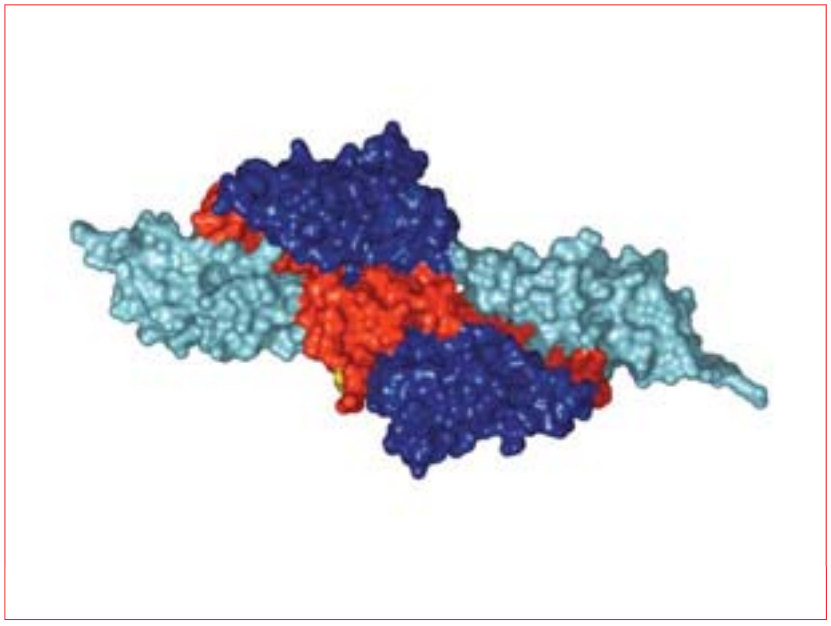
"This structure was particularly interesting because it gives us a direct look at how titin molecules from either side enter the Z-disc and are bound to each other," Matthias says. "Previously it wasn't clear that the ends were linked to each other at all. Now we see that telethonin grasps onto two copies of titin, and that this linker molecule has an interesting property that we haven't seen before in a protein. The titin modules run in opposite directions, and telethonin links to them like a palindrome – a phrase that has the same spelling whether you read it from the beginning to the end or from end to beginning."

Proteins are known to bind to DNA or RNA in this way, but such a palindromic link hasn't been observed before in protein-protein interactions, Matthias says. Yet it is exactly what you would hope to find in a structure like the Z-disc.

"Mirror image sarcomeres on either side mean that overall, the Z-disc has to be symmetrical," Matthias says. "But it's so complex that we haven't been able to unravel how that symmetry might be organized. So it is interesting to discover a subunit that in itself is symmetrical. It's like finding a seed around which the whole structure might form. There may be other such structures, linking other proteins from the two sides, and we'll continue to look for them. But this confirms our image of titin as a scaffold around which the sarcomeres organize themselves." ■

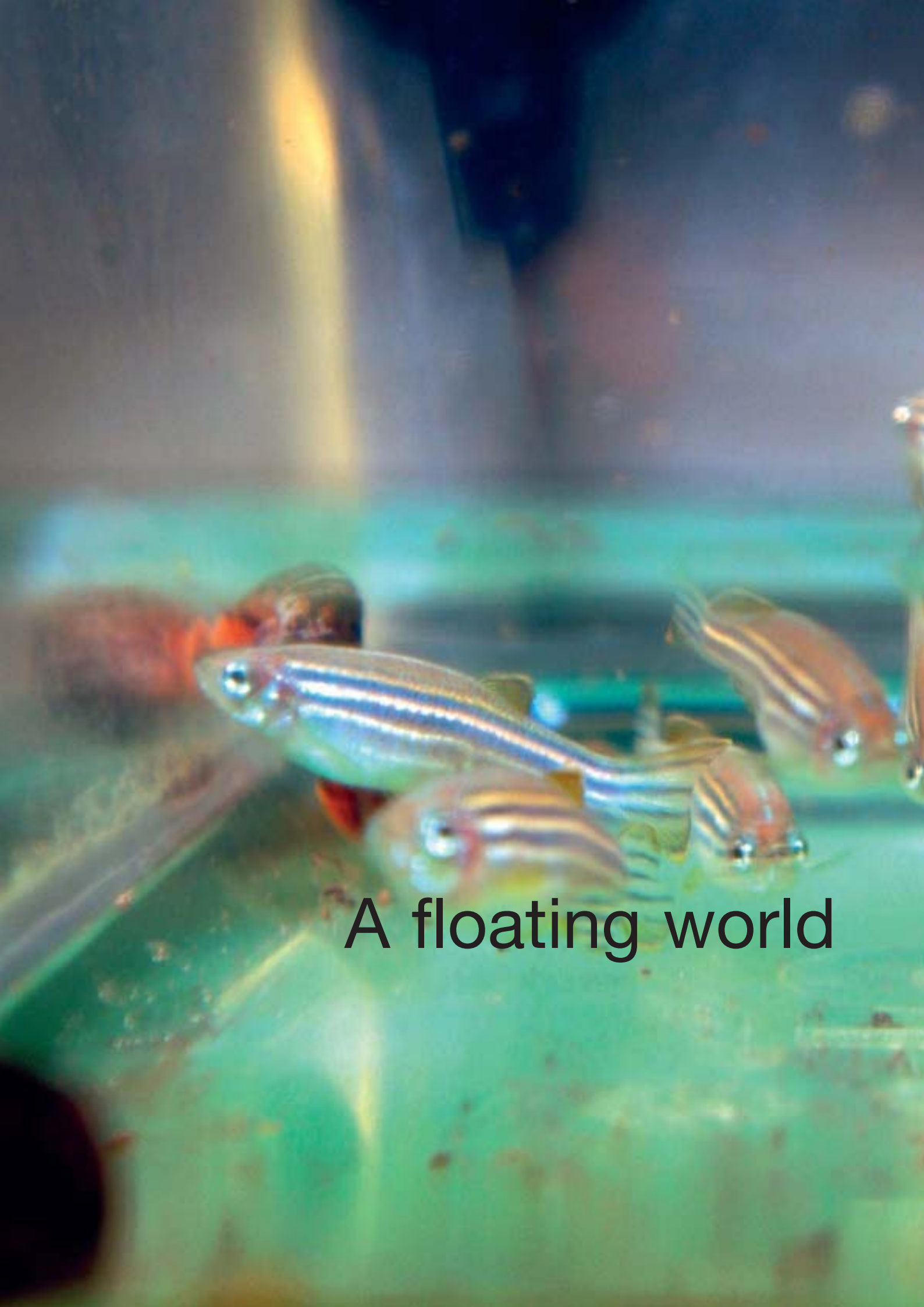




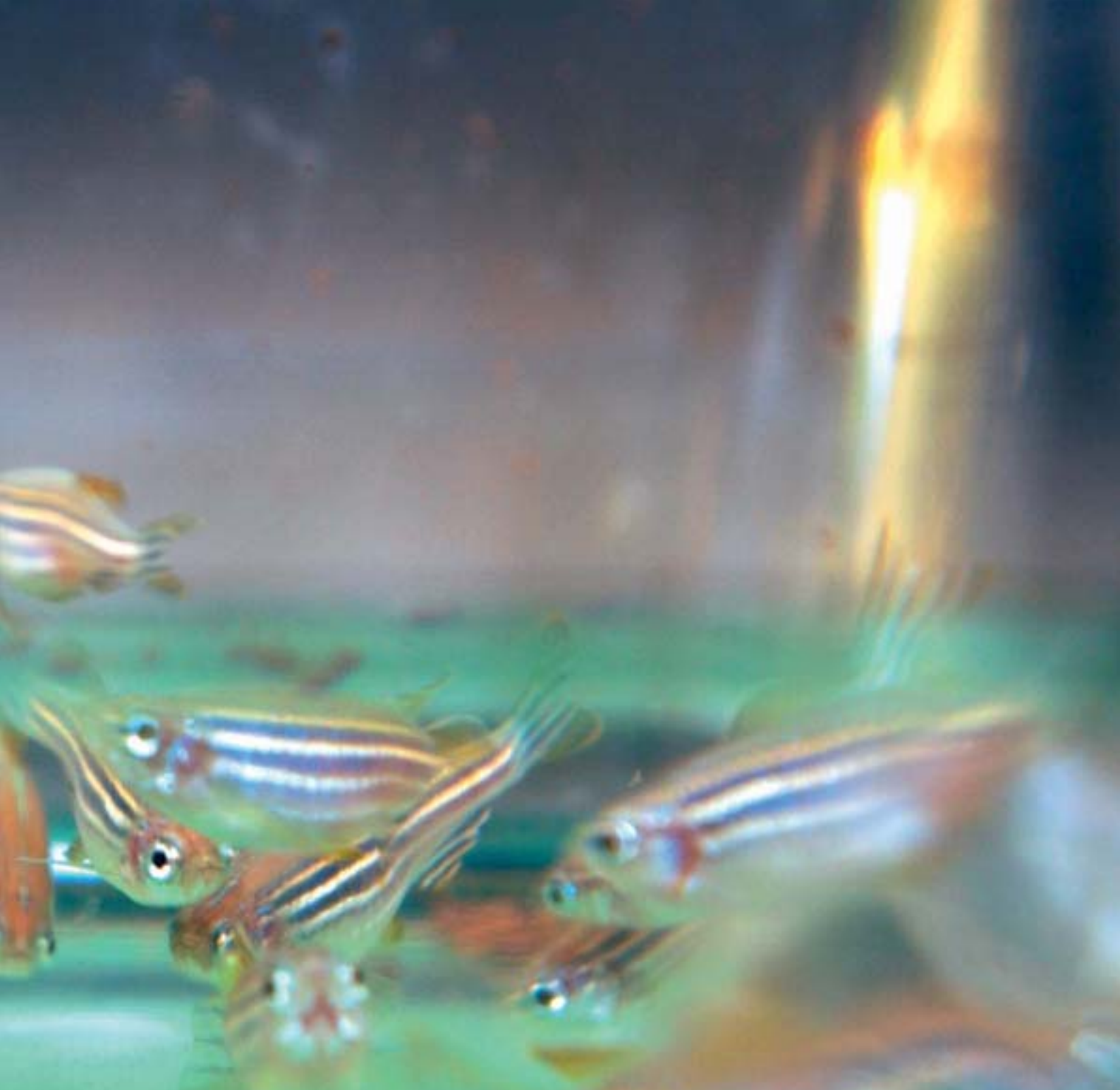


Above: EMBL's scientists build and operate beamlines and provide crucial services for users at the Hamburg Outstation. Petr Konarev and Manfred Roessle

Sarcomeres (far left) meet in densely-packed regions called Z-discs (red square). A close-up view of telethonin (red, middle picture) bound to two copies of the long, snake-like molecule titin (blue and green) suggests how symmetrical structures in the Z-disc might help organize sarcomeres. Near left: the strucure obtained by Matthias and his colleagues.



A floating world



IN 1895, working in his lab at the University of Würzburg, Wilhelm Röntgen discovered a new form of radiation that could pass through everything but metals. It also passed through the body unobstructed, unless deflected by mineralized tissues such as bone. Scientists began applying these X-rays to all sorts of materials, including crystals. Chemists weren't sure what crystals were, but they thought they might be three-dimensional lattices of molecules, stacked in symmetrical patterns; if so, they might deflect X-rays in ways that would reveal the arrangements of their atoms. Paul Knipping and Walther Friedrich soon proved that this was the case by putting a photographic plate behind a crystal and bombarding it with X-rays. The plate captured symmetrical patterns of spots.

Such *diffraction patterns* provide a deep look into the structure of a crystal. If it is well-ordered, a few copies of the molecule will be arranged into what's called a unit cell, and billions of copies of these will be stacked in arrays. This produces the regularity of the spots captured by photographic plates or modern detectors.

But the use of X-rays in biology isn't limited to crystals; they can also be used to study molecules in liquid environments more similar to that of the cell, says Dmitri Svergun. Along with Michel Koch and colleagues at the Hamburg Outstation, he has been working to perfect a technique called *small-angle X-ray scattering* (SAXS). This can yield information about molecular structures that is hard or impossible to obtain using crystals.

Dmitri explains the difference between what can be learned using the two methods. He compares the situation to shining a light through an aquarium full of fish. (A crystal would be more like rows of herring, packed in a tin.) "Even if you suppose that all the fish were the same size and shape, in water they would be turned in all sorts of directions and swimming, which would give their bodies different orientations," he says. "A diffraction pattern is an average of all of those situations. With a crystal you're averaging things that have just a few positions in the unit cell, and that cell is repeated billions of times. This creates a high-resolution pattern that often permits us to detect the precise position of every atom."

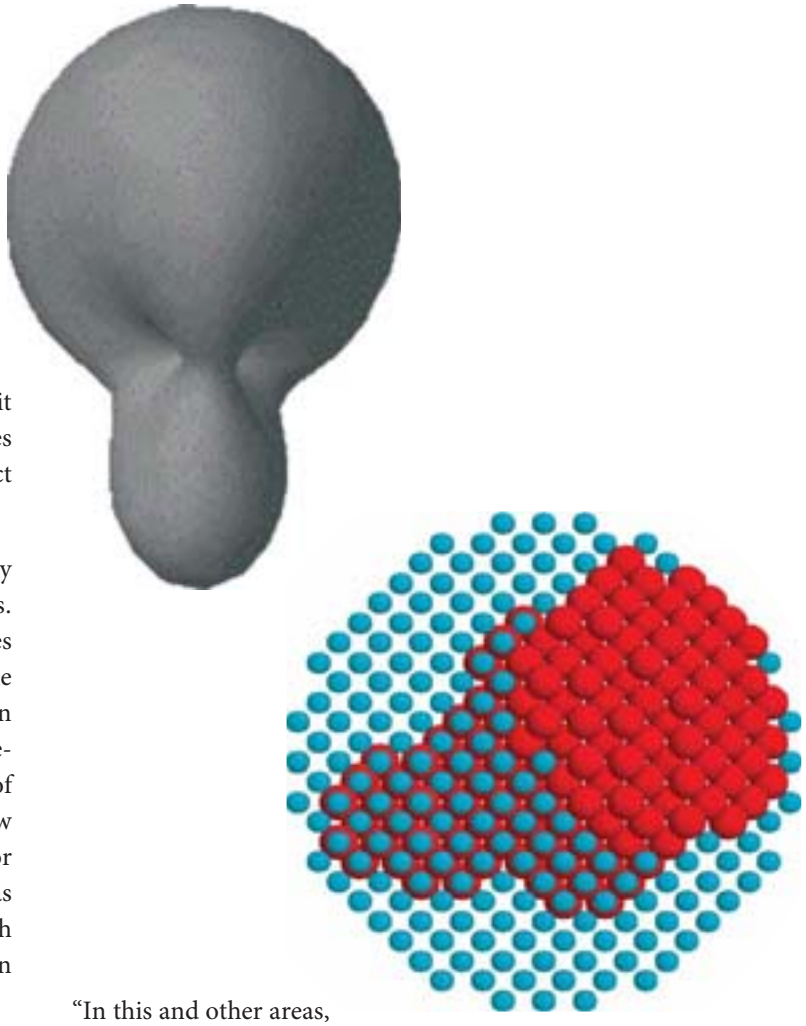
But proteins don't normally live in a lattice, he says; they move, bind to partners, work in huge complexes. Capturing all these different states would require scores of different crystals, many of which would be impossible to obtain – multi-component machines usually bind in disorderly clumps rather than neat arrays required to create crystals in the first place. SAXS can give a picture of larger objects and their dynamic nature, for example, how proteins change when they lock onto other molecules, or their reactions when treated with a drug. His team has made great strides in analyzing the data produced by such experiments; he thinks that even more information can be teased out.

Increasingly, groups throughout Europe are coming to Dmitri for help using the technique with their projects. Alongside EMBL labs, such as that of Winfried Weissenhorn in Grenoble (see story on page 130), a network has formed of scientists who hope to use and improve the method at laboratories in France, the UK, Germany and Italy. The collaborations form the basis of an EU project called SAXIER; the European Commission is providing half of a 7.2-million-Euro budget. Dmitri is coordinating the project.

"SAXS performs better the 'brighter' the beamline," he says. "We're ideally situated because of future plans here on the Outstation campus: our host, the German Electron Synchrotron Radiation Facility (DESY) plans to dedicate a large storage ring in Hamburg, PETRA-III, to produce synchrotron radiation. This will put us at the brightest synchrotron source in Europe."

* * *

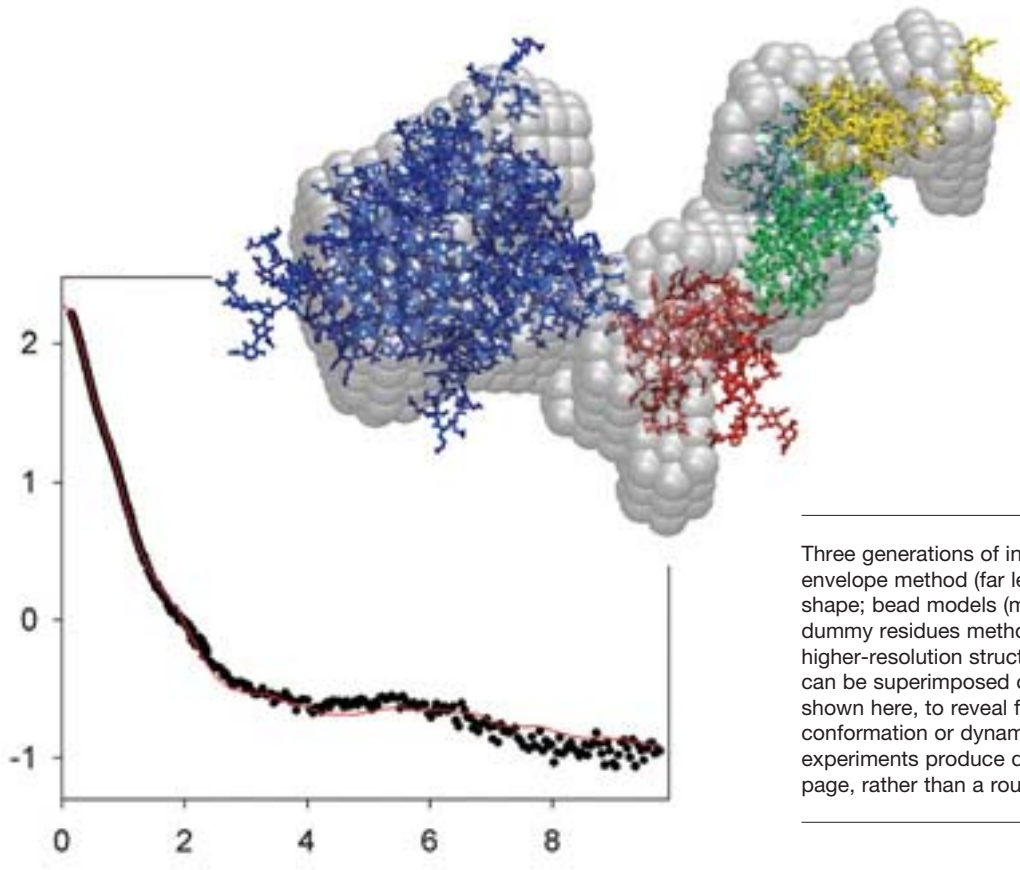
Diffractions of crystals typically contain thousands of reflections, arranged in a circular pattern; a SAXS experiment produces a one-dimensional stream of measurements. Although this makes the resolution from SAXS much lower, it provides information about forms and processes.



"In this and other areas, it often succeeds where other methods fail," Dmitri says. "We can watch changes in a molecule as it switches between inactive and active forms." With Paul Tucker, they investigated a gene-activating protein called PrrA from the tuberculosis bacteria. In crystals, PrrA has a compact form; with SAXS, they watched it become more extended when it was activated by another molecule.

Another use of the method is to study the structure of complexes. "Crystallography has often given us pieces of data that have to be fitted together – for example, there may be a high-resolution structure of one domain of a protein, or one member of a complex," Dmitri says. "We want to observe these pieces in their context, and how they behave over time."

The groups of Christoph Müller and Iain Mattaj had used crystals to solve the structure of *importin-beta,* a molecule that transports other proteins into the nucleus. Crystal studies had revealed the conformation of the molecule connected to a cargo, and in its unbound state. But importin can bind to many types of proteins, and it may do so in different ways. Using SAXS, Dmitri worked with Elena Conti's group to reveal both the general principles and specific elements of different import and export complexes.



Three generations of interpreting SAXS data: the envelope method (far left), giving a basic outer shape; bead models (middle), and the new dummy residues method (grey balls, above). If higher-resolution structures are available, they can be superimposed on a SAXS image, as shown here, to reveal features of a protein's conformation or dynamics in solutions. SAXS experiments produce data like the plot on this page, rather than a round diffraction pattern.

The same method can be used to get a time-lapse view of how complexes are assembled. Another use is to fill in missing parts of structures. It is often necessary to remove parts of molecules – such as flexible regions or loops – to get them to form into the symmetrical arrays of crystals. SAXS can show where some of these pieces fit into the structure.

One of the main challenges in developing the method has been to try to turn low-resolution information about a molecule in many different orientations and conformations into the clearest possible pictures. "Fifteen years ago we began with an 'envelope' method that could trace the outer shape of a protein, but gave few other details," Dmitri says. "The next step was what we called a 'bead'-model, where we portray a protein as a set of round balls, which have the size of our maximum 'resolution'. To obtain clear structures, we used a type of reverse prediction. If we knew in advance what a shape looked like, we could predict what sort of SAXS diffraction pattern it would make. This is backwards from the real situation in which we need to go from the pattern to the structure. Since we can't do that, we program the computer to assemble balls into a shape, then test the pattern that this shape would lead to in a real diffraction measurement. The computer does this again and again, gluing balls together in different shapes, each time changing things so that it gets closer to the real scattering pattern."

That method was better, he says, but still far from the goal. It held only shape information. "It's like seeing the rough silhouette of a fish, but not being able to say where the eyes are, or the mouth, or any other specific parts. We've now moved to a procedure called the dummy residues model, where we add what we know about the chemistry of a protein. We know that it's made up of a linear sequence of amino acids, and this adds constraints: certain things have to be a certain distance from each other. This means we can try to assign real parts of the protein to specific regions."

The dummy residue method uses balls, like the bead model, but now they represent real amino acids. The computer tries combination after combination until it has created a hypothetical protein that matches the diffraction pattern as closely as possible. It doesn't have to try all combinations – only those that are conceivable within a real protein.

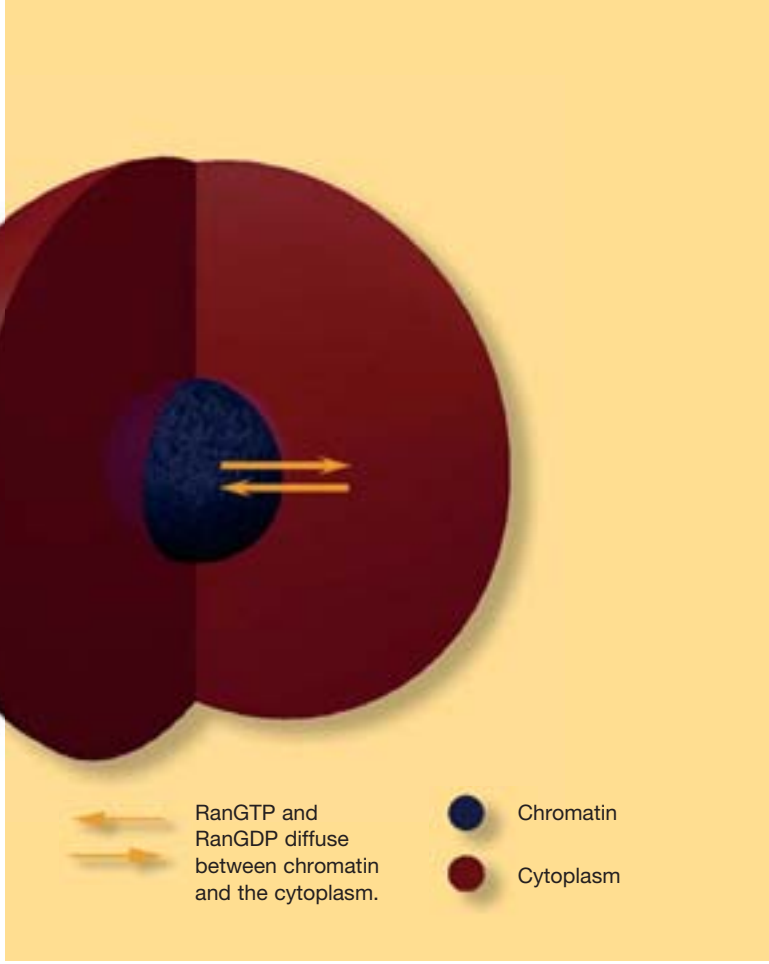
In isolation, SAXS doesn't provide much of the information that structural biologists would like to know about a protein. Its real power comes not only when combined with high-resolution methods like crystallography or nuclear magnetic resonance, but also with electron microscopy and biochemical information. SAXS is applicable to an extremely broad range of protein sizes, from individual molecules to large macromolecular machines, and will show scientists the shapes to look for in microscope images. ■

The architecture of space

Maïwen Caudron,
Eric Karsenti and
Philippe Bastiaens

FROM THE PATTERNS FORMED BY FLOCKS of birds and schools of fish to the food-finding behavior of an ant colony, the living world is full of complex dynamic patterns Until the nineteenth century, the only explanation for such phenomena was the idea that nature was directed by an intelligent creator. But as chemists, physicists and naturalists progressively unraveled the organization of matter and the living world, they began to understand that incredibly complex forms could arise spontaneously from dynamic interactions between objects. Changes in the elements or their properties can yield a wealth of patterns. Thus the V-shape of a flock of birds probably arises from a simple set of behaviors whereby birds try to fly toward the centre of the group, while adjusting to the speed of their neighbors and keeping a safe distance away. Changing any of these parameters – for example, if birds fly at different speeds, or if they don't aim for the centre – would alter the pattern.

For years, Eric Karsenti's group has been looking at the architecture of *microtubules* in this way. These structures help give cells their shape, hold organelles in their places and serve as transport routes through the cell. And microtubules perform another essential function: as the cell divides, they tow copies of chromosomes in opposing directions, to regions where two daughter cells will form.



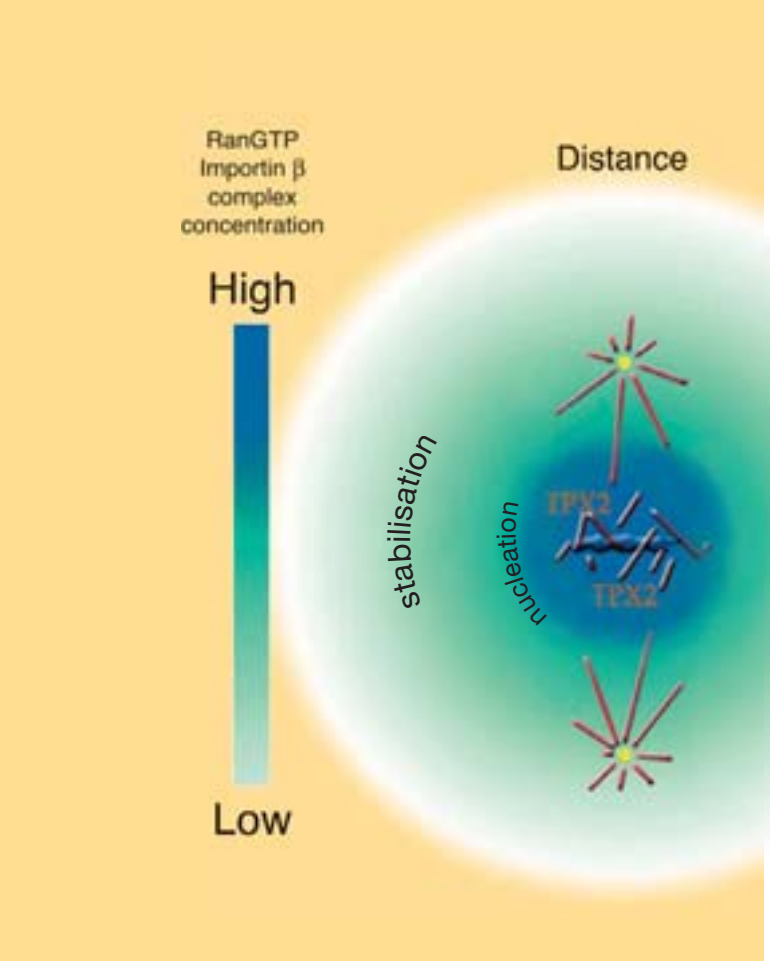
Doing so requires a major restructuring of microtubules. Each of the tube-like fibers is composed of protein building-blocks called *tubulin*, stacked in ringed rows the way bricks might be used to construct a round tower. Between cycles of cell reproduction, microtubule-building begins near the cell nucleus, at structures called *centrosomes*, and proceeds outwards towards the periphery of the cell. When cell division begins, the entire network of fibers is broken down, and then reshaped into a spindle-like form in the central region of the cell. Chromosomes are positioned at the equator of the spindle , linked by microtubules to poles on opposite sides.

Incredibly complex forms can arise spontaneously from dynamic interactions between objects. Changes in the elements or their properties can yield a wealth of patterns.

What prompts the cell to break down the network and build this form? What is different about the cell as it divides, and what coordinates the behavior of the spindle? Eric and his colleagues have been trying to answer these questions with a combination of physical studies of molecules, experiments in cell extracts, and simulations that assign properties to virtual proteins and generate rules by which they should act. The result is an eerie imitation in the test tube and on the computer screen of what happens in living cells.

Until a few years ago, most attempts to understand the spindle started with the poles. Usually these two outer positions are occupied by a centrosome, the source of microtubules under normal circumstances, thus it was logical to assume that filaments were being built outwards from there. But some types of cells build a spindle without centrosomes, so something else had to be managing the construction of the structure. Eric and his colleagues shifted their attention to the central region of the spindle, and the chromosomes. They coated microscopic beads with chromatin – the mixture of DNA and other molecules found in the nucleus – and dropped them into extracts derived from dividing cells. Suddenly microtubules formed around the beads, and they even adopted the shape of a spindle.

Using computer simulations, Eric and François Nédélec showed that poles formed through simple rules, based on the activity of *motor proteins*. There are different types of motors, each of which travels down microtubules in one



Concentrations of RanGTP bound to the protein Importin-beta form a gradient in cells which influences the behavior and stability of microtubules. Under high concentrations, subunits are added onto the microtubules; lower concentrations cause stabilization.

direction – either towards the growing tip, or upwards towards the source. As the motors move, they sometimes link to more than one microtubule, which pulls the strands together. This sorts filaments into bundles aimed in the same direction.

Something in chromatin was getting the process started. The labs of Eric and Iain Mattaj pinned the effects down to a molecule called *RCC1*, whose activity caused tubulin to *nucleate* (link into chains) and form microtubules.

"We were left with two competing theories about the construction of the spindle," Eric says. "One view proposed that things started from the poles and moved inwards; our work suggested that things might be working the other way around. Maybe these two views could be brought together. So that is what we've been working on."

* * *

Spindle-building requires that microtubules and their components behave differently in different places. "Explaining their behavior requires understanding what induces tubulin to start to form microtubules, what keeps them from breaking down once they have reached a certain length, and how they develop into the form of the spindle," Eric says. "For example, growing microtubules are very dynamic – they are continually added onto and broken down at the tips – which explains why they are usually fairly short. The fact that they are longer in the spindle, and are stabilized near the chromosomes, has to be accounted for. If they behaved normally, they might never stay together long enough to attach."

Several years ago, Eric had the idea that some of this behavior might be due to the fact that the chemistry of the cell varies from place to place. For example, certain molecules are produced at specific places and diffuse outwards, dropping in concentration farther from the source. This forms gradients that likely affect a molecule's behavior. Just as moving to a higher elevation would change the behavior of an animal – because it would have to live off different plants and animals – filaments might be responding to the presence of new molecules and their environments. The question was how to observe this. Eric met with Philippe Bastiaens, who was bringing new microscope techniques into the lab, and they started experiments with a protein called *stathmin*. This work allowed them to capture the first images of a gradient in the cell.

Just as moving to a higher elevation would change the behavior of an animal, filaments might be responding to the presence of new molecules and their environments.

Could the same approach be applied to microtubules? Iain's group had long been working on the differences between the chemistry of the nucleus and its surrounding compartment, the cytoplasm. A major distinction is the behavior of a small molecule called Ran, which exists in two forms: a high-energy state called *RanGTP* and a low-energy state called *RanGDP*. RCC1 loads RanGDP with energy and transforms it into RanGTP. Since RCC1 is cemented to the chromatin, this creates a zone of RanGTP around chromosomes and in the nucleus. In the

cytoplasm, a molecule called RanGAP unloads the energy.

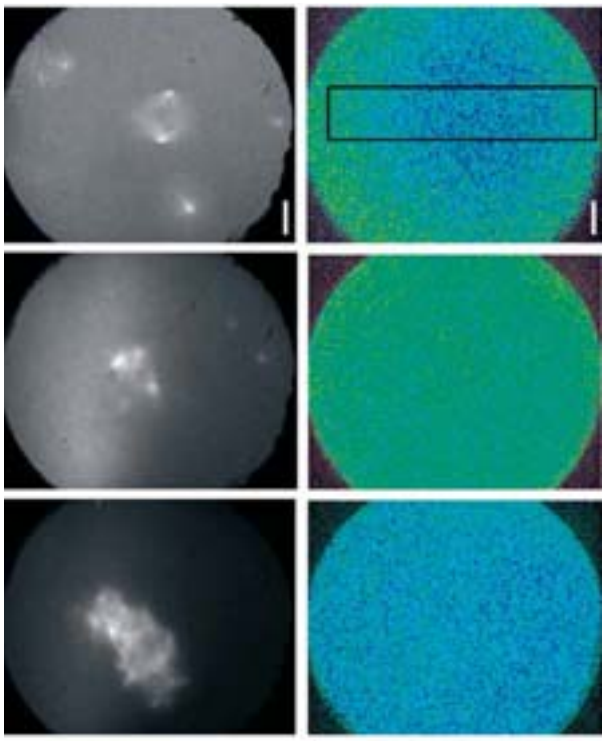
"This made us think that the different behavior of microtubules in different parts of the spindle might be governed by Ran," says Maïwen Caudron, a postdoc in Eric's lab. "But under the microscope we could see this wasn't a sufficient explanation by itself. We saw that even at a short distance from the chromosomes, there was almost no free RanGTP, so it couldn't account for a long-range structure like the spindle alone. It also wouldn't explain why the spindle becomes so asymmetrical."

Add to the puzzle the fact that RanGTP doesn't directly influence microtubules. That's done by other proteins, which affect how quickly tubulin forms filaments and how far they grow before they start breaking down again – the filaments are very dynamic and instable without factors to stabilize them. Iain's group has helped explain Ran's function by showing that cargoes of proteins that are brought into the nucleus to do jobs release factors that nucelate and stabilize microtubules. Once RCC1 loads Ran with energy around chromosomes, it can bind to a cargo and release from it the active factors that help build microtubules.

"Very little free RanGTP drifts away from the chromosomes," Maïwen says. "It quickly gets converted to the low-energy form. But it's often bound to other molecules, like importin-beta (cargo), and in that form it is found much farther away."

What the scientists needed was a chart of the geography of RanGTP linked to different complexes around chromosomes. Maïwen, Philippe and Eric then set up experiments using cell extracts and they began trying to observe gradients of such complexes. Philippe uses a laser microscope to study the physical properties of molecules. When the instrument illuminates fluorescent molecules, they give off energy, which makes them visible. This has another effect, because the amount of energy that a protein releases can be measured very precisely and the light itself has a given signature. When fluorescent RanGTP interacts with a second fluorescent molecule, like a fluorescent form of importin-beta, there is an exchange of energy between the two fluorescent molecules that results in another signature.

Using this method, called *FRET* (*fluorescence resonance energy transfer*), Philippe could chart both the location and the activity of RanGTP. But it was extremely difficult to observe gradients. "We ran into all sorts of technical problems," Philippe says. "The extracts were thick, and if there was a signal it was getting lost. And the laser had to be tuned very precisely because we discovered that light itself was having an effect on the molecule's activity. We were running into one of those 'Heisenberg' situations where the act of observation was changing the behavior of what we wanted to look at."

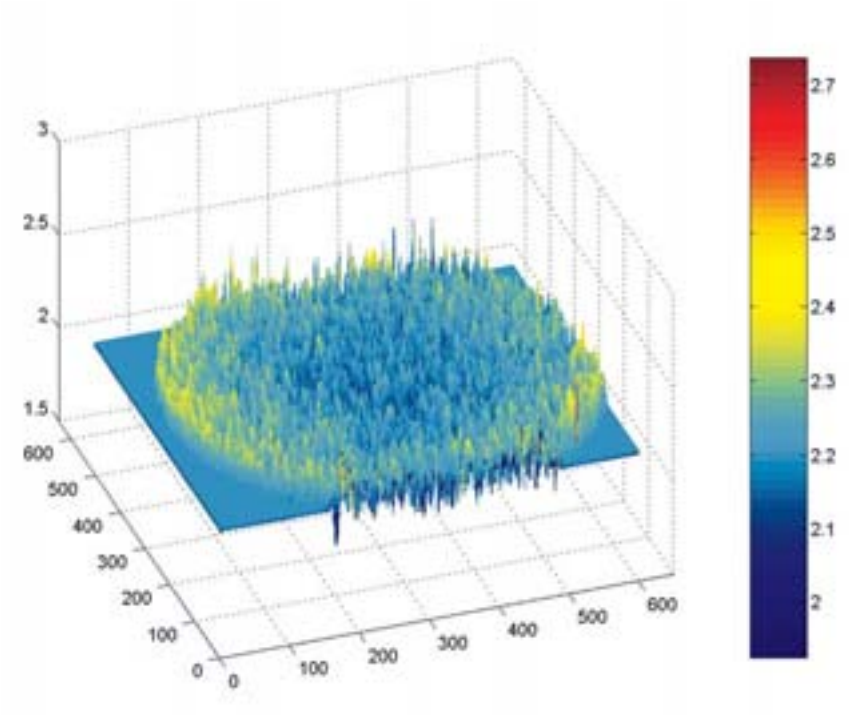


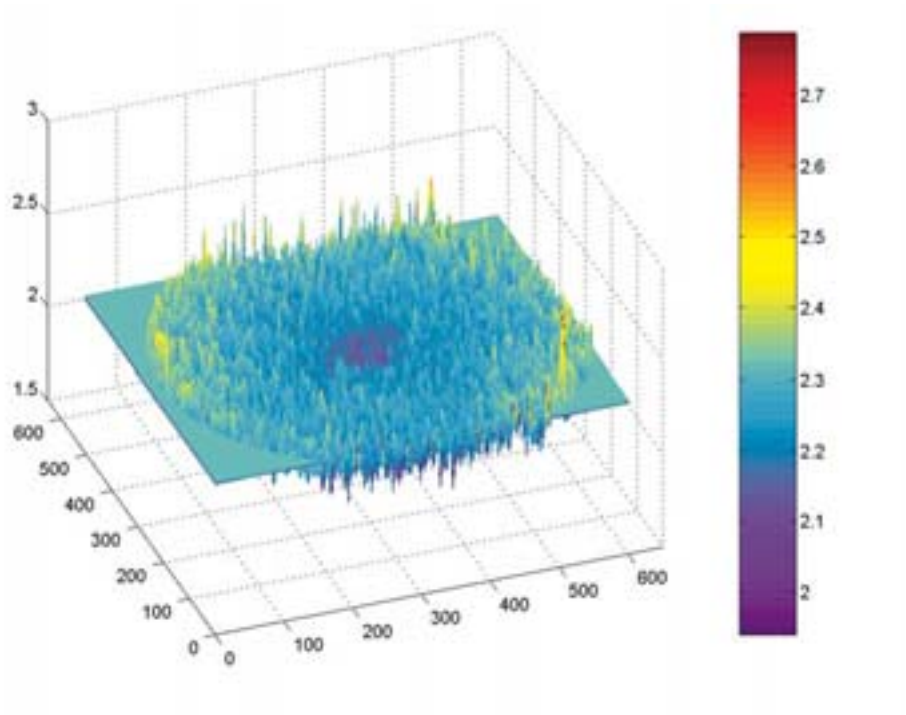
Frustrated, Eric and Maïwen and Philippe got together and decided to lock themselves in the lab until they could solve all the technical problems. "Then things went very quickly," Philippe says. "Eric was trying out various manipulations of the extracts, and we changed the wavelength and intensity of the light... A really important step was to make very thin samples of the extracts by pressing them between two layers of glass."

The result is a concentration of molecules which dissipates outwards in a symmetrical sphere. That symmetry gets broken up by the self-organizing activity of microtubules and motors.

Suddenly everything worked. The scientists obtained a map of RanGTP's interactions with importin-beta. There is a very high concentration of the molecules close to the chromosomes, which tapers off with distance, creating zones with different chemistries. Could this explain changes in microtubule behavior?

Left page: The left column of images shows the behavior of microtubules; the right column reveals the concentrations of RanGTP-Importin beta complexes under which the microtubules formed. Under normal conditions (top row) a gradient that forms around chromatin promotes the formation of proper spindle-shaped microtubules. If Ran is inactive everywhere (middle row) or active everywhere (bottom), microtubules don't form the proper shape.





This page: another representation of the gradient. Other molecules that affect microtubules can be used to change the normal gradient (above) to other shapes such as the one at left; this will also change the shapes of microtubules.

Maïwen returned to the cell extracts and began manipulating concentrations of molecules. She could change the gradient with RCC1, which generates RanGTP, and RanGAP, which transforms it to the RanGDP form. This had immediate effects on the shapes that microtubules formed. Changing the zones disrupted the ability of microtubules to grow and find chromosomes and destroyed the elegant symmetry of the spindle.

"In dividing cells, the nucleus is gone," Eric says. "The contents of the cell become mixed up, and this might turn into a homogeneous environment if it weren't for the chromosomes. They perturb things, along with molecules like RCC1 that are attached to them. The result is a concentration of molecules which dissipates outwards in a symmetrical sphere. That symmetry gets broken up by the self-organizing activity of microtubules and motors."

This picture, he says, moves us a lot closer to a picture of how the chemistry of the gradient determines these types of behaviors. "It also places the burden of maintaining the organization of the spindle on controlling the shape of the gradient during cell division. Now we need to take the next step and understand the factors that shape it." ■


BALLENA

Michael Knop (back right), Aleksander Benjak (red), Celine Maeder (green, in front), Nicole Rathfelder (white), Christof Taxis (middle), Reinhard Mayr (green, back), Massimiliano Mazza (back, middle), Deepankar Pal (left), and Peter Maier (front left)

To spore or not to spore

EAST OF HEIDELBERG, the Neckar River cuts a crooked valley through the lush forests of the Odenwald, one of the most scenic waterways in Germany. Hills overlooking the river are dominated by medieval fortresses with thick walls and high towers, often within hailing distance of one another. In the Middle Ages, the survival of a village was usually coupled to its capacity for self-defense: a position on high ground, and walls and gates to protect its citizens during a long siege.

Yeast cells don't have drawbridges to raise, but when the environment poses a challenge, they hide behind walls in the form of spores. This survival strategy is common to many simple organisms and depends on environmental factors.

"In yeast this process starts when cells are deprived of nitrogen and fermentable carbon," says Michael Knop, group leader in Heidelberg. "The organism does it in a special way because a single cell can generate between one and four spores, by creating compartments inside. Each spore holds a *haploid* genome – one copy of each chromosome rather than two, the way an animal egg or sperm cell has half of what is needed to build a whole functioning cell or organism. When the emergency has passed, that information is mixed together again – the yeast version of sex – to create cells with double-pairs of chromosomes."

A castle has watchmen to sound alarms, and generals to decide when the time has come to fill the moat and boil the oil. Michael and his colleagues have been looking for the decision-makers that guide the formation of spores in the cell. The fact that yeast decides not only whether to make spores, but also how many to make, means that there is a strong connection between the environment and the machinery that duplicates, splits up and relocates chromosomes.

The previous story shows how proteins and microtubules create a self-organizing system that helps the cell divide. Michael's lab has discovered a similar phenomenon of self-organization in the way yeast form spores, which has an impact on the long-term survival of an entire population of cells.

* * *

Spores are created within a "mother" cell as separate compartments with their own genes, wrapped in membranes. This happens in several stages. First, the mother cell duplicates its entire genome. Baker's yeast has 16 chromosomes which come in pairs, like those of animal cells; after this round of duplication, it has two *diploid* genomes (two sets, each containing pairs of each chromosome). Because of the way division happens, there are often slight differences between the chromosomes of two sets, but not within a pair. These differences have an impact on genetic diversity – they eventually lead to the development of unique, individual cells. In a second step, the pairs are split; now the cell has four haploid sets which contain all 16 chromosomes, but only one copy of each.

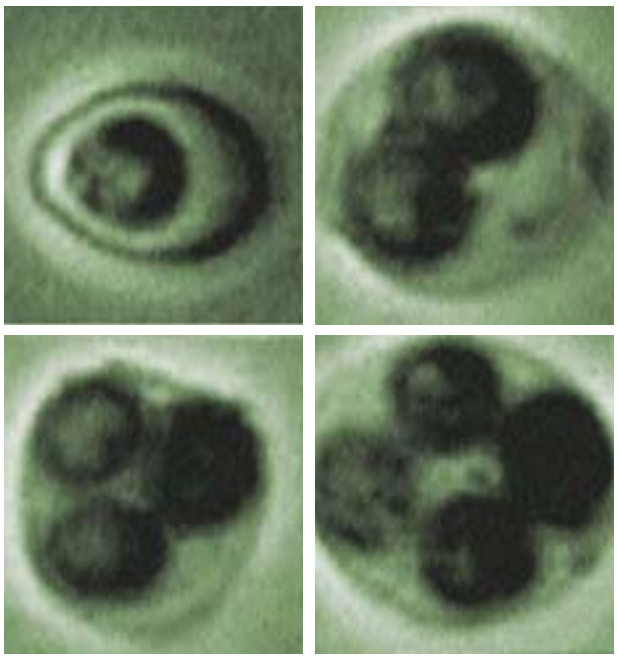
Moving all of this DNA to the right places in the cell depends on a spindle-shaped structure made of microtubules (see previous story). This can happen because as chromosomes are copied, the cell also copies a structure called a *spindle pole body* (SPB); microtubules grow from the four SPBs, which act as anchor points towards which the chromosomes are towed.

When making spores, an extra structure called a *meiotic plaque* (MP) is added to the spindle pole bodies. Several years ago Michael helped analyze the components of the plaque. The study showed that three proteins were critical to its main function: acquiring membranes and helping to shape them into compartments for spores. Once the MPs are in place, membranes form pouches that will surround and enclose each spore.

"What is curious about this whole process is that sometimes the cell makes four spores, and sometimes three, two or even only one," Michael says. "It takes more energy and nutrients to make four spores, and that depends on what's going on in the environment. So somehow there is a decision-making process that links signals to the environment to the process of replicating DNA and enclosing it in membranes. No one knew how this was being done – when the decision was being made, or what molecules were controlling it."

Yeast cells may form between one and four spores (below); the number is determined by environmental factors.

Right: Fluorescently labeled proteins allowed Michael's group to follow the decision-mking process whereby cells make spores. Bright green spots reveal active spindle pole bodies to which meiotic plaque proteins have been attached; these will lead to spores.



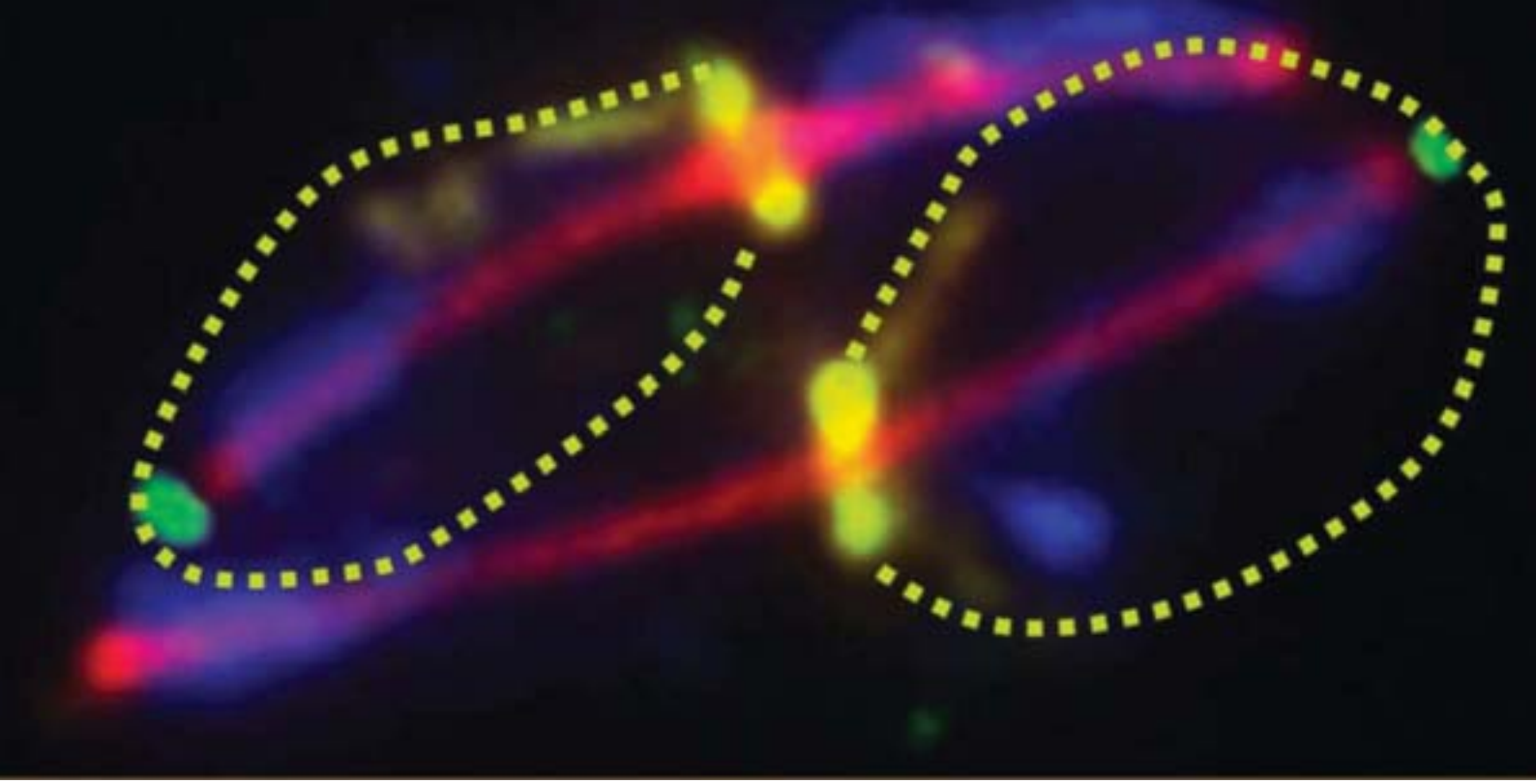
Scientists have identified some of the molecules involved in defining the number of spores that are made, and have discovered that SPBs play a crucial role. Now postdoc Christof Taxis and other members of Michael's group have obtained a glimpse into the complete decision-making machinery that links the environment to spore-forming behavior. Like the formation of the spindle that divides DNA in our own cells (see previous story), this is a feat of self-organization which follows from the nature of the elements involved and a set of rules that governs their behavior.

"One idea that has been around is that the number of meiotic plaques determined the number of spores that would be made," Michael says. "But until we discovered what they were made of, MPs couldn't be seen under the microscope. Once we knew that three crucial proteins were involved, we could tag them with fluorescent markers, count them, and watch their behavior over time." The group had a method of synchronizing the spore-forming behavior of many cells by carefully controlling the presence of energy in the form of acetate in the medium in which they were grown. They discovered that cells produce exactly the same number of spores as the number of MPs that are formed.

But how was this connected to the duplication of DNA? "You would think the two processes had to be connected," Michael says. "The cell ought to know how much DNA to make, otherwise it might end up with spores that contained incomplete sets, and that would be a disaster. What we've seen is that cells always make enough DNA to create four spores; if they suddenly decide to make only two, the extra DNA is lost and is somehow broken down."

The scientists wanted to test whether changing the number of MPs would influence the number of spores. They did this by changing the amount of building materials available to create plaques, making strains of yeast with one, two, or four copies of the three crucial MP genes. There was a direct correlation between the number of MP proteins and the number of spores that developed. "So we wanted to see what happened when the supply of MP building materials was cut off," Michael says. "We deprived the cells of energy and discovered that they reduced the production of the three crucial proteins. This means that there are sensors for the environment that regulate the genes for these molecules; they determine how much MP material will be built, and thus how many spores are made."

"Another reason why this is interesting is that the amount of nutrients in the environment varies in an 'analog' way,"



Michael says. "But the cell turns this into a 'digital' output – it chooses between one, two, three or four spores. We've shown that this transformation happens at the level of the MP components. Depending on how many are made, there's a self-organizing effect on how many plaques are built, and thus how many spores form."

At this point the researchers decided they knew enough about the process to try to model it with the computer. They created a system that could do three things: first, it translated an amount of energy available to the "virtual cell" into predictions about how many MP components would be made. Next it predicted how many MPs would be assembled from these components, and finally, it suggested how entire populations of cells would behave under these different conditions.

The scientists tested the model by trying to predict how real cells would behave if some of the conditions were changed – for example, if they were made less sensitive to changes in the environment, or if they had extra copies of some of the genes. The simulations matched outcomes of real experiments, and Michael is hopeful that the system will now help them understand what goes wrong in other types of mutants.

* * *

Spores not only ensure that yeast can survive hard times – they also influence the "quality" of future generations of cells. The way that the spores are made has an influence on evolution.

When a yeast cell copies its entire genome, and then splits it into four packets, the result is four sets of 16 single chromosomes – call them A1 and A2, B1 and B2. A1 is virtually identical to A2, and B1 is almost identical to B2. There are larger differences between either A and either B.

Michael and his colleagues wondered what happens when only two spores are made. The cell might be left with twins – two As, for example – rather than an A and a B. "The latter option is better for the long-term, because it preserves genetic richness," Michael says. "This means that if there's a mistake such as a mutation in a gene, it will only occur on one chromosome, and the healthy copy of the gene on the other chromosome might protect the organism."

The scientists created special strains of yeast, which permitted them to track how the spores recombined. When the cell built only two spores, there were nearly always an A and a B.

"The number of MP components shows the cell how many spores it has the energy and material to build," Michael says. "If it can only afford to build two copies, then it keeps those that give it – statistically speaking – the most diversity."

Thus the environment works through yeast genes to help the organism adapt to changing conditions. By influencing the construction of membranes, key proteins help ensure yeast survival and also the diversity of its genome. ■


Rush hour
on the nano metro

Arne Seitz and Thomas Surrey

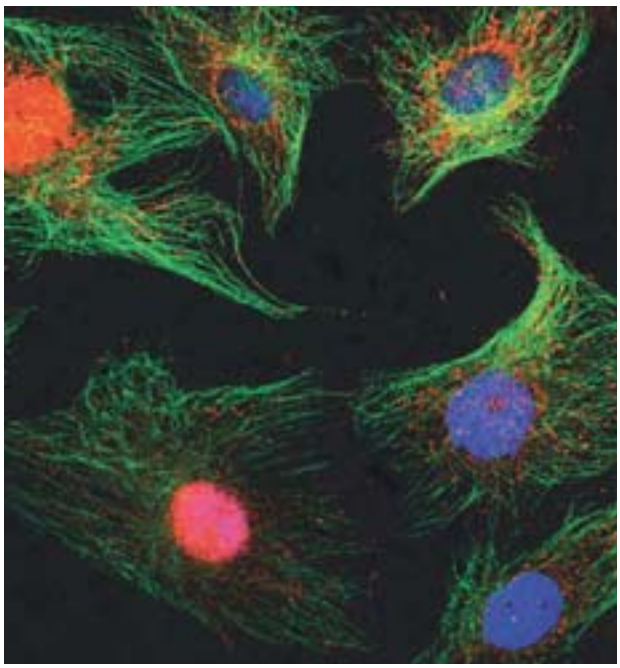
7:30 A.M.: traffic on the London Underground is building to a peak. At every stop, harried underground conductors shout over the loudspeakers: *Let passengers off the train before boarding. Don't try to crowd into a car that's already full.* The rules of physics seem to be off. No matter how many people have been crammed into a space, you can always cram in a few more.

There are rush hours in the cell as well, and they also happen along tubes – the microtubules. Here the passengers aren't inside the tunnels; they sweep along the outside. In contrast to the London Underground (hopefully), things manage to travel along the same tracks simultaneously in both directions. Thomas Surrey wants to know how that can happen without crashes, even under the most crowded conditions.

He's not living in London, but in Germany, where immense traffic jams called *Staus* are a sport, a way of life. Scientifically, his group in Heidelberg is focusing on a molecule that doesn't seem perturbed by Staus: *kinesin*, a motor protein which tows molecular cargoes to their proper destinations in the cell. Using a new microscope technique, Thomas and Arne Seitz, a postdoc in his group, have just learned that kinesins are patient users of the cellular railway system.

"One reason we picked kinesin is because its mechanics are well-known," Thomas says. "A typical motor takes about 100 'steps' along a microtubule before it lets go. As it moves it tows its cargo along on a flexible line. It tows things like protein complexes, molecules packed in vesicles, mRNAs, viruses, and organelles."

Careful structural studies of kinesin have revealed that it has two "heads" which dock onto the surface of the microtubule in alternation. A chemical transformation takes place; one head is released, and it swivels over the other in a "hand-over-hand" motion. This explains the walking motion, which is driven by the energy molecule ATP.



Two *heads* which *walk* in a *hand-over-hand* fashion? Thomas laughs and shrugs. "That's how we describe it," he says.

As it walks, the molecule competes for – footing? – with a wide range of motors and other proteins, all single-mindedly intent on delivering their cargoes. Very little is known about what happens when they encounter each other. You would think they would change speed, or stop altogether, Thomas says, but these questions haven't been addressed in a controlled and systematic way.

Arne and Thomas combined two microscopy techniques to better watch motor proteins directly, at the level of single molecules. Arne attached semiconductor nanocrystals – otherwise known as *quantum dots* – to kinesins and observed them using a method called *total internal reflection fluorescence microscopy*. The dots have several advantages over other types of tagging, such as fluorescent proteins like *GFP* (see previous story). Those molecules cast off their illumination in a short burst, and are then bleached and can no longer be seen. This means that observations are always an average of the behavior of many different molecules. Quantum dots are brighter, which makes them stand out against a background of other fluorescent molecules. They fluoresce for much longer, and so an individual dot can be tracked for extended periods.

Arne and Thomas decided to test two different conditions. First, they wanted to know if simple overcrowding changed kinesin's behavior. Under some circumstances in the cell, thousands of kinesins might be moving along

the same track. The scientists loaded microtubules even further, almost to the saturation point.

"Individual molecules move at different speeds," Arne says. "We thought that the presence of so many motors might slow the faster ones down. Instead, what we found was that motors did not really fill their track completely as long as they were in motion."

No matter how many motors had a chance to jump onto the track, the researchers couldn't get those molecules to form a traffic jam. In spite of the numbers, kinesins continue to move at a normal range of speeds, a feat that would make any transportation engineer envious.

The second experiment mixed normal kinesins with a variation of the motor engineered to be incapable of walking. This mutant lands on a microtubule and gets stuck there for a while, staying ten to twenty times longer than a normal kinesin would occupy any one position, then falls off. "What happens when the normal, fast-moving kinesin comes along?" Arne says. "Will there be a crash? Will the fast one be thrown off the track?"

In contrast to his experiences with the German highway system, Arne found normal kinesins to be careful and respectful drivers. When they encounter an obstacle, blocking a position that they need to take the next step, they wait patiently for the slowpoke to leave. Then they continue their walk until they encounter the next obstacle or reach the end of their run. This shows that during the waiting period, they must be strongly bound to the microtubule surface.

"The difference between this situation and that in the normal cell is that cellular proteins rarely stick to the microtubule and stay there, the way our mutant kinesins do," Thomas says. "Normally a motor doesn't have to wait very long for its next foothold to get free. This shows that kinesins are able to do so, however; they wait on the microtubule without letting go. It's a good solution to the crowding problem."

Could the findings be of any practical use?

"Oh, absolutely," Thomas says. "We received funding for the project from the German government. They want to apply what we've learned to the German Autobahn." A moment's pause, and he smiles. "I'm just kidding." ■

The microtubule system (above left, in green) plays a key role in organizing the cell's contents. Thomas and his colleagues are studying how motor proteins move along single microtubules (right).

Eavesdropping on the cell

A FEW YEARS AGO, the United Kingdom began building a system to monitor traffic on its highways. The idea was to set up sensors that would capture a dynamic picture of how many cars were on the roads and how smoothly they were moving. The data was fed into a centralized computer and could be used to detect traffic jams, warn motorists of problems ahead, improve the timing of traffic lights, and solve other types of problems.

Scientists would like a similar monitoring system to watch the flow of chemical information in the cell as

Heike Stichnoth, Alen Piljic, Adrian Neal, Carsten Schultz, Sirus Zarbakhsh, Andreas Füssl, Christiane Jost, Andrea Giordano, Tatsiana Skrahina, Oliver Wichmann, and Amanda Cobos

things unfold. Molecular signals steer a wide variety of processes: they coordinate the timing of events during cell division, pass along critical information that tells cells how to develop properly, alter the sensitivity of nerves, and often become disrupted during disease.

Several years ago, Carsten Schultz and others realized that getting a handle on certain diseases would require new, very sensitive types of probes able to monitor intracellular signaling events. His group has been helping to design them. "Many diseases involve a communications abnormality, and it's crucial to see exactly when and where in the system this occurs," he says. "It's not enough to say 'There's a traffic jam somewhere, everyone should stay off the roads.' If we want to have any hope of repairing problems without damaging healthy processes that are going on, we'll have to be able to monitor the complexity of cell signaling as precisely as possible – and we'll also have to monitor the effects of drugs on this signaling network."

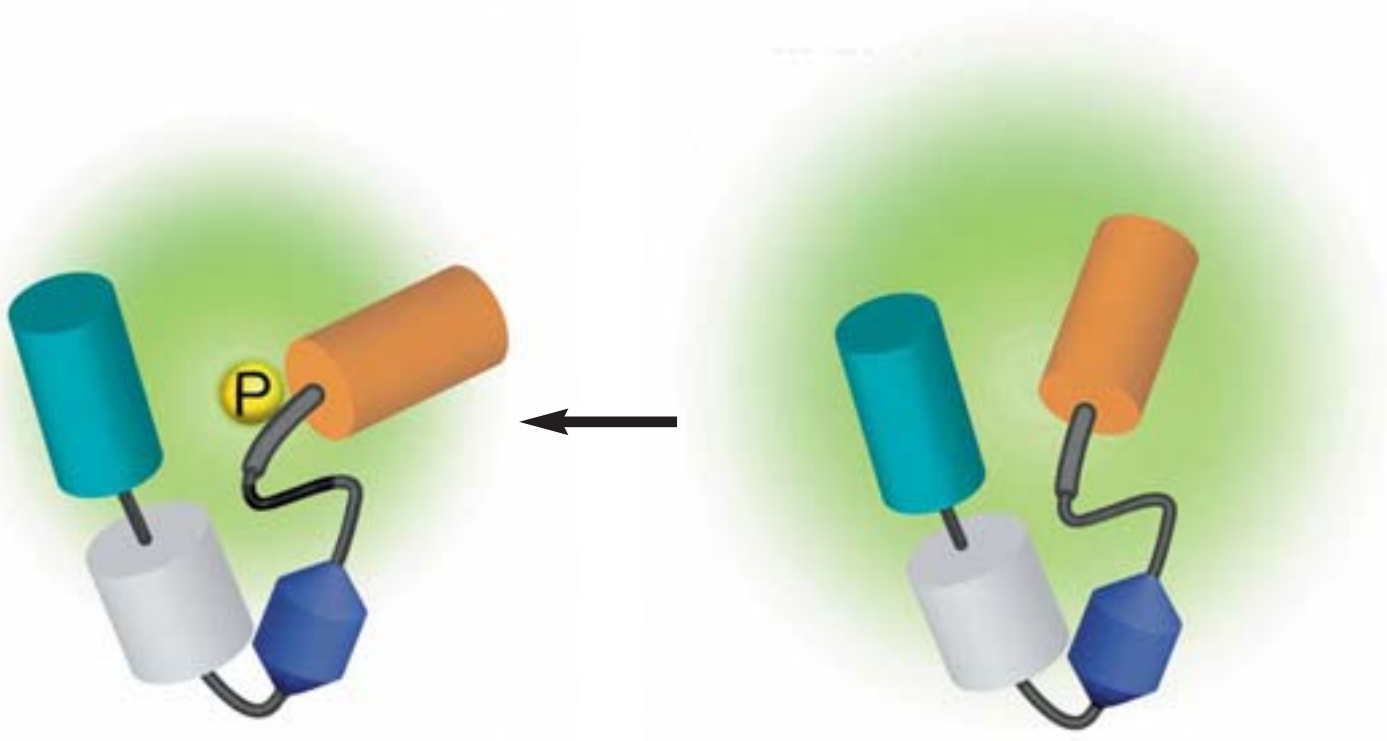
In the last ten years, several probes have been developed based on fluorescent tags that have been genetically added to proteins. The method called *intermolecular FRET*, for example (see page 24) allows scientists to detect when two proteins bearing different tags interact with each other. Two years ago, Carsten's group created a signal monitoring molecule which had two tags in the same protein. When the molecule was activated by another protein, it changed its conformation, bringing the tags close enough together to increase its FRET signal.

Now Andreas Schleifenbaum, Justin Brumbaugh and other members of Carsten's group have created a new variation on this sensor that can be used as a roving monitor for a particular type of cellular communication network called *phosphorylation*. This is one of the most common types of signaling in the cell: it involves transferring chemical modifications called *phosphate groups* between proteins.

"In a collaborative effort with the structural biology group of Michael Sattler, funded by the Volkswagen Foundation, we came up with a molecule that would show a conformational change whenever phosphorylation occurs," Carsten says. "This molecule, *pleckstrin*, is common in various types of blood cells, but it's absent from almost all others. Where it occurs naturally, it receives a signal from a signaling molecule found in all types of cells, called *protein kinase C* (PKC). If we put pleckstrin into other types of cells, it would also receive this signal, probably without interfering with other signaling molecules as they go about their business."

Pleckstrin had more advantages. For instance, it turned out that pleckstrin likely undergoes two different conformational changes, opening more possibilities to give it sensor functions.

Most signaling molecules contain a *recognition sequence* – a sort of lock that can only be opened with a particular key. Usually only one other molecule has this key, which



is how the cell keeps complex signaling networks from interfering with each other all the time. Pleckstrin comes with a very generic lock that can be phosphorylated by many different proteins. It could be adapted and specialized by adding particular locks, making it sensitive to only one type of signal.

"When pleckstrin is activated, two of its modules shift positions and are brought together," Carsten says. "We added a green fluorescent probe (GFP) to one module, and a yellow probe to the other. This means we could use FRET to detect the different activity states. When we tested the probe in living cells, it was very clear when PKC signaling occurred."

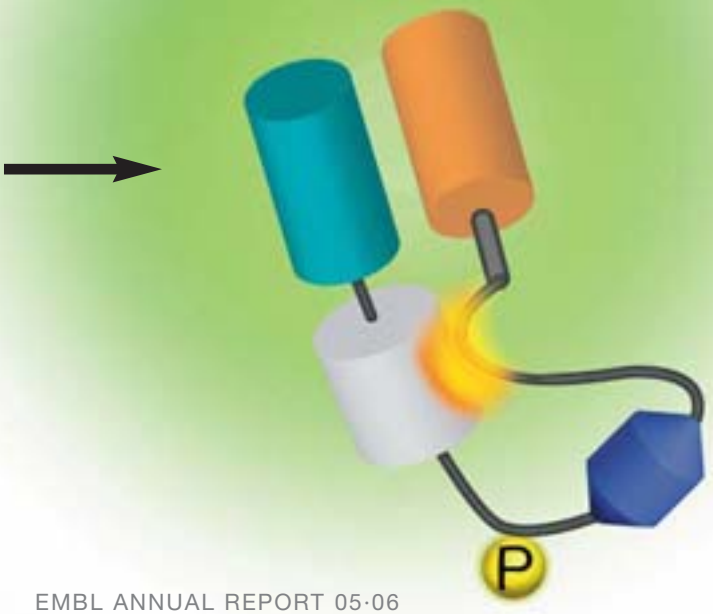
They called this modified version of pleckstrin *KCP-1*, and now they wanted to alter it to see if it could be used to detect signaling via another molecule, called *protein kinase A* (PKA). "The first thing we tried was to replace the PKC activation site with one that was known to be activated by PKA," Carsten says. "This was only partially successful. The outcome was usually a molecule that was still sensitive to PKC, or to neither. It was a little like changing the locks of your house so that only people with specific keys can get in – either nothing had changed, or we were locking everybody out."

Back to the drawing board in the lab. The scientists started to make more dramatic alterations to KCP-1, removing sections close to the target area. They came up with a version called KCP-2 which didn't do any better at sensing new signals – apparently using another change in the arrangement of the fluorophores within the molecule.

"We had somehow fundamentally changed the probe's structure – maybe we hadn't initially understood what was happening to bring the fluorescent tags together," Carsten says. They took their samples back to Michael Sattler, who helped study them using NMR. Experiments revealed that a region of the tail of the protein was probably responsible for the original conformation change. Without the tail, only a basic change was visible – monitored by KCP-2.

The scientists created more spinoffs of the probe, this time with alterations in the tail area. Introducing a PKA-sensitive sequence made a version that was sensitive to PKA and PKC signals, giving responses in different directions. By eliminating PKC activity, the probes were only activated by PKA.

Carsten says that what they learned should now be applicable to creating a wide range of new probes, each specific to a particular type of signaling. "One success of the project has been to create an artificial protein that acts as a FRET sensor based on an intrinsic remodeling of the protein conformation," he says. "There's an advantage to doing this with a single molecule, rather than the way it's often done, with an artificial construct. And our findings about the tail should make it relatively straightforward to make new probes, even to monitor completely different types of chemical signals. Eventually, these tools should allow us to make a much finer dissection of the cell's signaling pathways." ■



Modifying the new sensor molecule KCP-2 (middle) gives scientists a way to monitor two signaling pathways in the cell. One pathway changes the conformation of the molecule so that it gives a weaker FRET signal (left); another produces a stronger signal (right).

Alan Sawyer, Mike Spiegel, and Melanie Leuener in the Monoclonal Antibody Core Facility at EMBL Montorotondo.

An antibody assembly line

MANY OF THE MOST POWERFUL METHODS of the biotechnology revolution were created by turning molecules from one organism into tools that could be used to study or alter another. This has been the case with antibodies. These proteins are produced by our bodies; they bind tightly to other molecules, which suggested that they could be used to locate and observe proteins in living systems. Scientists now use them in hundreds of different ways to carry out basic and medical research, but creating each one is time-consuming and costly. Alan Sawyer and his team in Monterotondo have now developed a method to do this much more quickly and inexpensively.

Antibodies are created within white blood cells called *B cells*. As these develop from more generic cells, each shuffles around components of antibody genes in random ways, so cells produce these proteins in an almost unlimited number of forms. In the end, each fully-developed cell has its own unique antibody. Each is capable of binding to a unique partner called an *antigen*, which might be a protein on a surface of a virus or another foreign molecule. The randomness behind their production is crucial: it gives the immune system the chance to confront completely new challenges. When an antibody finds a match, the cell is stimulated to divide and create more copies of itself and these in turn churn out and secrete more copies of the antibody. These bind to other copies of the invader and attract cells that can destroy it.

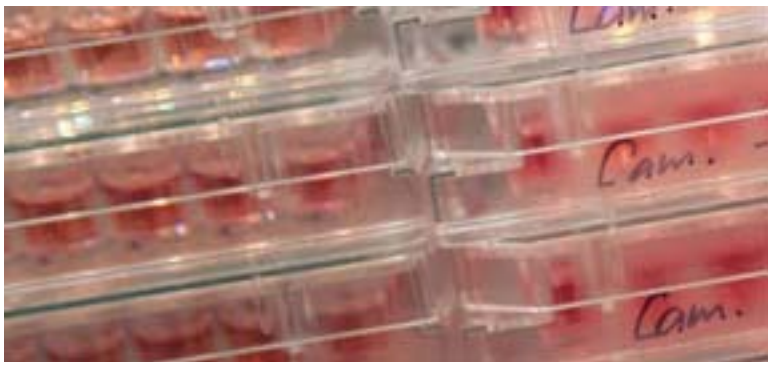
There are many contexts in which scientists need to create a molecule that can bind to another – why not have B cells do it for them? "It was a nice idea, and theoretically it could work, but there was a problem," Alan says. "If you need antibodies for a particular molecule, you can raise them by injecting the molecule into a rabbit, for example. But there's no guarantee that the antibodies you get will be truly specific for this one antigen – they may bind to other proteins as well. And if you do get a B cell that produces the exact molecule you need, it's impossible to grow the cell in cultures, which is necessary to obtain sufficient, pure quantities."

In 1975, working in an MRC laboratory in Cambridge, César Milstein and Georges Köhler discovered a way to circumvent this problem, work for which they won a Nobel Prize in 1984. By fusing a B cell that produced a specific protein to a particular type of cancer cell called a *myeloma*, they could create *monoclonal* antibodies – unlimited, identical copies of the molecule stemming from a single mother clone. "These fused cells are a sort of marriage where the B cell brings along the capacity to make antibodies, and the myeloma reproduces indefinitely, in cultures," Alan says. "We've been making them basically the same way for two decades. Creating each one is a time-consuming process, with a lot of hands-on care. It has to be done very precisely, but it's the same job, over and over again, which is hard on a technician. Even a good facility can only turn out so many per year."

A few years ago, Alan began thinking about how the process might be streamlined and automated. EMBL's scientists needed antibodies all the time, usually turning to outside companies to get them. The Laboratory was in the

Federico di Masi working on protein chip-spotting equipment.

Below and right: the automated monoclonal antibody pipeline. B cells are fused to myeloma cells, robotically transfered into wells and grown, and printed onto protein microarrays. Cells which successfully produce the desired antibody are grown in culture, and the molecule is then extracted and purified for use.



process of creating Core Facilities to operate some central services – the same could be done with antibody production.

"This fit with EMBL's style of developing technologies," says Christian Boulin, who heads the Core Facilities. "EMBL has a history of innovations that begin within research or technical groups, and are then further developed as services or facilities. This goes back to the early days of the lab; one of the first examples was building new types of detectors for use at the beamlines. There have been several other cases: Jacques Dubochet's group developed cryo-electron microscopy; Wilhelm Ansorge's group designed new types of high-throughput DNA sequencers; Ernst Stelzer and his group made innovations in confocal and other types of microscopes; Matthias Mann and Matthias Wilm helped turn mass spectrometry into a tool for proteomics. Many of these projects planted the seeds for our current Core Facilities."

Alan started to assemble his antibody pipeline in Heidelberg; then PhD student Federico di Masi from Wilhelm Ansorge's group jumped on board. "The first phase was spent ironing out the robotics and proving the principle," Alan says. "Since our move to Monterotondo we have entered a stage where we've solved most of the problems and are now into high-throughput production." Not just for EMBL, he says. If things continue to improve, the group will be able to turn out an average of 450 antibodies per year. That's close to the current annual needs of researchers throughout Europe. "Although those needs are constantly expanding to meet capacity as the costs drop," Alan says. "We are essentially expanding the market."

* * *

The innovations introduced by Alan and his team involve ways to automate, or semi-automate, most of the labor-intensive steps in creating monoclonal antibodies. First, rather than challenging a single mouse with a single antigen, they inject the same mouse with several. Its immune system is capable of responding to several targets at the same time. They extract B cells from the spleen of the mouse and fuse them to myeloma cells using state-of-the-art robotics. The hybrid cells are further robotically transferred into "wells" – 96 separate compartments in one plastic dish.

This setup is important, Alan says, because of the new method the scientists have created to solve one of the most time-consuming steps: screening to see which cells are producing the correct antibody. "Basically what we have are millions of cells, only some of which produce the molecule we need," Alan says. "When Federico joined the project, he brought along expertise in making microar-

rays – protein chips. We've turned that technology into a way to find fusion cells that make the antibody."

As cells grow in the wells, the scientists prepare the protein microarrays: glass slides onto which a "lawn" has been uniformly coated. A robot dips needles into the 96 wells, collecting samples, and prints them onto this protein layer. If cells in one of the wells produce the proper antibody, it will bind to the antigen as a small spot. They still have to be detected – so now a fluorescent molecule is added that recognizes successful antibody-antigen pairs. "The neat little trick we thought of," Alan says, "is to put in a mixture of dyes. Antibodies come in different forms called *isotypes*, and each dye is specific for one of them. This allows us to detect not only the specificity for the antigen, but simultaneously what kind of antibody is binding. The presence of a spot shows us that an antibody is bound and the colour of the spot shows us the antibody's isotype." The fluorescent spots are detected with a confocal laser scanner, and it shows the scientists which well contains the fusion cell that they need. All that remains to be done is to extract it and grow it in culture, then extract and purify the antibody.

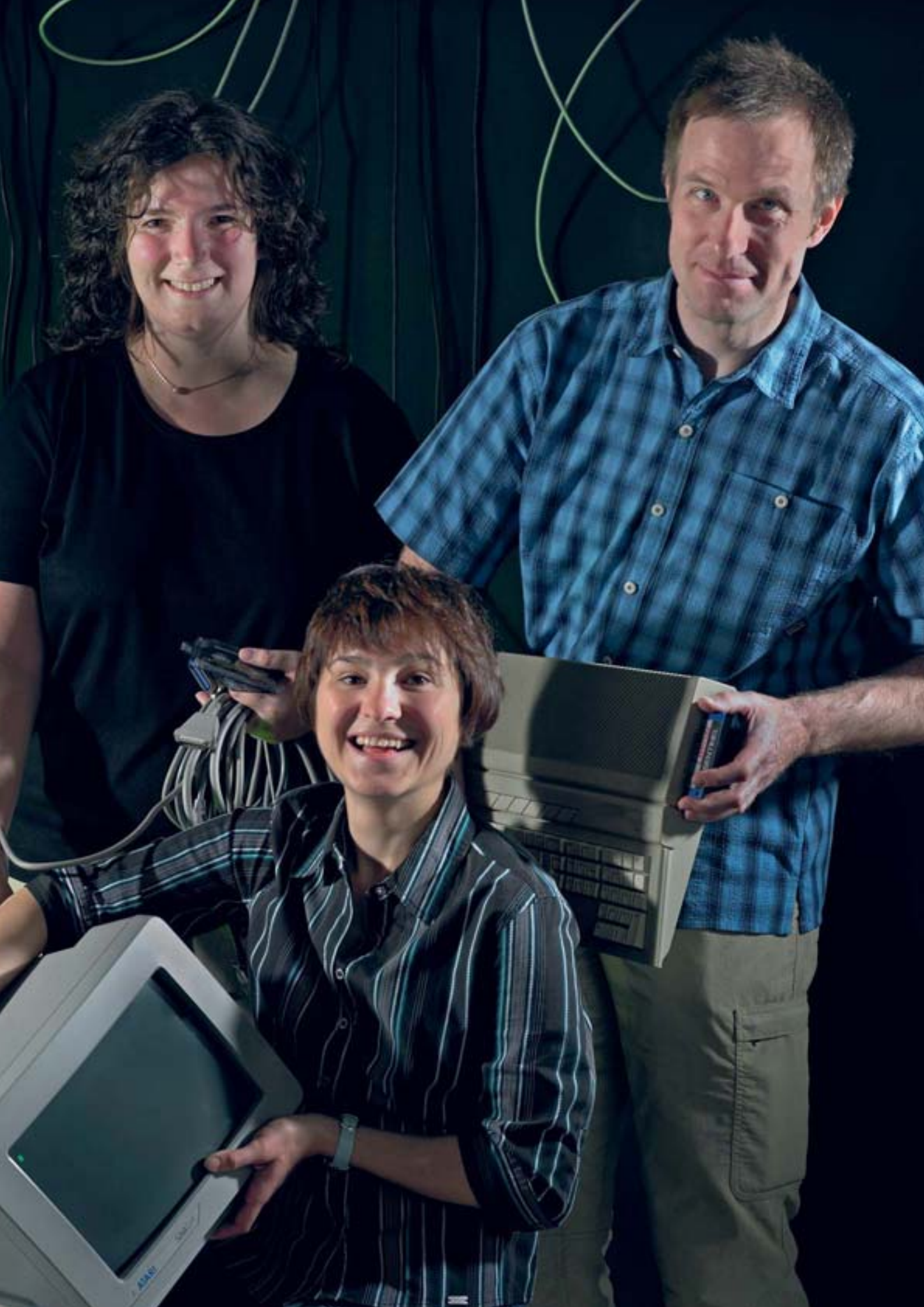
Would the method also work in a high-throughput way? As a proof of principle, the team injected 80 separate antigens into eight animals – ten apiece – and checked to see how many antibodies were produced using the method. Within six weeks of the initial immunization, the group had obtained 68 antibodies, a success rate of 85%.

"To put this in perspective," says Alan, "in a full run of 80 antigens, testing every sample in triplicate, we are looking at 460,800 individual tests in less than 24 hours. That would be impossible using the current detection methods."

The microarray screening method is more sensitive than methods that have been used in the past. "We would have missed a considerable number of hits; this allows us to find successful antibodies that have been produced in lower numbers," Alan says.

The work has attracted the interest of major pharmaceutical companies whose scientists need large numbers of different antibodies, and EMBL's technology transfer company EMBLEM is exploring ways to arrange exchanges of services and expertise. Antibodies have a bright future, Alan says.

"In the genomic age, we need high-throughput technologies that can inform us about the functions of all the molecules we've discovered," he says. "DNA chips and other high-throughput technologies can tell us whether a cell is making RNAs, but not whether it is making a particular protein. Antibodies are still the method of choice for screening tissues not only for their presence but also their activation state, and we're removing the most significant technological obstacles to creating lots of them." ■



Bernadett Papp (back), Tanya Klymenko and Jürg Müller

Memories of silence

WHEN THE FIRST PERSONAL COMPUTERS appeared in the early 1980s, they had no hard disks and very little memory. Turned off, they forgot everything but the most rudimentary bits of the operating system; most applications had to be loaded from floppy disks every time you started the computer.

Why don't cells forget their "programs" when they undergo cell division? Two new daughter cells have the same "hardware" as their parent, but also operate some of the same "software": patterns of using genes. Cells on the path to becoming neurons express one pattern of genes; those becoming muscle, skin, or blood express other patterns. These have to be maintained for many rounds of cell division. By activating some genes and silencing others, they acquire distinct properties.

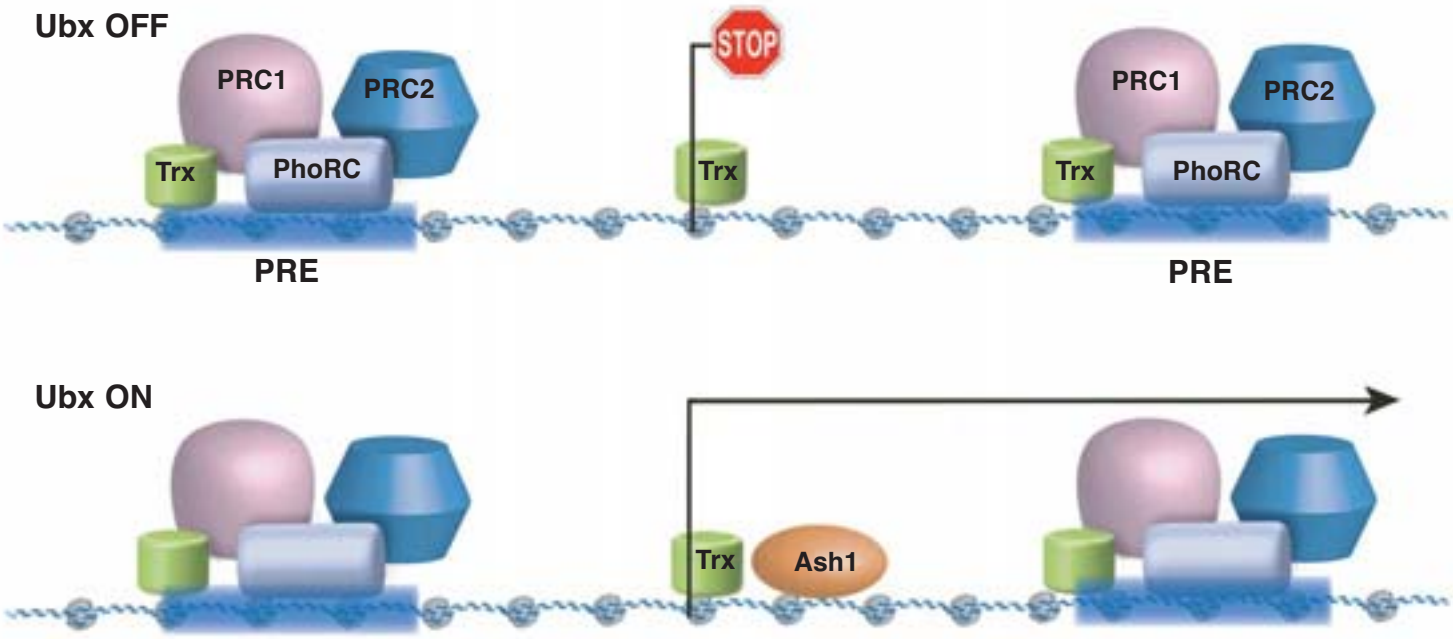
Jürg Müller and his group in Heidelberg hope to understand the mechanisms that underlie this type of cellular memory. They know that it depends on how proteins manipulate chromatin: DNA and the molecules it is bound to in the nucleus of cells. Computers rely on microchips and hard disks to manage memory; cells use a language of chemical markers attached to chromatin.

* * *

In the nucleus of every human, animal and plant cell, the double-helix DNA strand is wrapped around spool-like units of eight *histone* proteins. Scientists knew that this spool structure, called a *nucleosome*, plays a key role in activating or inactivating genes. Each histone has a tail which protrudes from the spool and is subject to a large number of chemical modifications; some proteins attach those chemical tags, others bind to them or strip them off again. This often changes the activity of genes. Studies over the past few years have revealed that the attachment of *methyl groups* at particular positions on histone tails often leads to genes being switched on or off.

Having the right combination of genes on and off in a given cell is crucial to the development of organisms, and thus the methylation code likely serves as a sort of patterning memory. To decode the programming, scientists are looking at *HOX* genes, discovered in the early twentieth century, which play a key role in the development of embryos. Activated in a specific order in the right tissues, these molecules help organize and structure all animal body plans.

Over long phases of development, HOX genes need to be kept active in the cells that need their expression but it is equally important to keep them silent in other cells, despite the constant interruption of cell division. Studies of mutant flies have shown that *Polycomb* (PcG) and *trithorax* (trxG) groups of proteins play a key role in maintaining these states. TrxG molecules are responsible for keeping genes active whereas Polycomb proteins usually silence genes by binding to target gene DNA at sequences called *Polycomb Response Elements* (PREs). Both types of molecules act by attaching methyl tags, but they do so at different sites on histones. Yet how this generates on and off states is largely unknown.



"People have recently come to believe that one of the first steps in silencing is for PcG proteins to methylate histones at PRE sites," Jürg says. "That chemical change supposedly attracts other PcG proteins specifically to PREs;and these then repress genes. We now believe that this process happens quite differently."

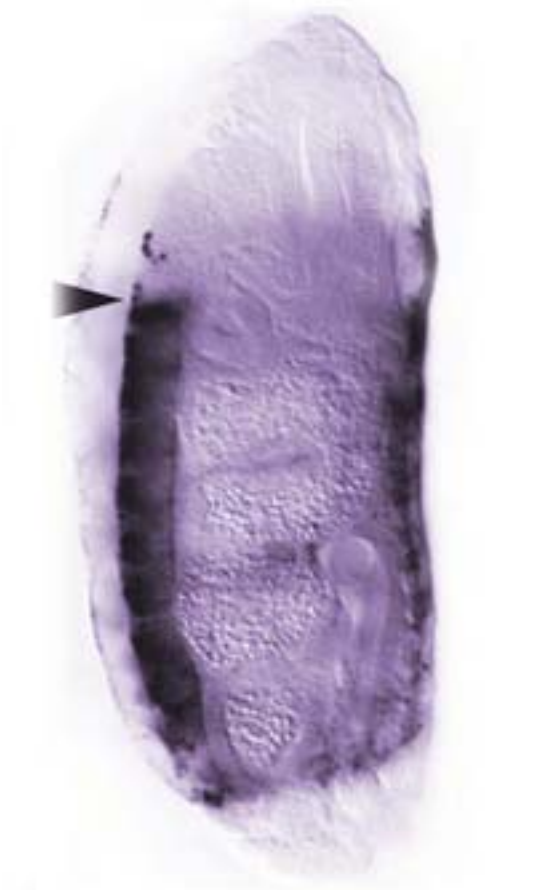
Bernadett Papp, a PhD student in Jürg's group, wanted to clear up two questions: Are PcG proteins only bound when genes are silenced? And what parts of genes are actually methylated? She chose a HOX gene called *Ubx*, which is on in some tissues and off in others, with important effects on development. She compared the binding of PcG proteins and the histone methylation of Ubx in the two states.

To discover what proteins were bound to Ubx, Bernadett used a method called *ChIP* (described on page 156). She made some surprising discoveries. PcG proteins were bound to the gene whether it was turned on or off. This meant that the binding of these proteins alone wasn't determining the on or off state of the gene; other molecules had to be involved in regulating its activity. Consistent with this, Bernadett discovered that not only PRE sites, but histones along the entire gene were marked with PcG-type methyl tags. Was this important, or was the methylation of some regions of the gene more crucial than others?

She found an important difference in one region: when Ubx was on, the PcG-methyl markers were absent in the protein-encoding part of the gene. This region was therefore playing a crucial role in differentiating between the two states.

The same couldn't be said for the presence of PcG proteins, which were bound to Ubx whether it was on or off. But did TrxG proteins behave the same way? Bernadett's experiments uncovered a TrxG protein called *Ash1* which was only present when the gene was on. Suddenly she had a switch.

"Ash1 blocks the attachment of PcG-type methylation marks to the coding regions," she says. "We tested this by removing Ash1 from the fly. Automatically, PcG-type methyl tags were added to the coding region of the gene, and Ubx was shut off in cells where it is supposed to be on. Conversely, when we blocked the methylating activity



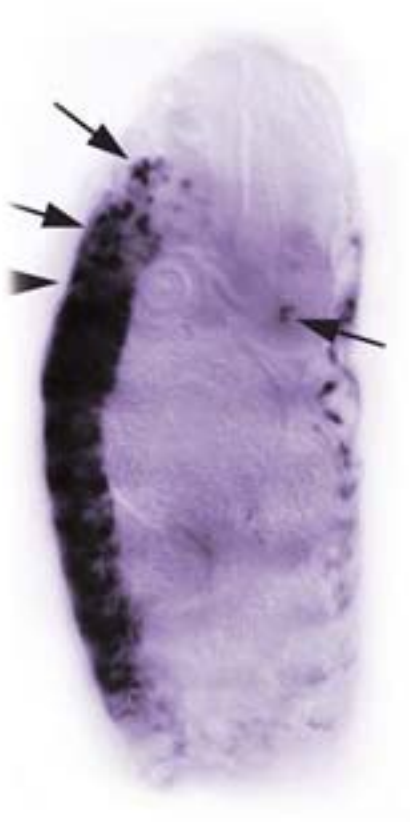
Left: PcG proteins like PhoRC are bound to some developmentally crucial HOX genes all the time. When the protein Ash1 binds, repressive methyl tags are no longer added to key positions in the gene, and molecules which should be silent are turned on.

Below: When a fly embryo is 14-16 hours old, Ubx protein is produced in cells of the central nervous system. The left image shows the normal expression pattern (dark purple band). When the protein dSfmbt is removed, the Ubx gene is no longer silenced in important cells, and they produce the protein in inappropriate places (arrows on the right).

of PcG proteins, cells began turning on Ubx when it shouldn't be."

* * *

PcG proteins only bind to PREs, which are very local sequences that are often located far from the coding regions of genes. How do they influence what happens so far away? Little was known about how PcG molecules bind or behave; Tanya Klymenko, a PhD student in Jürg's group, wanted to find out.



She chose a PcG molecule called *Pho* because among 15 different types of PcG proteins, it is the only one that binds directly to DNA. This suggested it might be acting as an anchor, seeking out and binding to a PRE and then helping to assemble other proteins on the site.

Tanya purified protein complexes containing Pho in search of cellular partners that might say something about its functions. Members of Matthias Wilm's mass spectrometry group lent a hand in identifying the other molecules attached to the protein. Unexpectedly, they found that Pho is not associated with any of the usual suspects (the already known PcG proteins). Instead, Pho operates in two different machines: one including factors that help remodel nucleosomes, and the other containing a protein called *dSfmbt* that had never been studied.

Tanya and her colleagues focused on dSfmbt because she found that it was also attached to PREs. First the scientists demonstrated that it had an important role to play. "When this gene is removed from the fly, HOX genes are switched on in tissues where they should be silent," Tanya says.

How does dSfmbt help silence genes? "The protein contains modules called MBT repeats," Tanya says. " In collaboration with Wolfgang Fischle , we discovered that MBT repeats bind to histone tails that have a particular kind of methylation."

However, they knew that a PRE doesn't have to be methylated to attract dSfmbt; they had strong evidence that it is brought there by Pho. "One possibility would be that dSfmbt, tethered to the PRE, scans the methylation status of flanking histones," Jürg says. "We imagine that it recognizes nucleosomes with particular methyl marks and docks onto them. This draws the PRE, and the other molecules attached to it, close to those nucleosomes. Now other proteins in the complex can go to work on the histone tails in these regions. Some of them can attach even more PcG-methyl marks. Others probably act as roadblocks to the gene-transcribing machinery."

This is only a model, Jürg cautions, but if true, it might help to explain the memory behavior provided bythe PcG/trxG system. Bound at the PRE, Polycomb proteins generate extended stretches of chromatin that are covered with PcG methyl marks. This keeps a gene switched off, and the high levels of methylation on histone tails are stable enough to survive the "reboot" that takes place after cell division. ■

Complementarity and the fates of cells

IN HIS BOOK *THE DOUBLE HELIX*, James Watson recounts how he experienced a flash of insight that would change biology forever. Watson, Francis Crick, and their colleagues in Lawrence Bragg's laboratory in Cambridge had been struggling to understand the nature of DNA. They ordered models of the chemical building blocks of DNA, called nucleotides. Watson was fitting the pieces together and noticed that when nucleotide *A* was joined to T, it had the same size as *G* joined to *C*. Just as the rungs of a ladder ought to be the same length, this was crucial to understanding measurements that had been made of DNA by Rosalind Franklin and her colleagues. If pairs of nucleotides, also called base pairs, had identical lengths, they could be stacked. If that was the case, and they were then linked by "handrails" of sugar on the outside, twisting along the way, they would form a double-helix ladder. This explained patterns that Franklin had obtained by exposing DNA to X-rays. Crick came in, Watson explained what he had found, and off they went for a drink at the Eagle Pub. The rest is history.

The structure of DNA is based on *complementarity*: an A always binds to a T, and a G to a C. Thus the complement of the DNA sequence *AAAA* would be *TTTT*, and the partner of the string *GAACCT* would be *CTTGGA*. Watson and Crick's groundbreaking publication of their discovery concluded with one of the most famous understatements of scientific history: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." Chemistry had suddenly been linked to cell division, heredity and evolution.

Complementarity is essential to understanding much of what happens in the cell. It explains not only the linkage between two strands of DNA, but also how RNAs (also made of nucleotides) bind to DNA and to each other. This chemical principle has led to a wide range of new technologies to monitor what happens within cells.

Understanding these principles and then improving the technology to manipulate and sequence DNA led to the genomic era. Something similar is happening today as scientists continually discover new ways that RNAs influence the life of the cell and put these findings to use in new technologies. From very diverse starting points, EMBL groups from all over the Laboratory are converging on RNA. The resulting discoveries are dramatically altering our view of the information contained in genomes and the complexity of cellular processes. ■


Meeting Point

Xouri Georgia, Irina Charapitsa,
Gäelle Legube, Sascha Mendjan
and Asifa Akhtar

Business at the meeting point

IF SOMEONE EVER GIVES YOU an alligator egg, you get to choose the sex of your future pet: incubating it above 33 degrees Celsius will produce a male, and cooler temperatures will deliver a female. The two sexes of alligators have identical chromosomes, and the sex of their embryos is determined by temperature and other environmental factors.

Hundreds of millions of years of evolution have created a different system for defining the sexes of flies, humans and many other species. Female cells have two copies of the X chromosome, while males have an X and a Y. The other chromosomes are found as pairs in both sexes. A single human X chromosome contains about 1000 genes; Y probably carries just 78. (This varies widely between species – a male kangaroos has but one lonely gene on its Y chromosome.) The two Xs give females double copies of hundreds of genes for which males only have one. Normally this would cause female cells to produce twice as many proteins from X-chromosome genes, which would be fatal because the two sexes need most of these molecules in equal amounts. They survive thanks to the co-evolution of *dosage compensation* mechanisms that maintain the balance.

Scientists have believed that this happens in two fundamentally different ways: in flies, molecules attach themselves to the single male chromosome and double its output. In humans, the opposite happens; one of the female X chromosomes becomes inactivated. "Switching on and off groups of genes in regions of DNA is crucial for organisms in other ways," says Asifa Akhtar. "For example, changes in gene expression guide the development of different types of cells, leading to the creation of tissues and organs in animals. So in addition to studying how dosage compensation is managed between the sexes, we are using it as a model to understand some of these other processes."

Now a discovery by her group may give new insights into the management of the X chromosome and other cases of dosage compensation.

* * *

Scientists have known for several years that the way in which DNA is arranged influences whether genes are kept "silent" or are used to make RNA and proteins. In the cell nucleus, DNA is cluttered with thousands of different types of proteins in a form called *chromatin*. Some of these proteins help fold and pack the genetic material into huge bundles. This can create roadblocks which prevent other molecules from gaining access to regions of DNA and transcribing genes into RNA molecules.

A few years ago, scientists discovered another method by which cells deactivate genes and sometimes regions of a chromosome. The membrane that surrounds the nucleus is full of proteins, too, and they can latch onto DNA and sequester it near the border of the nucleus, the way yarn might become entangled in a thorny rosebush. "This was first observed in yeast cells," Asifa says, "and the effect was usually to silence genes. So the periphery of the nucleus has obtained a reputation as a gene-silencing zone."

On the trail of differences between the sexes, PhD student Sascha Mendjan and other members of Asifa's group were studying proteins called *MSLs* that are produced in male flies. MSLs joins a complex of proteins called the *dosage compensation complex* (DCC) which helps increase the output of the single X chromosome in males. Thousands of these complexes dock onto X genes and double their transcription into RNA.

"Last year we discovered that the cells of vertebrates also produce most of the components of DCCs and combine them into a very similar machine," Asifa says. "That was puzzling to us because most vertebrates control dosage in the opposite way – by tuning down the second female X chromosome. Why would evolution conserve a machine in vertebrates that they don't need? Does it have a completely different function, or is it also involved in dosage control?"

Getting a handle on this question meant taking a closer look at the DCC. The scientists knew that this "molecular machine" contained at least five proteins, but there might be more. Sascha and another PhD student, Mikko Taipale, were able to extract the protein complexes from the cells of fly embryos and humans and identify their components.

"We found something very surprising," Sascha says. "First of all, we confirmed that the machines are built from the same basic molecules in fly and human cells. But then we discovered a new component – a protein known as *Mtor* in flies. The complex in humans contains a close relative of Mtor called *TPR*."

These molecules were familiar to the scientists because they have been the subject of intense study in other labs. They are found in *nuclear pores*, structures in the membrane of the nucleus; pores act as gateways to the surrounding cellular compartment, the cytoplasm. A further component of the DCC, the protein Nup153, is also found in nuclear pores. In 2001 Iain Mattaj's group discovered that alongside its role in building the pore, Nup153 acts as a doorkeeper: it helps shuttle molecules in and out of the nucleus.

Why would a nuclear pore protein be found in the DCC? How crucial was this molecule to the machine's function? Jop Kind, a PhD student, used a method called *RNAi* to remove Mtor from the insect cells. "Suddenly MSL proteins were no longer clustered in their typical location on theX chromosome," Asifa says. "Either they were no longer bound to the X chromosome, or the X chromosome itself was sprawling across the nucleus. Whichever was the case, male X genes were no longer producing double doses of proteins."

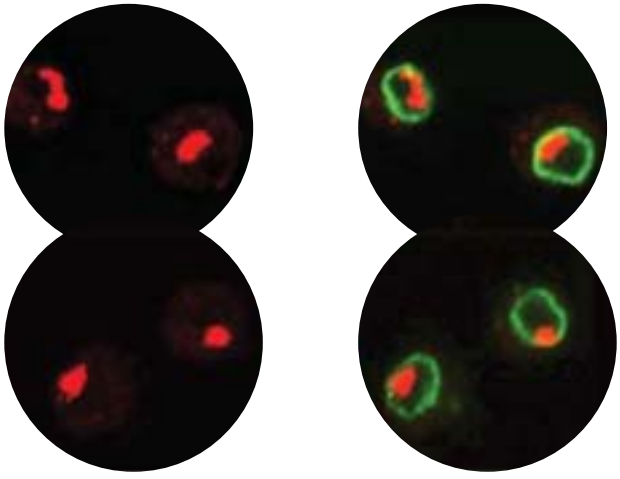
* * *

The scientists don't yet know why this happens. One hypothesis is that nuclear pore proteins may help generally organize chromatin in the nucleus, and this is subsequently required by MSL for dosage control. Another possibility is that by linking to other proteins in the DCC, Mtor helps tether the X chromosome to the membrane. Studies in yeast have suggested that DNA at the edge of the nucleus often contains a high number of inactive genes, and recently scientists have observed cases in which the region seems to have an activating effect as well. It might be serving as a region to control large blocks of genes. "But these are only some possibilities that need to experimentally tested," Asifa cautions.

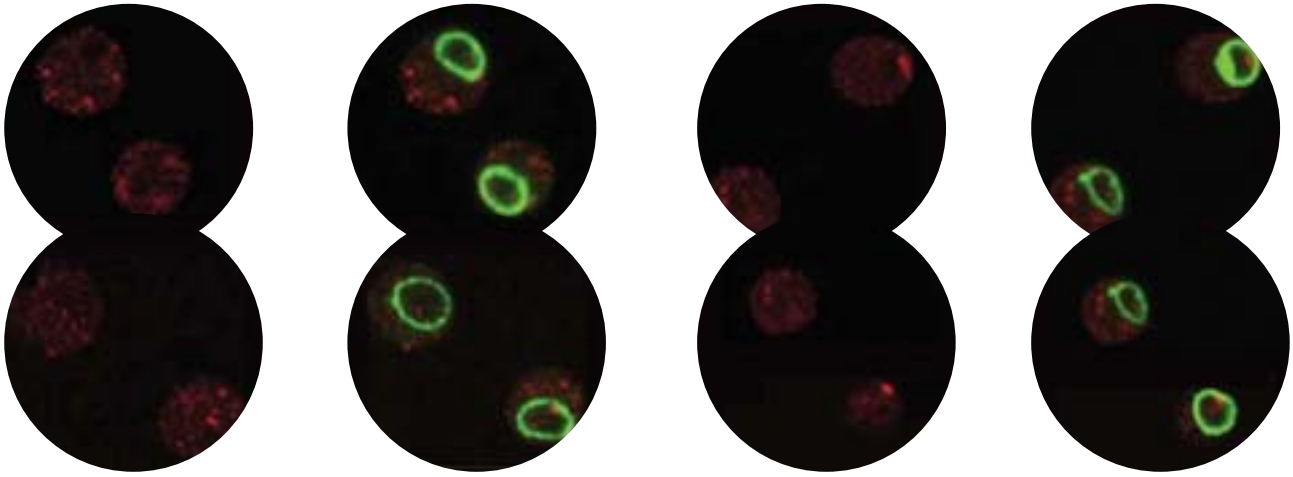
In this scenario, removing Mtor from cells might cut the mooring lines between the X chromosome and the membrane. The nuclear membrane, particularly the region of the pores, might make a good "meeting point" to assemble molecules that are needed to transcribe genes and ship them quickly outwards to the cytoplasm. This has to happen for mRNAs to be translated into proteins.

"Although mammals and flies solve the X-Y problem in different ways, it's interesting that a protein machine has been kept intact for at least 600 million years," Sascha says. "Whether it has also maintained a role in sex-related dosage compensation remains to be seen."

In research, it is often hard to distinguish causes from effects. It's tempting to see this type of dosage compensation simply as something necessary to maintain the existence of two sexes in flies and other animals. But the pieces of the DCC evolved long before sex, and were originally used for other things, possibly even dosage compensation in other contexts. Without the ability to regulate the productivity genes, there would probably be no Y chromosome. So from one point of view, the DCC is the servant of sex. From another, dosage compensation is the *cause* of it. ■



Nuclear pores create channels through the membrane surrounding the nucleus. Asifa's group showed that a nuclear pore protein belongs to the dosage compensation complex in flies, which binds to the X chromosome (above). This protein complex helps regulate genetic differences between the sexes. Microscope studies (below) revealed that the X chromosome is brought to the membrane. Without the nuclear pore protein, this localization doesn't happen, and dosage compensation breaks down.




EMBL

Kent Duncan, Claudia Strein
and Karsten Beckmann

Heads and tails

IN FLIES, MSL AND OTHER PROTEINS help regulate differences between the cells of the two sexes. Docking onto the single X chromosome of males, they increase its output in a sort of genetic affirmative action, helping genes become as productive as their relatives on the two X chromosomes found in females. But why wouldn't the same thing happen in female flies? Why wouldn't MSL and the rest of the dosage compensation complex (DCC) land on both chromosomes, once again leading to a double dose?

One reason, says Matthias Hentze, is that cells of female flies don't build a functioning DCC machine. "Most of the parts are present, but one crucial component is missing – the protein *MSL-2*. Getting a detailed look at why female cells don't produce this molecule has kept us busy for several years."

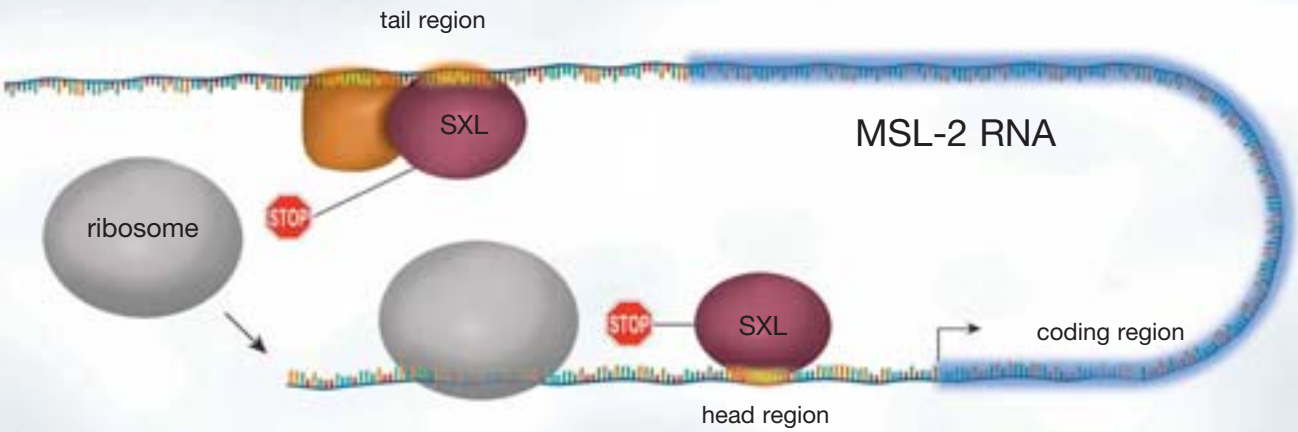
Creating a protein involves multiple steps, including transcribing information from DNA into a messenger RNA, and then using this molecule as a template for creating proteins in the process called translation. The cell can stop the synthesis of a molecule anywhere along the way. Until about a decade ago, most of the control was thought to happen during or just after transcription in the nucleus of a cell. A few cases were known where translation was blocked, but these were regarded as exceptions. It costs the cell energy to produce an RNA – why go that far to make a molecule that never gets used?

Matthias and his colleagues have helped demonstrate that stopping translation is also common and has important functions. For example, a pool of RNAs put "on hold" gives the cell a quick way to respond to environmental changes, or to infections. Proteins can be called up from the pool simply by removing whatever is blocking translation. Without such a system, the emergency response is more complicated: it takes more time to activate new genes and make RNAs from scratch.

MSL-2 is deadly to females, and it gets blocked before its translation into protein. Both sexes produce the RNA template, but in females this molecule goes no farther. This prevents the dosage compensation complex from being built. So understanding how cells cope with sex means understanding what happens to MSL-2 RNA. Matthias and his colleagues knew that a protein called *SXL* docked onto this particular RNA and was somehow responsible for the interruption, but they didn't understand how.

An important clue came with the discovery that SXL could bind at more than one position on the RNA. Karsten Beckmann, a PhD student in the lab, started a project to investigate what this meant. He draws a scheme of an RNA molecule to explain what he found.

"An RNA is a string-like molecule; as you move from the head to the tail, there are regions which have different functions," he says. "The part that encodes a protein is here in the middle. Other molecules bind to the regions on either side of it to help control translation. On the head side, a platform of molecules gets built; this serves as a landing area for the translation machine called the *ribosome*."



Before translation begins, the RNA is drawn into a loop, putting the head and tail regions side-by-side. If the platform at the head isn't built properly, for example, the ribosome never starts to work. If it is able to dock, but something obstructs its path as it scans the RNA, it will fall off.

The circular form means that even molecules bound to the tail of the RNA can sometimes interfere with the beginning of translation. Yet Karsten knew that SXL binds at both ends of the RNA in female cells, which seemed like overkill. Was this necessary?

In a series of experiments, Karsten showed that if SXL is bound only to the head, translation is sharply reduced, but some MSL-2 protein slips through. "Even relatively small amounts of the molecule can be deadly in females," he says. "The only way to stop the protein entirely is to have SXL bound at both places – so what's going on at the tail is crucial. One hypothesis we had was that the loop shape might bring the copy of SXL at the tail close to the protein at the head – were they joining up, somehow, as a roadblock to the ribosome?"

Answering this question required a close-up look at how the ribosome moved. Karsten managed this using a technique which is called toeprinting because it allows scientists to track where the ribosome has come into contact with the RNA. The results were a surprise.

"The tracks were quite different in the two situations," he says. "It means that the ribosome responds differently to a copy of SXL at the head than one at the tail. The first SXL usually manages to stop translation, but if it doesn't, the second will definitely succeed."

What's unusual about this case, Matthias says, is not so much that multiple copies of a molecule are working together to block a process, but that they do so in quite different ways. "SXL has become refined through evolution to serve as its own backup device. That's a subtle solution to a problem where small errors can be fatal."

The same thing could obviously be accomplished with two different proteins, he says; a second molecule could be enlisted to serve as backup. But that would pose a problem familiar to anyone who works with machines: two parts represent two opportunities for things to break down. Doing both tasks with one SXL is an all-or-nothing solution. If it fails, the female fly dies at an early embryonic stage. If it works, the organism is set for life. It's such a logical solution that Matthias and Karsten are sure nature has hit upon it more than once. They're now looking for other examples where a single protein does double-duty.

Such a case has already been found – one which has nothing to do with dosage compensation. The discovery was made by another EMBL group (see page 60).

* * *

As Karsten picked apart the functions of SXL, another member of Matthias' lab was achieving another breakthrough regarding control over MSL-2. Postdoc Kent Duncan was about to discover that the copy of the molecule at the tail doesn't accomplish its job alone; it needs an assistant. Matthias and his colleagues, EMBL alumnae Fatima Gebauer and Marica Grskovic, had already suspected this: three years ago, they discovered something interesting about the tail region of the RNA. While this area of MSL-2 was acting as a landing pad for SXL, it was also doing something else.

"Fatima and Marica did experiments that made changes in the code of SXL's landing site on the tail of the RNA," Matthias says. "The changes weren't stopping SXL from binding. Even so, somehow they made it impossible for the protein to carry out its backup function. A likely scenario was that something else was able to bind and help out. Changes in the code prevented that from happening."

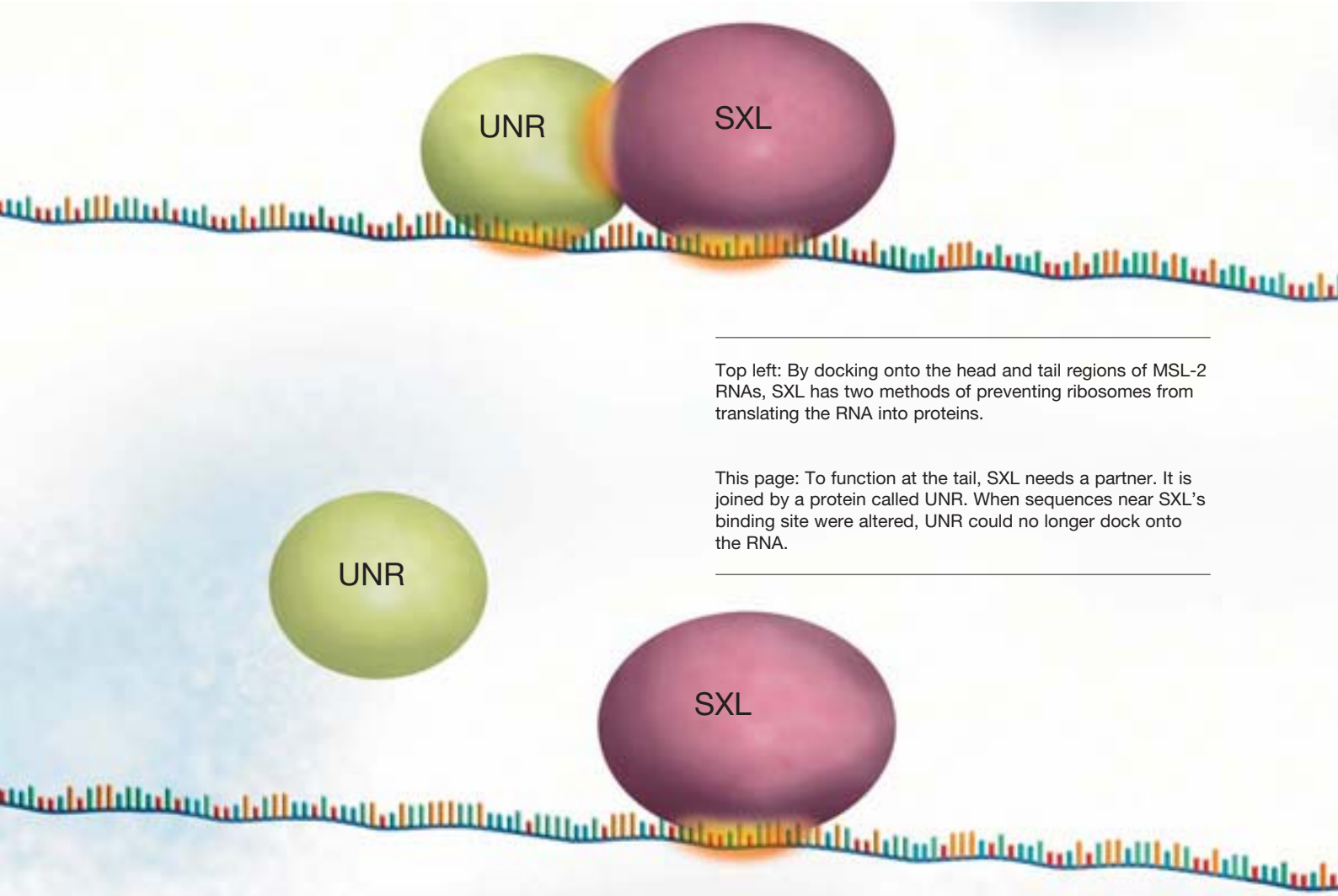
Kent and his colleagues tried a new method (developed in collaboration with Iain Mattaj's neighboring lab) to look for SXL's elusive partner. They used genetic engineering techniques to create a molecular "bait", and then went fishing. The bait consisted of a small piece of MSL-2 RNA containing the area where SXL binds and additional regions where the second molecule was suspected to bind. Technician Claudia Strein dropped it into a test tube containing extracts from fly cells, hoping to catch SXL and its mystery partner. Then the team extracted the bait and analyzed their catch with the help of Matthias Wilm's group.

As they had hoped, SXL's helper was identified, a molecule called *UNR*. "The identity of the partner was a surprise," Kent says. "In flies, nothing was known about the function of this protein. But a very close relative in mammal cells has an opposite effect: instead of helping switch off translation, it actually enhances the production of proteins when it's bound to an RNA."

UNR was bound to SXL and the msl-2 RNA in the test tube, but did it also function to control dosage compensation in the fly? The hypothesis would be difficult to test in an organism, because deleting UNR protein was fatal – to both males and females. "Another possibility was to study its function in cell cultures," Kent says. "But for that we needed a 'female' cell line – one that made SXL and used it to shut down msl-2."

The fact that no one had found this before didn't mean that it didn't exist, so Kent and his colleagues went searching for just such a cell type. They found one in a cell line that has been used in labs for other purposes – a stroke of luck, Kent says. The scientists then eliminated UNR from the cultures and watched as the cells lost control of the dosage compensation machinery. "We're now using this system in collaboration with Asifa Akhtar's group to study other aspects of sex-specific biology in cell culture," Kent says.

That UNR should behave completely differently in flies and mammals is unusual, but Kent believes this has to do with the special nature of its partnership with SXL. It may also be due to the context of events: in building a machine, parts have to be installed in the right order. If SXL arrives first on an mRNA and UNR is mounted later, the effect may be the opposite of installing UNR on the RNA alone, or in combination with other molecules. ■



Top left: By docking onto the head and tail regions of MSL-2 RNAs, SXL has two methods of preventing ribosomes from translating the RNA into proteins.

This page: To function at the tail, SXL needs a partner. It is joined by a protein called UNR. When sequences near SXL's binding site were altered, UNR could no longer dock onto the RNA.

Matthias Hentze, Anne Ephrussi
and Marina Chekulaeva

Building on molecular foundations

IN MARCH 2006, during a renovation of the Church of the Holy Ghost in Heidelberg, workers uncovered the bones of victims of the Black Plague that swept through Europe in the fourteenth century. More exciting to archeologists was the discovery of the older foundations of a Romanesque church under the floor. While it is well known that great Gothic cathedrals were frequently built on the sites of earlier churches, no one had suspected that to be the case here. There was no historical record of an earlier church or the architect who designed it.

Early structures often influence the orientations of buildings that are later built on top of them. There is a biological parallel: the body of an embryo is built on foundations established in the earliest stages of its life. Much of the architecture of the fruit fly *Drosophila*, for example, can be traced back to events that occur when it is a single unfertilized egg (called an *oocyte*). Once fertilized, this huge cell will subdivide into smaller cells; the whole organism will remain about the same size until several hundred cells have been made. But long before that happens, some regions of the single oocyte are marked with specific molecules. This stakes out territories that will define the fly's head-to-tail axis (anterior to posterior).

One of the main architects of this basic building plan is a molecule called Oskar. This protein should only be produced at one pole of the oocyte, an area which will become the posterior end of the fly. If it is produced elsewhere, orientation will be lost, and the developing embryo will experience fatal anatomical deformities.

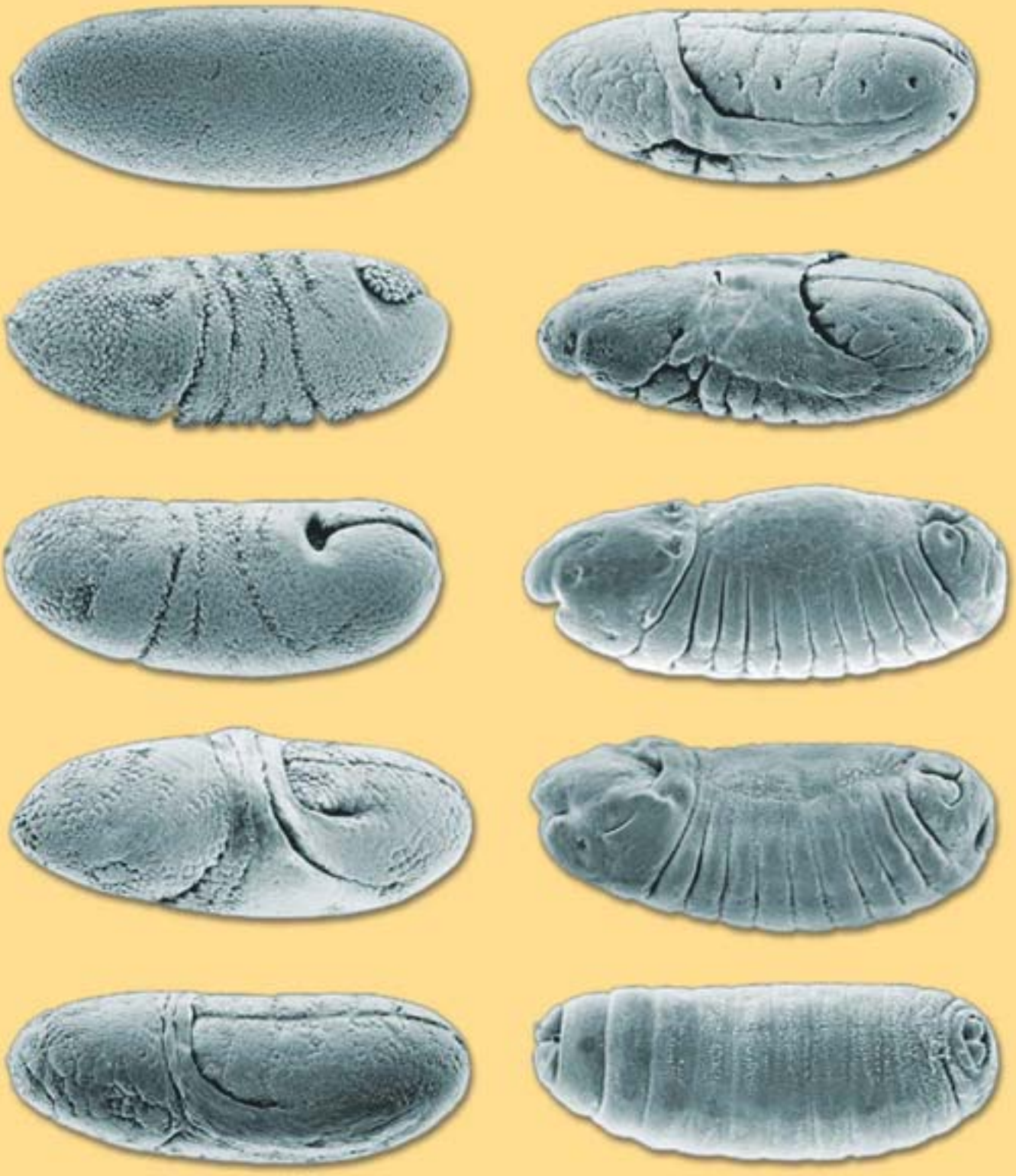
Why doesn't Oskar appear at inappropriate places? It's a question Anne Ephrussi has been working on for many years. Control of the production of proteins often happens just before they would normally be produced – during the handling of an RNA (previous story). That's also true of Oskar. *oskar* mRNA isn't translated until it has arrived at a precise location in the cell, and the fate of the fly depends on this.

* * *

The two end regions of a single RNA cooperate in determining its destiny and in this case that of an entire organism. Molecules that bind to these regions help attract the ribosome and determine whether it can read the coding information sandwiched in between and create a protein.

Now Anne's group, in collaboration with Matthias Hentze, has found that this also happens to *oskar* mRNA. "About ten years ago, it was shown that a protein called Bruno somehow blocks translation by binding to the tail region of the mRNA," Anne says. "More recently a second molecule was found, Cup, which is also required to prevent the translation of Oskar. Experiments suggested that Cup might bind to Bruno at the tail. We already knew that Cup docks onto a second protein attached at the head region of the RNA."

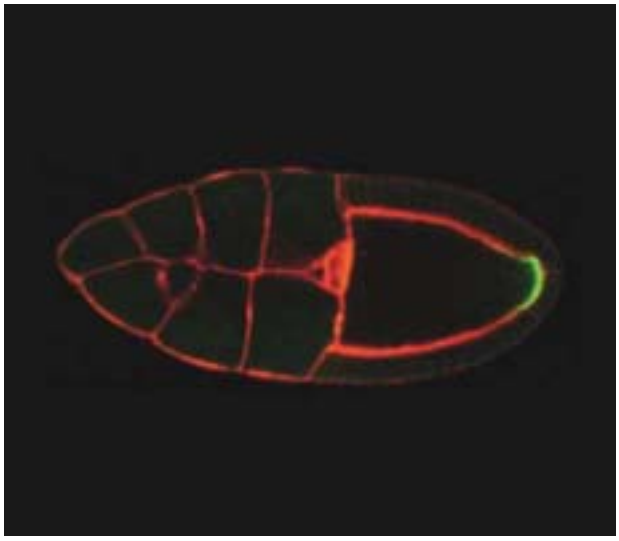
Cup might be an element in a bridge between the head and tail, which are drawn together as translation begins. Such bridges can influence translation, either by helping



the ribosome or by hindering it. In this case, the combination of Bruno and Cup was somehow obstructing the process.

Marina Chekulaeva, a PhD student in Anne's lab, wanted to know how. There was already a hypothesis that involved the way ribosomes are attracted and attached to mRNAs. "This requires a step-wise assembly of other proteins starting at the head end of the mRNA," Marina says. "Cup binds to a protein called eIF4E that acts at one of the early steps in translation. We believed that in doing so, it interfered with the completion of the next step, which is to link to part of the ribosome." Since Cup stopped translation only when Bruno was also around, the scientists assumed that the two proteins and eIF4E were creating a *ménage à trois* that caused the blockade.

Marina's first task was to figure out whether this model was correct. She used a method that was developed in Anne's group several years ago by Stefania Castagnetti. The system is based on extracts taken from fly oocytes, collected in the test tube. This mixture contains all the molecules needed to translate mRNAs, and by adding special *reporter* mRNAs, the scientists can observe when translation takes place. This provides an easy way to test hypotheses. For example, if *oskar* mRNA is put into this system, it won't be translated into protein, because Bruno is around and blocks the process. But if scientists change the code of the RNA so that Bruno won't bind to it, the molecule is translated. This is good evidence that Bruno is directly responsible for the interruption.



Molecules can be added to the extracts that prevent others from doing their jobs. For example, eliminating Bruno binding leads once again to translation of Oskar. Or drugs can be introduced that arrest translation at precise stages. "Previously we had all-or-nothing answers – either translation occurred or it didn't," Anne says. "This method tells us exactly at what step it stops. That's how Marina was able to see that the combination of Bruno and Cup did exactly what had been predicted: it brought things to a halt just before the ribosome attaches itself to the RNA."

Can Bruno stop translation without the help of Cup? Marina did another experiment with extracts containing a version of Cup that doesn't function. She was surprised to see that even working alone, Bruno could do the job.

"This meant that Bruno has two different methods of blocking translation," Marina says. "In the experiments with Bruno but a mutant form of Cup, we noticed that *oskar* mRNA was clumping up in huge assemblies of RNAs and proteins. We extracted these to find out what they were made of. It turned out that Bruno – probably with the help of other molecules – was acting as a sort of magnet for *oskar* RNAs, pulling them together in these huge knots."

The clumps also contain other proteins: more copies of Bruno, along with Cup (if it's available), and a third molecule called Me31B. Other experiments have shown that this protein can also block the translation of *oskar*.

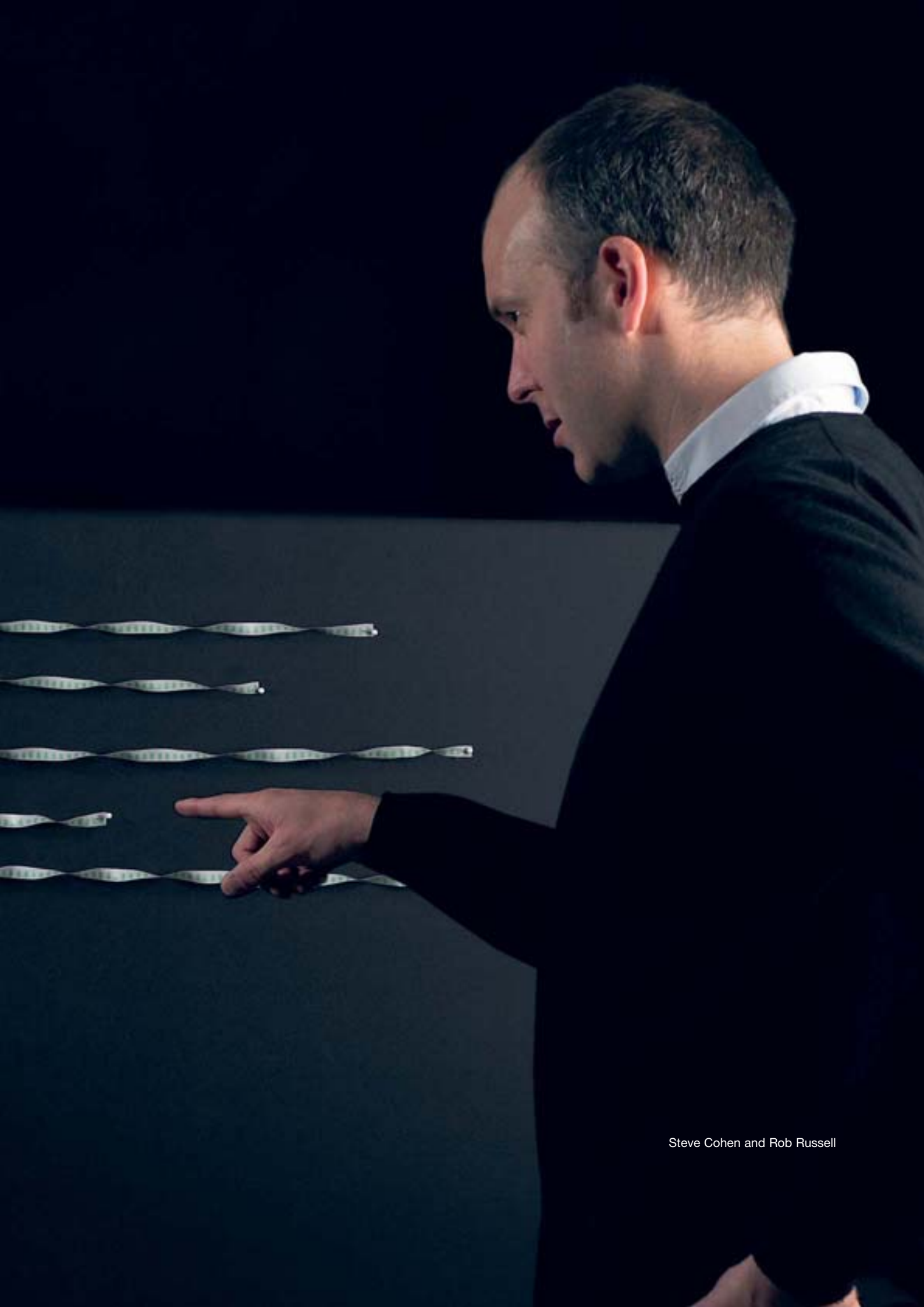
The study revealed one more interesting fact: "Why Oskar doesn't get translated until it arrives at the right place is one thing we needed to understand," Anne says. "Another is why the restrictions are removed when it arrives there. Our work with Cup shows how it might link the two processes. Cup forms a roadblock to eIF4E, which is a part of the platform built to get the ribosome placed on the RNA. But it isn't sending eIF4E away; in fact, it may attract eIF4E to the RNA in the first place. This means that once the RNA reaches the posterior pole of the cell, there will be a good supply on hand of one of the critical components necessary to get translation started."

Bruno and SXL (see previous story) each have two ways of blocking the translation of proteins from RNAs. It's interesting – but no accident – that within a few months of each other, two EMBL groups have come across different examples of this new principle. Scientists working on many different themes are converging on RNAs as a level at which cells regulate important processes in rich and subtle ways. ■

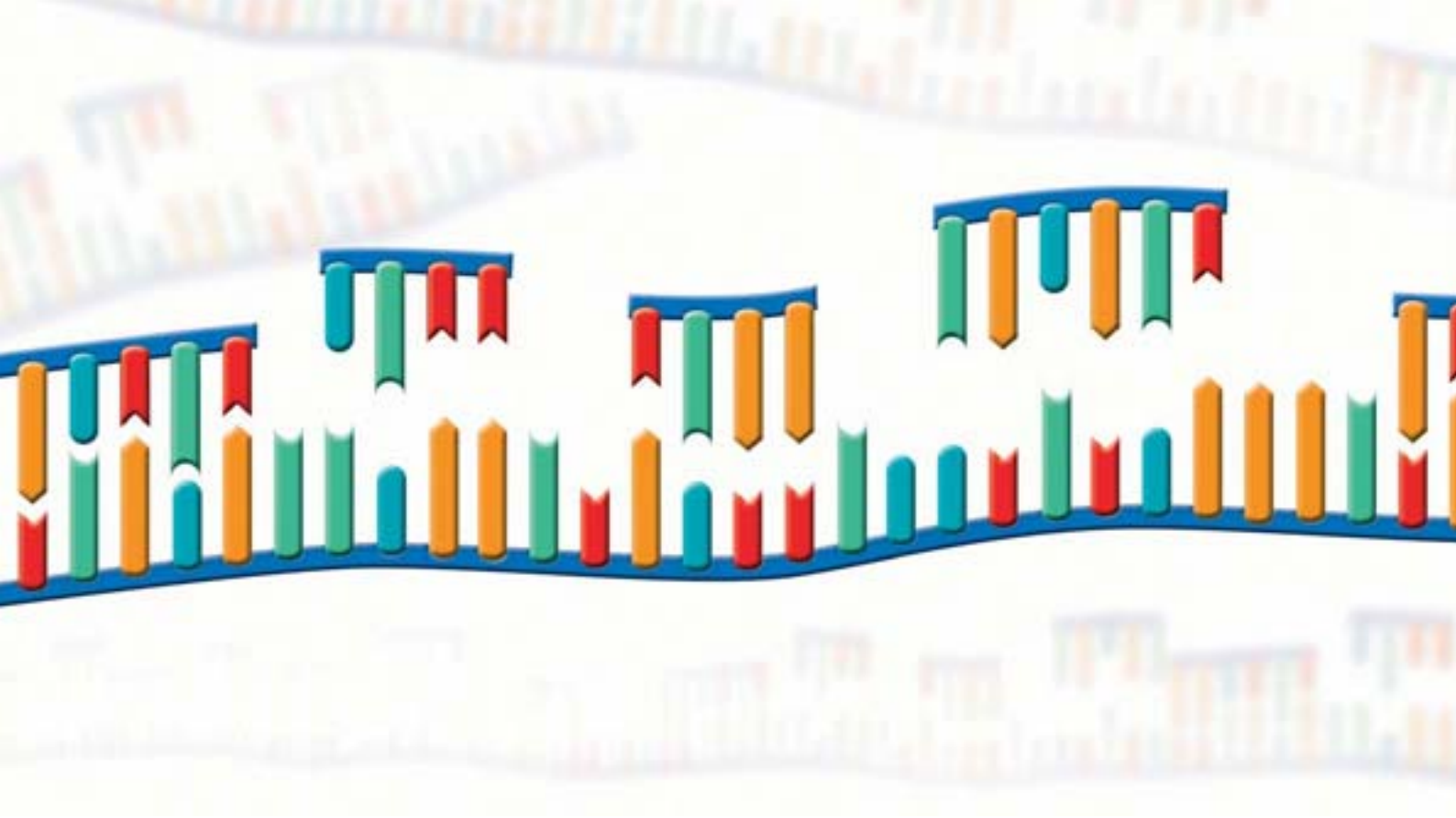
Left page: The development of the fly embryo over time. Each stage is built on patterning established in earlier stages.

This page: oskar mRNA must be moved to a particular position at one pole of the oocyte before it is translated into protein (green); otherwise, patterning will be disrupted.

Border guards and the evolution of tails



Steve Cohen and Rob Russell



THE THRESHER SHARK has a spectacular tail that often measures longer than the rest of its body. This enables the thresher to swipe at small fish and stun them, making them easier to eat. For millions of years, natural selection has been elongating the tail; length brings more meals, and better chances for survival and reproduction.

Natural selection has produced both long and short tails in animals, and Steve Cohen and Rob Russell are looking at ways that it has influenced the tails of RNAs. The previous stories in this report show how the region acts as a landing platform for other molecules that influence whether an RNA is used to synthesize new proteins. This type of control has evolved many biological functions, for example to balance genetic differences between the sexes and to help cells respond to environmental changes.

Additionally, the tail serves as a binding site for tiny molecules called *microRNAs* which also influence protein production. One effect of microRNAs is to recruit other molecules that block the translation machinery and cause the RNA to be degraded. A decade ago, these small molecules were considered an oddity; in the meantime, they have been found throughout animal genomes. The collaboration between the labs of Steve and Rob has steadily yielded insights into their functions; now the researchers have identified some general principles about cells' use of microRNAs. Along the way, the researchers have discovered something about their importance in the evolution of complex organisms.

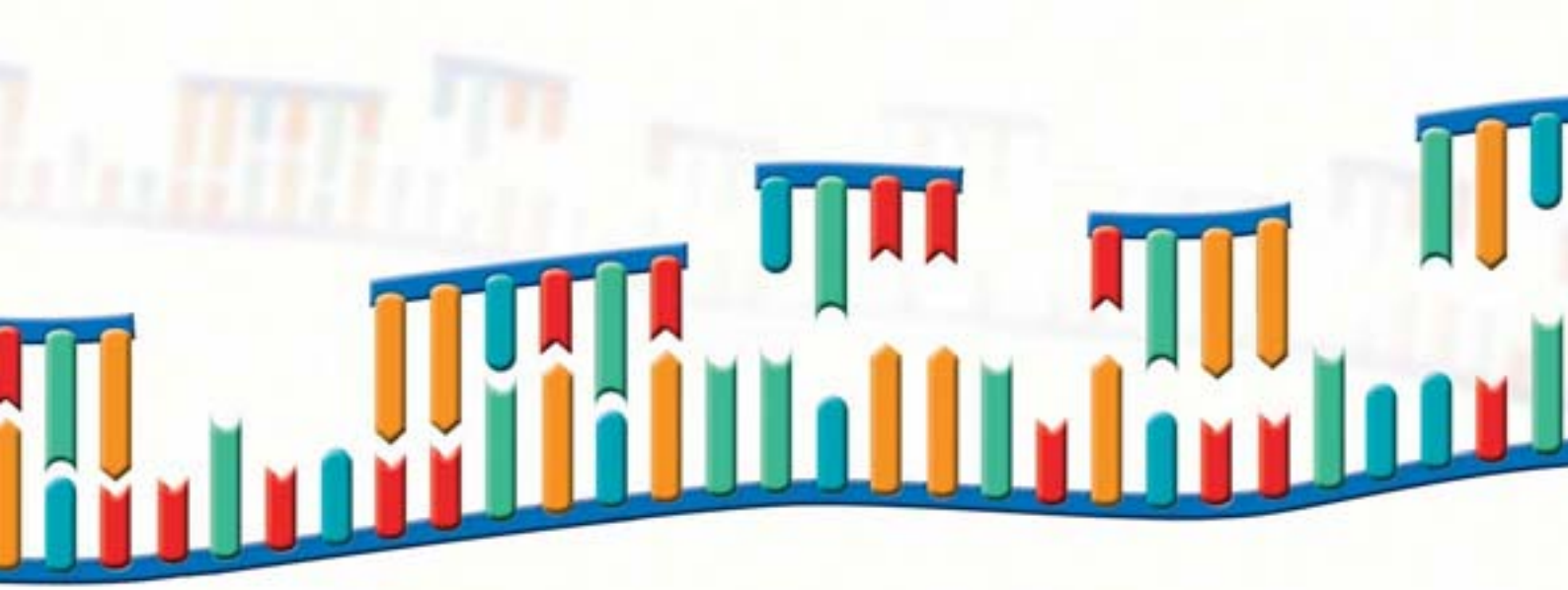
MicroRNAs do in fact bind to other RNA molecules, usually in the tail region, and this could partially or completely block translation.

* * *

In 1993, scientists discovered the first microRNAs in the genome of the worm *Caenorhabditis elegans*. These molecules were obviously not destined to become proteins – surprising because at the time, RNAs were still largely seen as means to an end, messengers rather than performers, the carriers of information needed to create proteins. Once an RNA was created, it was assumed that a protein would normally follow.

But RNAs were gaining recognition as an important step at which cells could control processes. Proteins could block translation; microRNAs might do so, too. Their sequences were often complementary to codes found in other RNAs. This suggested that a microRNA might seek out the complementary molecule and bind to it, the way that complementary nucleotides in DNA bind to form a double helix.

Researchers quickly found that microRNAs did in fact bind to other RNA molecules, usually in the tail region, and this could partially or completely block translation. The discovery was developed into an important new tech-



Complementary sequences permit microRNAs to bind to the tails of other RNA molecules, often with important regulatory effects. One tail may have many targets for several different microRNAs.

nology: scientists could shut down a particular RNA by designing a small, complementary molecule and inserting it into cells. Switching off genes and watching what happens to an organism has long been a key tool to understand their functions; now this method, called *RNA interference*, has become one of the most commonly-used tools to do so.

Getting a global idea of how microRNAs function has been difficult; their size makes them hard to spot against the background of billions of letters of code in animals' genomes, and it is equally difficult to find the complementary sequences in the longer RNAs that they target. Another difficulty in finding them is that they mutate quickly. Although many of the molecules have been conserved throughout evolution, each organism has invented a repertoire of its own; they are easy to make – often a change in a single nucleotide in the genetic code will create a new microRNA or a target. This means that it is hard to find them the way many genes have been discovered – by comparing molecules from different species.

Rob's group and other teams have circumvented some of these problems by developing new search algorithms. As a result, they estimate that microRNAs make up between one to five percent of animal genes. And while it was originally thought that most of these molecules acted on a single target, Alex Stark from Rob's group and Julius Brennecke from Steve's lab teamed up to show that the average microRNA has the potential to dock onto hundreds of RNAs. They also discovered that a single tail may contain targets for several different microRNAs. One particular molecule – the "thresher shark" among RNAs – has fifteen sites where ten different types of microRNAs can dock. In total, about 20-30 percent of animal RNAs seem to contain targets.

They estimate that microRNAs make up between one and five per cent of animal genes, and the average microRNA has the potential to dock onto hundreds of RNAs.

These surprisingly large numbers, and the fact that specific microRNAs and targets have been conserved over long stretches of evolution, suggest that they have an important function. How often do cells use them to block the production of proteins? Steve's lab had already shown that they help manage crucial aspects of cells like division and self-destruct programs called *apoptosis*. Now Julius and Alex began a wider analysis in hopes of uncovering their functions in the development of organs and tissues.

One discovery they made was a correlation between the length of an RNA and the number of docking sites for microRNAs it contained. They had expected to find that long tails typically had more binding sites than short tails, but why should the density of docking sites also be higher

Above: Expression patterns of microRNA-124 (blue) and one of its targets, sc (brown) at three stages of fly development: stage 9 (top, dorsal view), stages 11 and 14 (ventral view)

Right: microRNAs may have important functions in establishing borders in devleoping tissues. microRNAs are frequently expressed abundantly on one side of a border, and their targets abundantly on the other. These borders may be spatial – as shown on the right – or temporal – the same tissue may express a microRNA at some phases of development and its target at others.

on long tails? Evolution seemed to have packed as many targets as possible into long tails, while reducing or totally eliminating them in shorter tails.

There was also a clear difference in the jobs carried out by the two types of molecules. The short-tailed variety of RNA, with few targets, was being used in all types of cells, by proteins with crucial functions that were needed everywhere. RNAs with longer tails usually appeared only in specialized cell types or tissues.

Were microRNAs doing something specific to build specialized tissues and organs? Or had evolution actively eliminated microRNAs in crucial proteins? The scientists decided to find out by looking at experiments in which organisms' microRNAs had been removed. If they performed important control steps in the creation of specific tissues, this ought to lead to major defects in embryonic development. But this wasn't the case; the absence of microRNAs didn't seem to create serious problems.

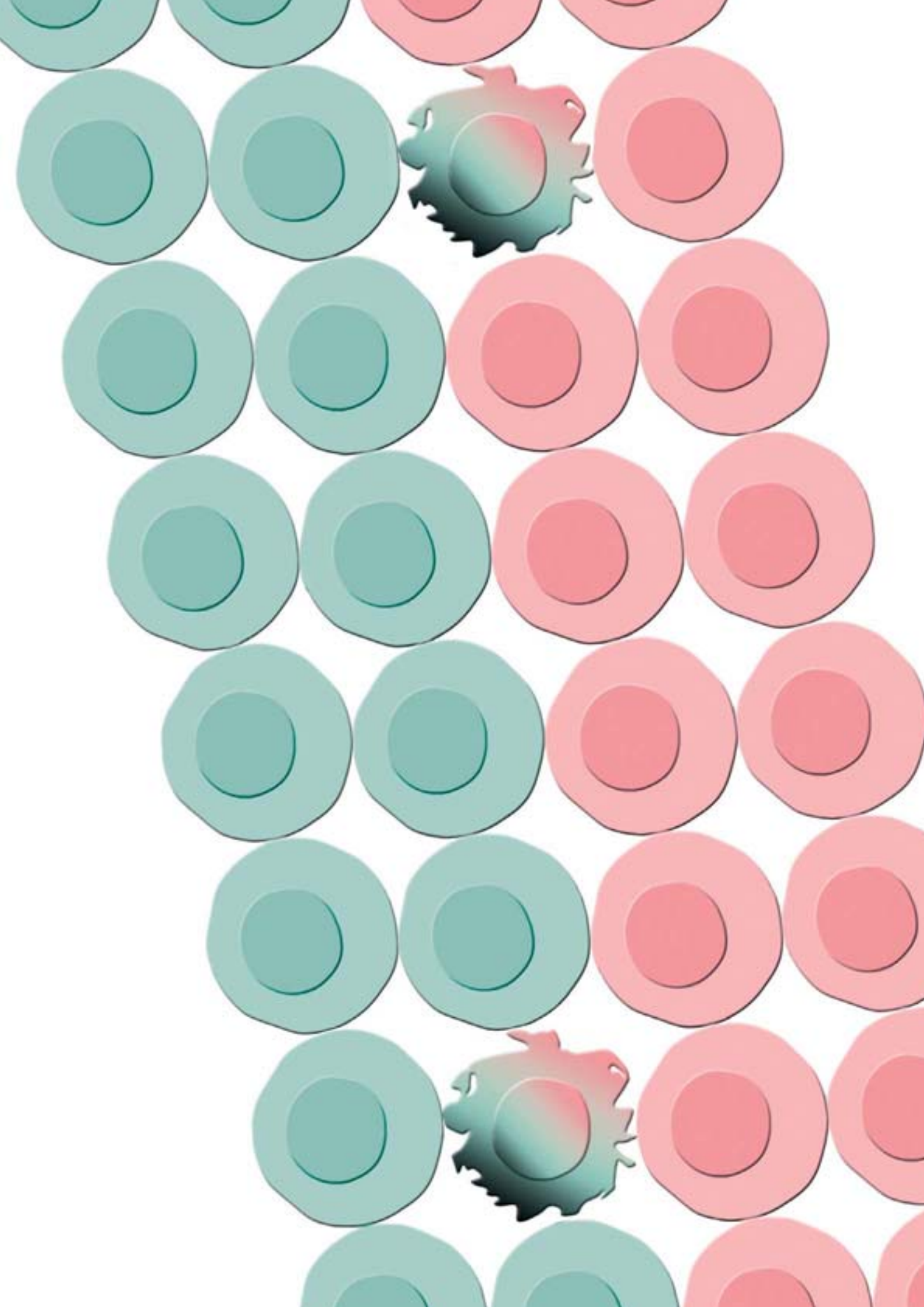
Julius and his colleagues thought of a solution to this apparent contradiction. To have any effect, both microRNAs and their targets have to be produced in the same cell. Studying developing tissues showed that this doesn't seem to happen very often. MicroRNAs and their targets are rarely abundant in the *same* cell. However, they are often abundantly expressed in *neighboring* cells.

This could have an indirect effect in shaping organisms. The cells of the early embryo are similar. As an embryo develops, molecular signals pass between its cells, instructing them to form compartments. This permits the creation of different cell types and eventually sophisticated organs. There have to be clear boundaries where cells on one side behave differently than those on the other – otherwise an organism would be little more than a mass of a single type of undifferentiated tissue.

MicroRNAs might be helping to form borders. If one cell produces a microRNA but not the target, and its neighbor does the opposite, a borderline may form. If either cell makes a mistake and produces both types of RNA, the microRNA will block the production of the protein. If these proteins happen to be important players in establishing cell identities – for example, if they are transcription factors that activate developmental genes – this type of control could help enforce boundaries and compartmentalization. This hypothesis is supported by the fact that transcription factors frequently contain multiple targets for microRNAs.

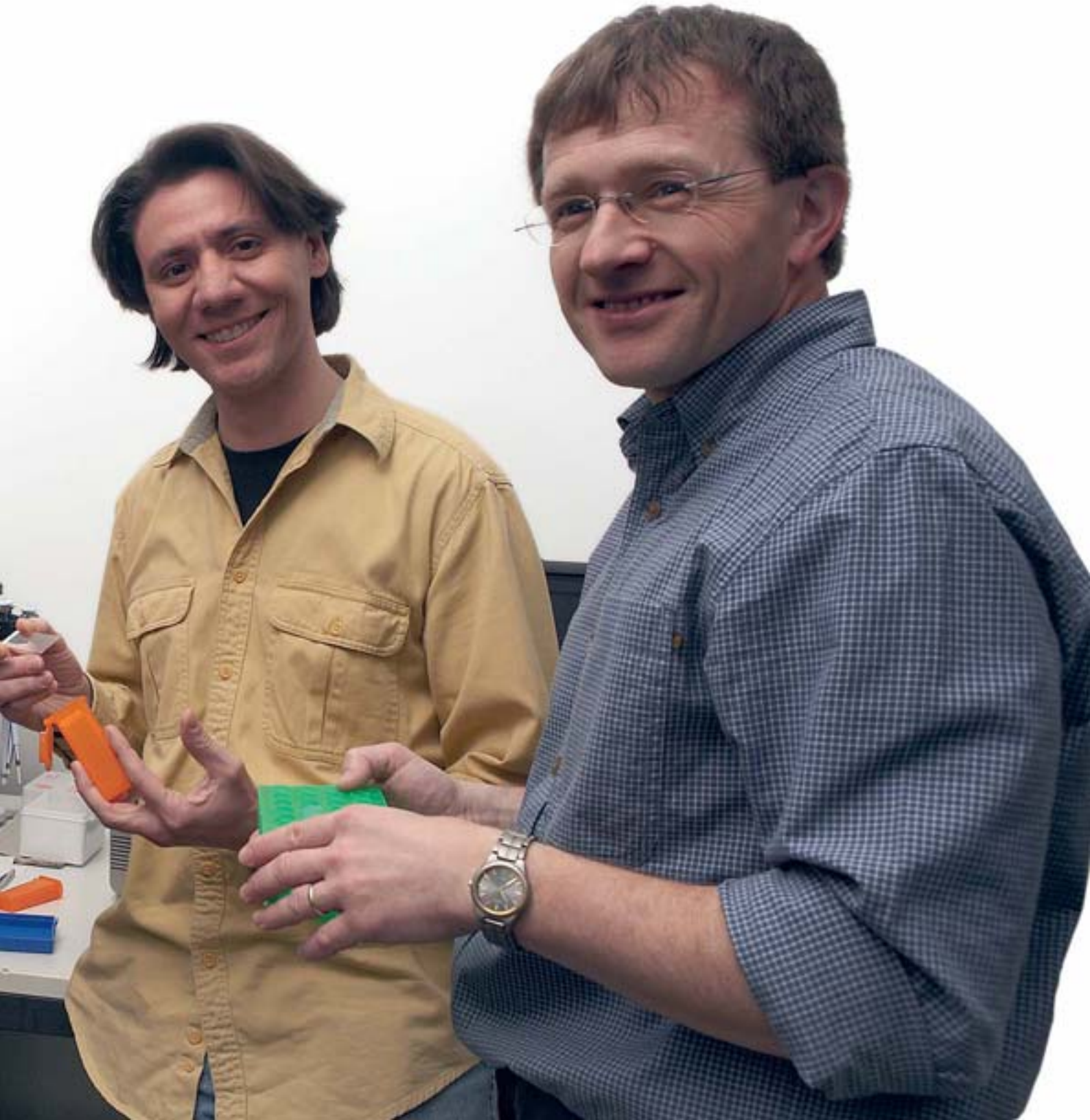
Evolution has produced both long and short tails. "What we've found suggests that microRNAs seem to be acting as a sort of backup system to protect the identity of tissues, a border patrol that catches trespassers," Rob says. "They block molecules that would lead to mistakes in development. RNA targets with long tails – very effective at blocking protein synthesis – are found right across from microRNAs."

Evolution also explains why proteins targeted by microRNAs usually haven't acquired critical cellular functions, he says. Natural selection usually favors the robust. Both microRNAs and their targets are so small that a single mutation is likely to eliminate them, and this probably happens frequently. It's best not to base the fate of an organism on a system that can be so easily perturbed. But as one system of control among many, a way of locking cells into their tissue identity, microRNA regulation has served animals over very long stretches of evolution. ■



Matthias Hentze, Martina Muckenthaler,
Mirco Castoldi and Vladimír Benes

A cellular census



A census can tell you all kinds of fascinating things about a country – for example, that the rate of illiteracy in Europe in 2000 was lowest in the world – only 1.3% among people at least 15 years old. Thirty years earlier it had been 6.9%, also lowest in the world. Or that Iceland had 4,015 researchers per million population in 1995, compared with 3,054 in Denmark and 2,831 in Germany. Information from a recent European census also allows you to find and compare the average salary of a scientist (if you can translate between national currencies).

A census of the cell can also reveal fascinating things. Several years ago, scientists invented the DNA chip, a method to interrogate populations of molecules in cells, to discover differences in the way genes are used in different types of cells. It has permitted scientists like Vladimír Benes, of the GeneCore Facility in Heidelberg, to create probes that can conduct a complete survey of the contents of a cell.

Normal DNA chips aren't well-suited to study microRNAs. It is difficult to find common experimental conditions under which many different microRNAs will bind equally well.

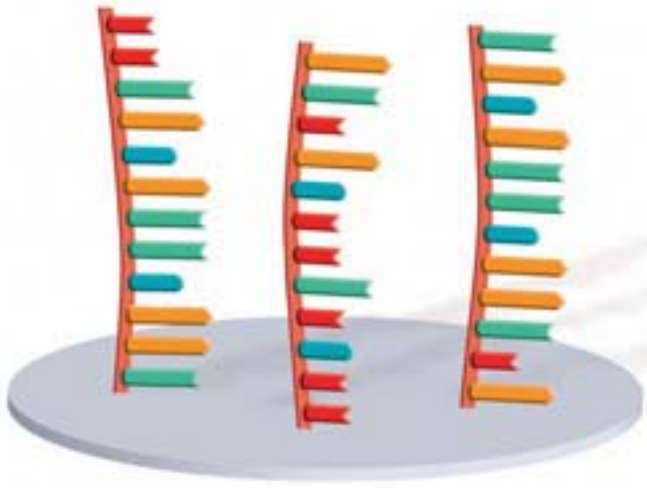
The technology is based on the principle that RNAs and DNAs which have complementary sequences bind to each other. Samples of DNA representing all of a cell's genes are used as probes. They are printed onto glass slides, and are then exposed to RNAs extracted from cells. An RNA will bind to DNA from genes with a complementary sequence. Because the RNAs have been tagged with fluorescent markers in advance, a laser can be used to illuminate the slide and detect probes which have made a catch. This lights up the genes that are active in a given cell, and by comparing different types of cells (for example, healthy and infected ones), scientists can discover differences in the activity of their genes.

This has become a powerful, commonly used technique that works well for "typical" RNAs. Martina Muckenthaler, a former member of Matthias Hentze's group and now a professor at the University of Heidelberg, developed DNA chips with the help of the EMBL GeneCore Facility to study diseases related to the body's uptake of iron. Now she and Mirco Castoldi, a postdoc in Matthias' group, have adapted the method to work with microRNAs. The project was carried out in the Molecular Medicine Partnership Unit, a joint research unit of EMBL and the University of Heidelberg, headed by Matthias and Andreas Kulozik of the university.

Making a microRNA chip was important, Mirco says, because these small molecules are now recognized as major players in how the cell regulates the activity of its genes. But for technical reasons, normal DNA chips aren't well-suited to study microRNAs. The binding of short molecules is finicky and highly sensitive to temperature. It is difficult to find common experimental conditions under which many different microRNAs will bind equally well.

The solution that the scientists came up with was to switch the DNA probes for similar molecules called LNA, which stands for *locked nucleic acids*. The difference between the two is chemistry. "A DNA molecule is built of nucleic acid bases connected to each other, like the rungs of a ladder," says Vladimír. "Those bases are linked

Normal DNA chips (below) use DNA with its normal "backbone" of phosphate sugars (red), but these probes are poor at capturing microRNAs. Changing the sugars (blue, right page) creates a new type of molecule called LNA which makes probes that bind strongly to microRNAs.



vertically by phosphate sugars – which act like handrails. LNA has the same rungs, but another type of sugar that leads to very strong binding between the RNA and the LNA probes." The LNA used on the chip, he says, was provided by the Danish company Exiqon.

This gives researchers the technology to carry out the same sort of census of microRNAs that has already been possible for longer RNAs, and the GeneCore facility is prepared to make chips for other EMBL groups. They expect wide interest because of the growing recognition of the importance of microRNAs in regulating cellular processes; they have also increasingly been linked to disease.

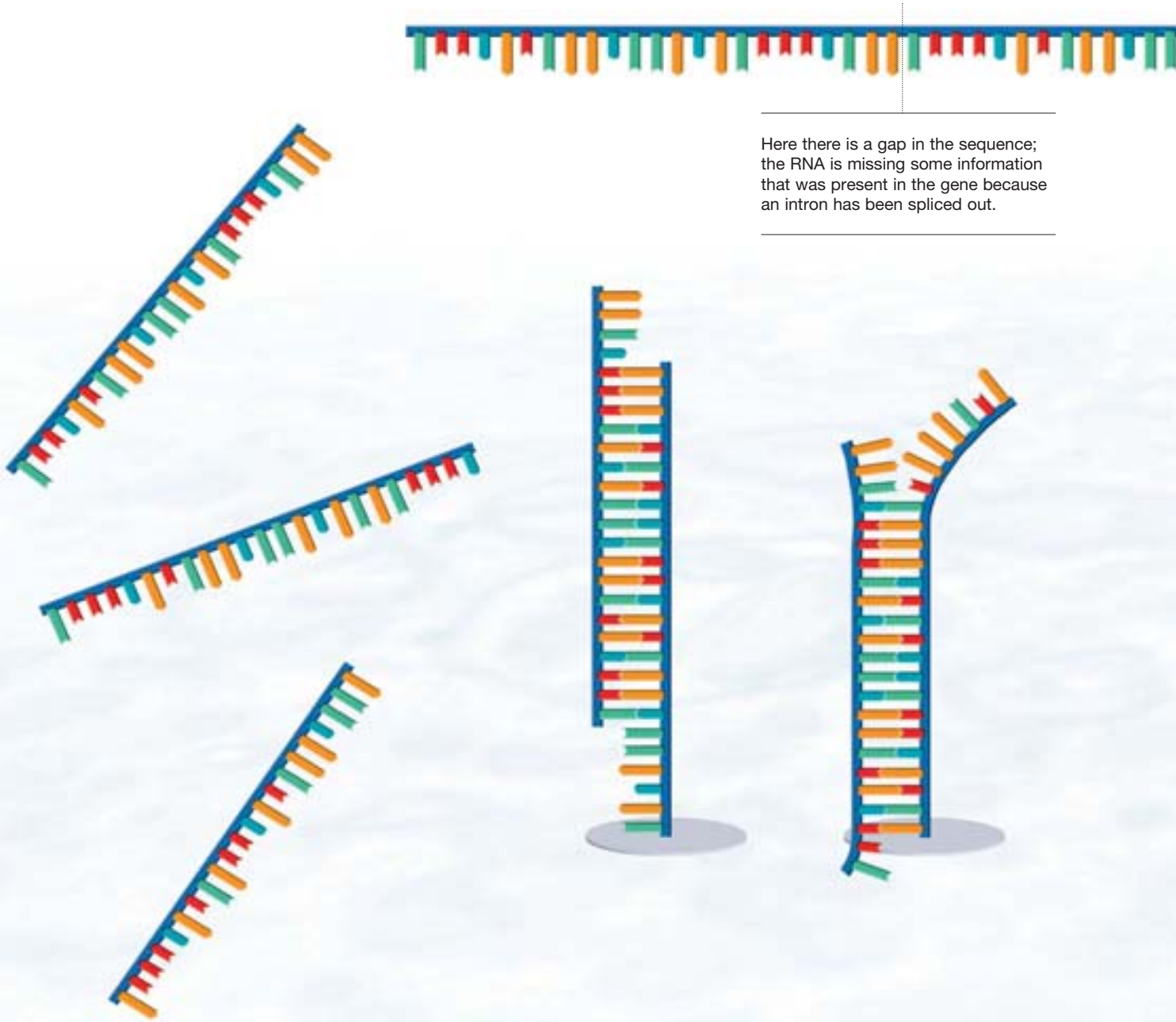
DNA chips have already helped scientists discover new molecular markers for cancer and other diseases, and we'll likely find equally important effects of microRNAs.

"The platform and the experimental protocols are very easy to handle and can be used in clinical settings," Martina says. "DNA chips have already led to the discovery of new molecular markers for cancer and other diseases, and we will probably find equally important effects of microRNAs. We already know, for example, that hepatitis C requires the cooperation of a cellular microRNA to replicate in human cells. That could suggest new possibilities for therapies."

The method is also attractive to medical researchers because it requires far lower amounts of samples than some other types of microRNA chips, which is an issue in samples taken from patients. Mirco, Martina and their colleagues have already used the technique to compare microRNA expression in tissues such as the liver and the heart. But this is just the beginning; they have already expanded the census to other types of cells and a variety of genetic diseases. ■

Lars Steinmetz, Marina Granovskaia, Wolfgang Huber and Sandra Clauder-Münster

Reappraising the genome



Here there is a gap in the sequence; the RNA is missing some information that was present in the gene because an intron has been spliced out.

WHAT'S IN A GENOME? Just a few years ago, most people would probably have answered "genes and junk," and there seemed to be an awful lot of junk. Upon completion of the human genome, scientists announced that only about two percent of the complete DNA sequence encoded proteins. Most of the rest appeared to be excess baggage, the leftovers of evolution. Surely some of it had a function – cells were known to produce some RNAs that didn't encode proteins and had regulatory functions. But what was the rest up to?

In a collaborative project with Stanford University, Lars Steinmetz of Heidelberg and Wolfgang Huber of EMBL-EBI have been trying to answer this question. They are using a new method called a *tiling array* to search for new functions in the complete yeast genome. A tiling array is a DNA chip, so the method is similar to DNA chips based on genes or microRNAs (see previous story). All of these methods contain probes of DNA on a glass surface to detect RNA molecules extracted from cells thus a tiling array study also shows what part of the genome is active under various conditions. But each of these DNA chip methods is like asking a series of yes/no questions: you only get an answer if you've posed the right question. Typical gene chips hold only probes for known genes, and the samples on a microRNA chip look for matches to samples preselected by a computational analysis of the

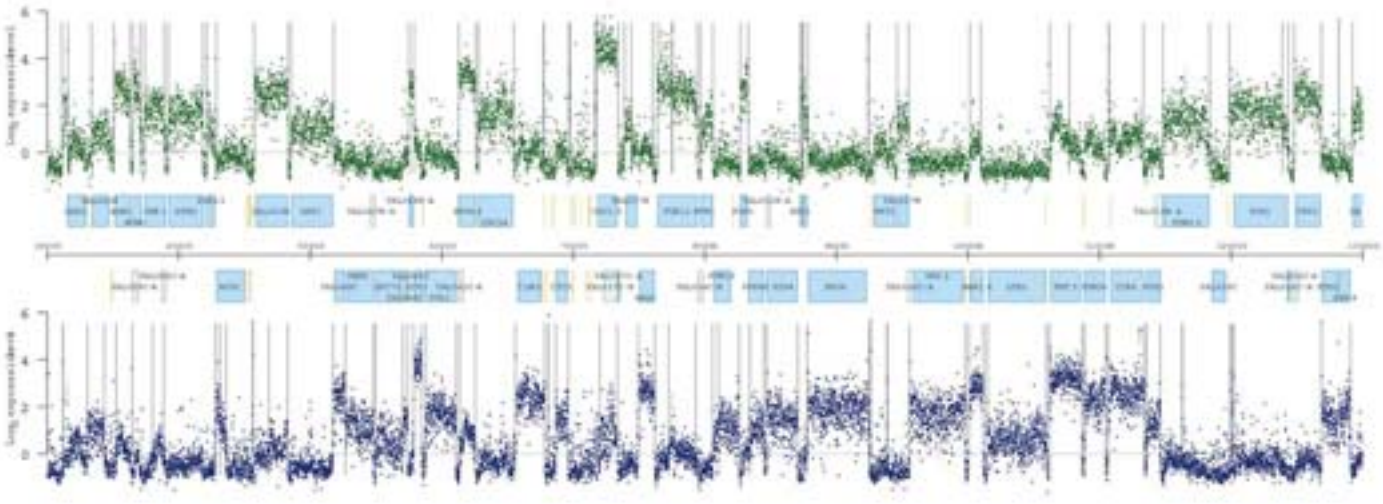
The tiling array used by Lars' group uses 25-nucleotide sequences from the yeast genome as probes. RNAs (represented by the long molecule along the top) are extracted from cells, cut into fragments, and allowed to bind to any probes with complementary sequences. The method reveals not only which RNAs are produced at particular times, but where they begin and end, and also whether sequences are missing – for example, when an intron has been removed.

genome. The result is like conducting a survey only of friends, and supposing that the results apply to everyone. This means that the assay is biased towards the sequences which are selected and placed onto the array.

"This will give us our first full look at the cell's complexity at the level of RNA. At the moment, a tiling array is the best way to find this out."

In contrast, a tiling array creates probes from both strands of the entire genome, even "junk" regions that have no known function. "There are several reasons to generate an array in such an unbiased fashion," Lars says. "One is that we might find new genes. Another is that it will give us our first full look at the cell's complexity at the level of RNA. It can tell us which bases in the genome are transcribed. At the moment, a tiling array is the best way to find this out."

As often happens in the development of new technologies, he says, at first skeptics wondered whether such arrays would yield useful results. Now the power of this technology is apparent. It promises to revolutionize what microarrays can reveal about genomes.



A readout from a tiling array experiment. Blue and green represent different strands of DNA. Dots show where probes recorded "hits" – RNA transcripts produced by the cell.

Tiling arrays have been made before on a smaller scale, for example to investigate the genome of cellular structures called *mitochondria*. Most scientists believe that these structures evolved from independent organisms – probably bacteria – which once took up residence in other types of cells and never left. Mitochondria have their own DNA, a much smaller genome which reproduces independently of the DNA in the nucleus. Recently scientists have begun interrogating larger stretches of DNA, such as whole genomes, but these studies have yielded unclear results because of a lack of precision and problems in interpretation. Lars and his colleagues took on these issues last year when they created a new "high-resolution" array that contains 6.5 million separate probes from the yeast genome.

"It's been hard to get a direct look at the untranslated regions of the RNA at the head and tail. With this study we could determine exactly where an RNA molecule begins and ends."

"The resolution comes from the number of probes making up each array and the overlap between consecutive probes as they map to the genome," Lars says. "It's a bit like trying to read a book by sampling the text. A standard DNA chip based on genes says, we know there is some content on page five, so we start at page five and grab 50 or 60 characters in the middle. Then we skip to page eight and do the same thing. The tiling array starts with the first letter in the book (really the first base of the genome), and captures the first 25 letters. Then we move down eight letters and take another sample of 25 letters – an overlap of 17 letters in the code. And we continue this way all the way to the end of the book. Then we do exactly the same thing with the second strand of DNA."

Each experiment using the array produces hundreds of megabytes of data – a nightmare for interpretation. Here Lars and his colleagues could draw on the expertise of Wolfgang's group at EMBL-EBI, who have been collecting methods needed to analyze microarray experiments in a suite of tools called *BioConductor*. The methods are particularly good at distinguishing meaningful data from noise – particularly important with short probes, in which the specific sequences that make up each probe affect how well RNAs from the sample bind to them. Poor binding leads to ambiguous results and lots of noise.

The analysis was important and complicated, Wolfgang says, because the tiling array shows where RNAs are bound – but not precisely what they are. The same DNA sequence can produce different RNAs, for example, when an RNA is spliced to remove an intron. Thus several forms of a molecule may be bound to the same probe, and it takes clever computational and statistical methods to understand what a "hit" means.

Overall, the study revealed that when yeast grows in a rich source of food, 84.5% of the entire genome is transcribed into RNA. This is substantially more than the protein-encoding part, which accounts for about 75% of the yeast genome. 16% of the bases that are transcribed in the genome had never before been observed or predicted.

The scientists also made some important discoveries about the structure of genes. "In many cases the coding region of the gene was well-known," Lars says. "But it's been harder to get a direct look at the untranslated regions of the RNA at the head and tail. With this study we could determine exactly where an RNA molecule begins and ends."

Wolfgang's analysis showed that some genes had an unexpectedly complex architecture. At times, parts of genes were expressed at different levels, indicating that different lengths of RNA molecules had been created from the same DNA sequence. Other unusual cases included single RNAs that seemed to encompass two neighboring protein-coding regions.

Another discovery was that the average tail region of an mRNA is longer than the head – 91 versus 68 nucleotides. That makes sense, Lars says, because the tail region is often packed with information that helps cells regulate when, where and how often an RNA is translated into protein. The longest tails were usually found in RNAs that encoded proteins which would be used in the mitochondria, the cell membrane or the cell wall. Long untranslated regions usually indicated that the RNAs were somehow being regulated, for example through the attachment of proteins, ther RNAs.

As well as discovering hundreds of new RNAs, and RNAs produced by reading the "second" strand of DNA, the researchers obtained new insights into the functions of these molecules. The length of an RNA's untranslated regions is related to its function and the region of the cell in which it operates. And what happens on the two strands is not independent. If in a particular region, both strands of DNA encode an RNA, their untranslated regions tend to be longer. "Antisense" RNAs made by transcribing the strand opposite another gene often seem to be involved in regulating other RNAs – which is logical, because they have complementary sequences to the second strand and thus the two molecules could bind to each other. This was suggested based on genetic engineering experiments several years ago, but so far the phenomenon hasn't been considered to have a serious role under normal conditions in the cell. Lars says that this study revisits the issue and suggests that antisense transcription, which the scientists have now observed extensively over the genome, could indeed have a regulatory role in yeast cells.

"This opens a new frontier," Lars says. "Yeast was the first completely sequenced eukaryotic organism, and people have had ten years to work on the information encoded in its genome. Even so, there is a vast amount of transcription detected by our study that was not known."

The same is true of other genomes, including humans'. Paul Bertone, who recently joined Nick Luscombe's group at EMBL-EBI, carried out a tiling array study of the entire human genome as a PhD student at Yale. In both studies, probes captured a large number of RNAs that hadn't been known to exist, including antisense RNAs.

The yeast data represents the most accurate transcriptional map of any eukaryotic organism, with far higher resolution. Even so, it's just a beginning.

The yeast data represents the most accurate transcriptional map of any eukaryotic organism, with far higher resolution, Lars says. Even so, it's just a beginning. He believes that there is much more information to be mined from the data that has been obtained. "And comparing these results to similar studies in other organisms, once they can be carried out at a sufficiently high resolution, will give us unique insights into evolution," he says. "Comparing complete genomes has already suggested that a lot of DNA beyond the protein-encoding content of genes may have a function than we have been able to observe. If we find that this information is transcribed in several species, it will give us a handle to start looking for its functions." ■

An RNA recycling centre

THE DNA IN EACH OF OUR CELLS has a life history that stretches back billions of years, to the last common ancestor of all life on earth and farther to its ancestor. RNAs and proteins have much shorter histories – they are created within our cells and usually end there after lifespans lasting from a few minutes to many hours. Then they are usually broken down and recycled.

RNAs get taken apart for various reasons. Being tagged with a microRNA usually leads to a molecule's destruction; sometimes the binding of a protein has the same effect. They are also broken down if they contain faulty information. Cells often recognize RNAs that have been transcribed from genes containing mutations, pulling them out of circulation before they can be used to produce harmful proteins. This system, called *nonsense-mediated decay* (NMD), doesn't always work; in some diseases, for example, defective RNAs escape.

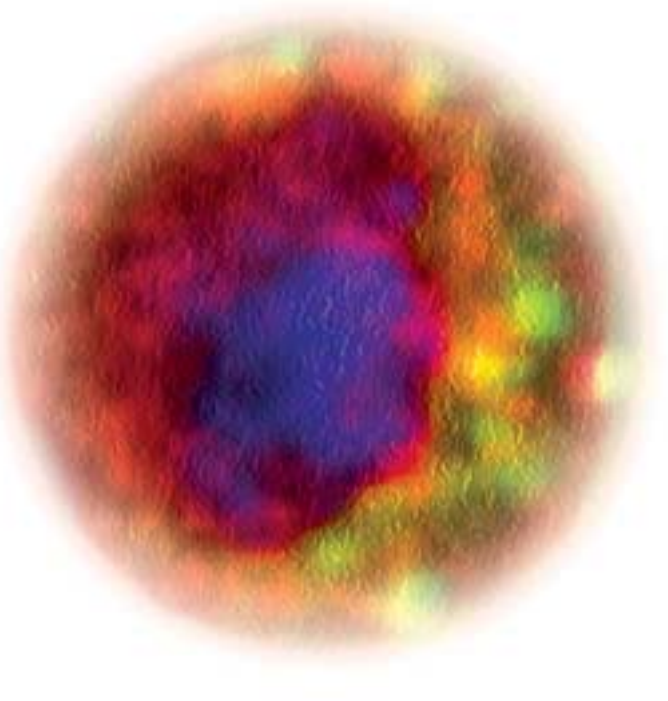
Elisa Izaurralde, Anne Ephrussi and Matthias Hentze have been studying the life histories of RNAs and some of the intricate connections between the ways cells deal with them. For example, they have helped to show that splicing in the cell nucleus helps determine an RNA's fate. Elisa and her colleagues recently found another connection in the way cells handle RNAs that need to be eliminated.

Shutting down NMD, by interfering with the molecules that govern it, didn't affect microRNA regulation. Clearly separate pathways were being used to cope with different types of RNAs.

Most RNAs that get broken down are brought to a structure within the cell called an *mRNA processing body*, or *P-body*. "These regions collect several different molecules that assist in breaking down RNAs into their basic components," Elisa says. Does a common end station mean that different types of RNA disposal share other common features? Elisa's group decided to find out.

P-bodies were discovered a few years ago when scientists found that several proteins involved in mRNA degradation accumulate in small structures in the cell cytoplasm. The RNA-binding protein *GW182* is a marker for these focal points. The protein also helps hold P-bodies together; without it, the structure falls apart. Recently PhD student Jan Rehwinkel and other members of Elisa's group discovered another function of the protein.

Elisa Izaurralde, Ana Sofia Eulalio, Jan Rehwinkel, Daniel Schweizer, Isabelle Behm-Ansmant, Eric Huntzinger and Pavel Natalin



"GW182 is necessary for microRNAs to shut down their targets, because if you remove it from cells, they no longer do so," Jan says. "The cells use the RNAs to create proteins, as if there weren't any microRNAs at all."

Was GW182 involved in other types of RNA breakdown and recycling? The scientists discovered that nonsense-mediated decay doesn't depend on the protein, because the process continues even if GW182 has been removed. So GW182 seems to be specialized in processes involving microRNAs.

A very small genetic change can suddenly turn an RNA into the target of NMD, or a microRNA, and that can potentially have big effects on an organism.

Jan and postdoc Isabel Behm-Ansmant followed up by investigating other molecules that participate in the microRNA pathway. In animals, a family of proteins called *Argonautes* (AGOs) were known to be involved. Isabel and Jan showed that one of these, AGO1, probably docks onto an RNA first, and then is joined by GW182.

How do the molecules function? When attached to an RNA, GW182 slows down translation and increases the rate at which the molecule is broken down. It does this by working with a complex that strips the "cap" off an RNA – a group of proteins built on the head region of the RNA.

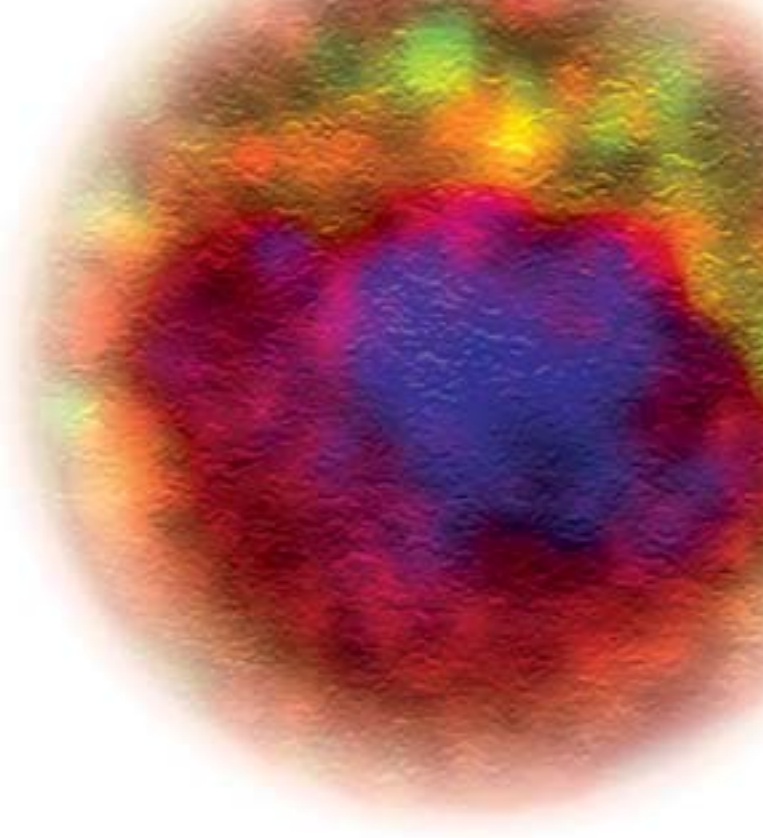
"Normally the cap forms a crucial link between an RNA and the machinery that translates it into protein," Elisa says. "Breaking this link is surely crucial in silencing the RNA. But there are some RNAs that function without the cap, which meant that microRNAs and GW182 shouldn't affect their translation."

MicroRNAs need GW182 to block translation; NMD doesn't. More experiments showed that shutting down NMD, by interfering with the molecules that govern it, didn't affect microRNA regulation. Clearly separate pathways were being used to cope with different types of RNAs, although they end up at the P-body recycling centres.

* * *

Evolution has spun off several types of RNA regulation. How far can all of these be traced back in the history of eukaryotic cells? The mechanisms are clearly ancient, because NMD, microRNAs and other types of translational control are found throughout the plant and animal kingdom and their eukaryotic relatives, like yeast cells. But when Elisa and her colleagues took a close look at these processes and their targets, they found that each organism seems to have developed its own flavors of regulation.

The machines that carry out NMD, they discovered, have been conserved over long periods of evolution. The molecules that carry out this process in diverse species are



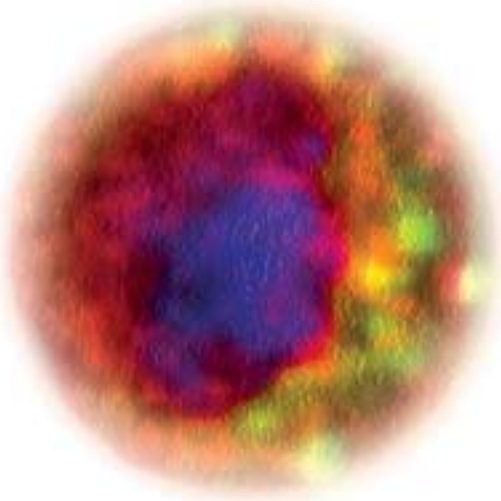
Microscope images reveal when and where the cell collects RNAs for destruction.

usually closely related to each other. But the targets, and their biological functions, vary from species to species.

One conclusion to draw is that fine-tuning molecules at the level of RNA is evolving quickly, and it may even contribute to the development of new species. A very small genetic change can suddenly turn an RNA into the target of NMD, or a microRNA, and that can potentially have big effects on an organism.

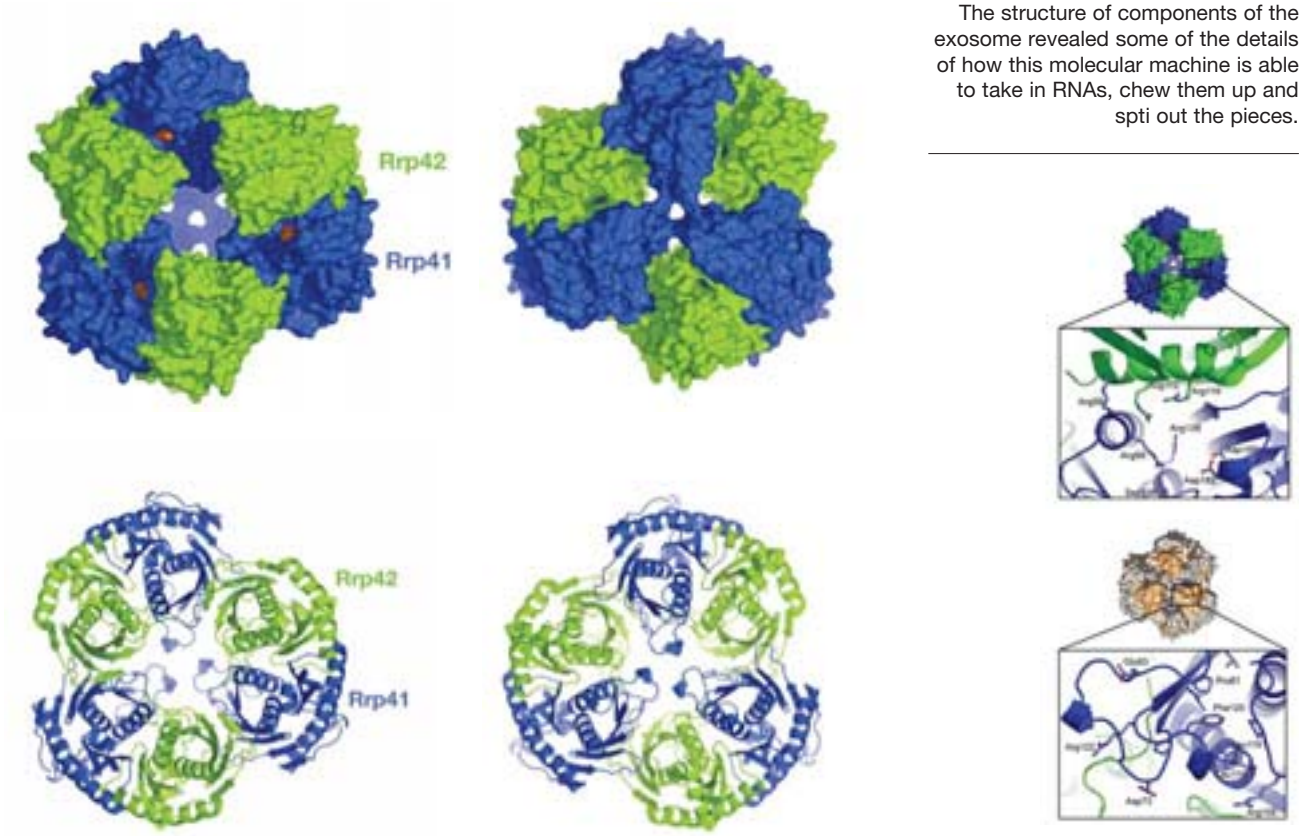
That in turn can help explain what are sometimes considerable differences between species with very similar DNA. Completing the genomes of humans and chimpanzees, for instance, has revealed how similar the genes of these species really are. There may be much more significant differences, however, in which proteins their cells express. And in species like ours, a considerable amount of that may be handled through the regulation of RNA (see story on page 182). ■

The machines that carry out NMD, they discovered, have been conserved over long periods of evolution. But the targets, and their biological functions, vary from species to species.



Elena Conti and Esben Lorentzen

Paper clips and shredding machines



The structure of components of the exosome revealed some of the details of how this molecular machine is able to take in RNAs, chew them up and spti out the pieces.

NOW THAT THE ELECTRONIC AGE has completely infiltrated most offices, it's hard to find a paper clip, and the paper shredder has mostly been replaced by the trash bin on the computer desktop. Cells have evolved, too, in the way they manage molecules that need to be destroyed. But they still have structures like a waste bin (see previous story), and Elena Conti's group has now gotten a close look at the cellular equivalent of the paper shredder.

Elena has been working on how molecules that dismantle RNAs recognize their targets. "This is handled in all organisms by a protein complex called the exosome," she says. "While the components of the machine have become increasingly complex over the course of evolution, its architecture has remained largely the same, and it continues to function in basically the same way. We've now taken a look at one of the oldest and simplest versions of the machine, which is found in one-celled organisms called *Archaea*."

The exosome grasps the tail of an RNA and begins chewing its way upwards, breaking the molecule down into nucleotides. At the heart of the machine are two proteins, *Rrp41* and *Rrp42*, which come directly into contact with the RNA. When that happens, they cut off a nucleotide; now the machine is ready to move up the RNA molecule and repeat the operation.

Esben Lorentzen, a postdoc in Elena's group, obtained crystals of Rrp41 and Rrp42 bound to each other. He wanted to see how this structure linked to RNA, so he soaked RNA molecules into the crystal. He also added *ADP*, which is produced when the exosome dismantles RNA. He took the crystals to an X-ray synchrotron beamline, where he captured a high-resolution picture of the structure. The results, Esben says, explain a great deal about the exosome's activity.

Rrp41 and Rrp42 pair up, and then three of these pairs link together to make a ring. While this study revealed the structure from Archaea, the exosomes of distantly related organisms like human cells also are built around such a ring. Interestingly, evolution has spun the original triplets of two molecules off into six separate proteins which form this structure in the exosome in our cells.

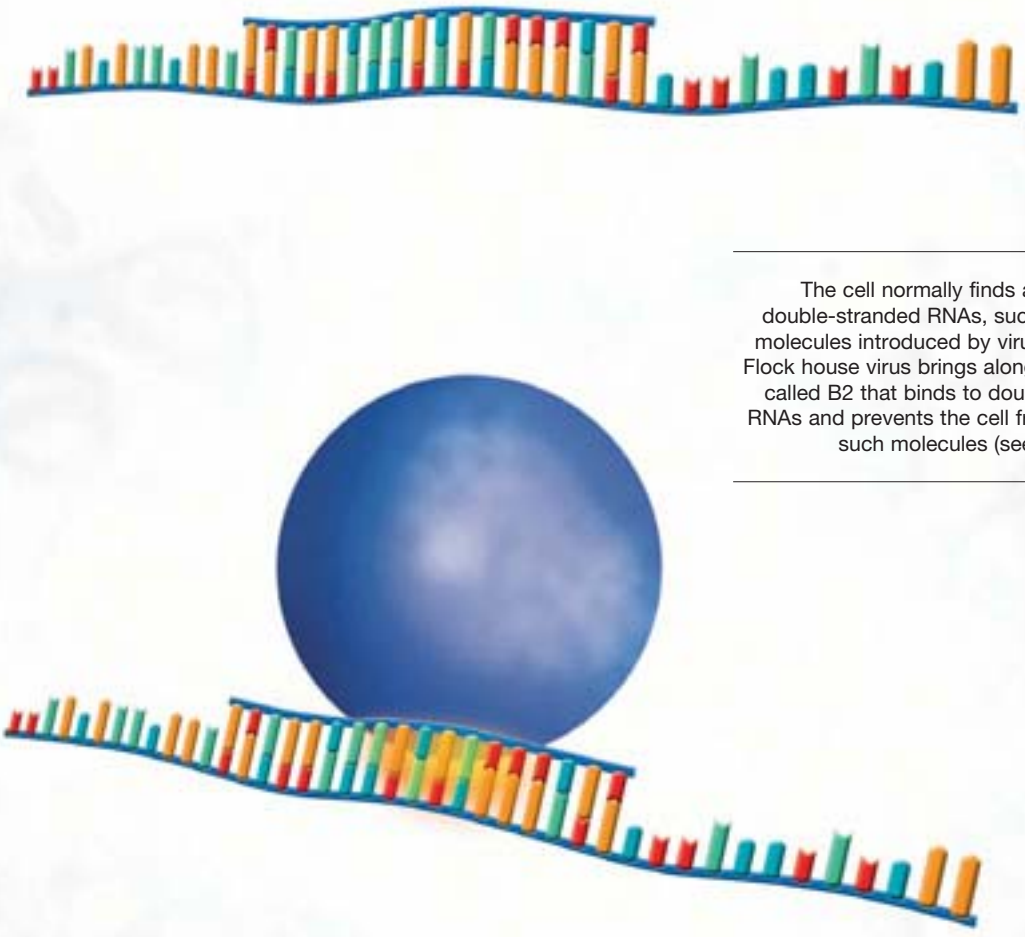
The hole of the ring is the active core of the exosome, where RNAs are handled. The study showed how Rrp41 and Rrp42 are bound to RNA and also how they link to ADP. RNAs are threaded into the hole and bind in pockets on the side of the ring. Esben and Elena could see atom-by-atom how the molecules fit together. In the case of RNA, not only the nucleotides interact with the exosome, but also the sugars that accompany them. "The sugars that link nucleotides in RNAs are different than those in DNA," Elena says. "This allows the exosome to

distinguish between the two types of molecules. That's important in organisms like yeast, because we think the exosome operates right at the sites where RNAs are transcribed from DNA. If it were capable of attacking both, it would damage the genetic material of the cell."

When an mRNA has been targeted for destruction by the exosome it is important that it not be released before being completely broken down into single nucleotides. The structural picture shows that this is achieved by holding on to multiple nucleotides of RNA at the same time. Once a nucleotide has been cut off the RNA, it has to move out of the way. Why do they move at all – why don't they get stuck at the active site? If that happened, Elena says, the machine would be blocked – it wouldn't be able to move up the RNA to remove the next nucleotides. Instead, the freshly cut end of the RNA is moved into the slot and the exosome dismantles the next part of the tail. Esben and Elena think that the cleaved nucleotides are ejected through channels along the side of the ring, and that the tail is brought into place with a sliding movement. "One interpretation of the structure is that the active site is more strongly attracted to nucleotides that are strung together, rather than loose fragments," Elena says.

This picture has revealed that RNA breakdown takes place in a shielded cavity that RNAs only reach when they are threaded down a central channel. Other parts of the exosome are responsible for locating the RNA and feeding it in. "The same type of channel is found in the proteosome, a complex that breaks down proteins," Elena says. "This suggests that it is important for cells to digest molecules in compartments – either in special areas of the cell, like P bodies for RNAs or lysosomes for proteins – or by creating protein machines with internal cavities."

* * *



The cell normally finds and destroys double-stranded RNAs, such as harmful molecules introduced by viruses. But the Flock house virus brings along a molecule called B2 that binds to double-stranded RNAs and prevents the cell from "seeing" such molecules (see next page).

Andreas Lingel, Elisa Izaurralde, Michael Sattler and Bernd Simon

When an RNA is produced by the cell, it is a string-like molecule, a single strand of nucleotides. Under some circumstances *double-stranded* RNAs are found in the cell, for example, when microRNAs bind to another RNA, or when scientists introduce a *small interfering RNA* (siRNA) into the cell to block the production of a particular protein. These double strands are then recognized by other molecules that prevent the RNA from being translated into protein, and they may destroy it.

"It is thought that RNA interference arose as an ancient defense mechanism against viral RNAs," says Elisa Izaurralde, group leader in Heidelberg. "Some viruses bring along RNA that has to form a double strand to replicate itself. Recognizing and destroying such molecules is a good defense strategy for the cell. There are also cases where the cell itself produces double strands, which also need to be blocked."

That means such viruses shouldn't be able to infect a cell which destroys double-stranded RNAs – but as can be expected, they have evolved ways to evade cell defenses. One that has done this is the *Flock house virus* (FHV), which infects insects but can also reproduce itself in leaves. When aphids eat the leaves, they catch the virus; later they deposit it in other plants, leading to infections of new insects.

FHV and the related *Nodamura* virus (NoV), which infects insects and mammals, bring along a protein called *B2* that can bind to double-stranded RNAs. By doing so, this molecule prevents cellular defenses from "seeing" the RNA and destroying it. Michael Sattler, Elisa and their colleagues recently took a close look at B2 using a method called *nuclear magnetic resonance* (NMR), hoping to discover how the protein works.

NMR molecules to be studied by exposing them to strong magnetic fields. The nuclei of the atoms within a molecule act like tiny compasses, which become "aligned" when placed inside the field. Scientists then probe the system by exposing it to additional magnetic fields, and

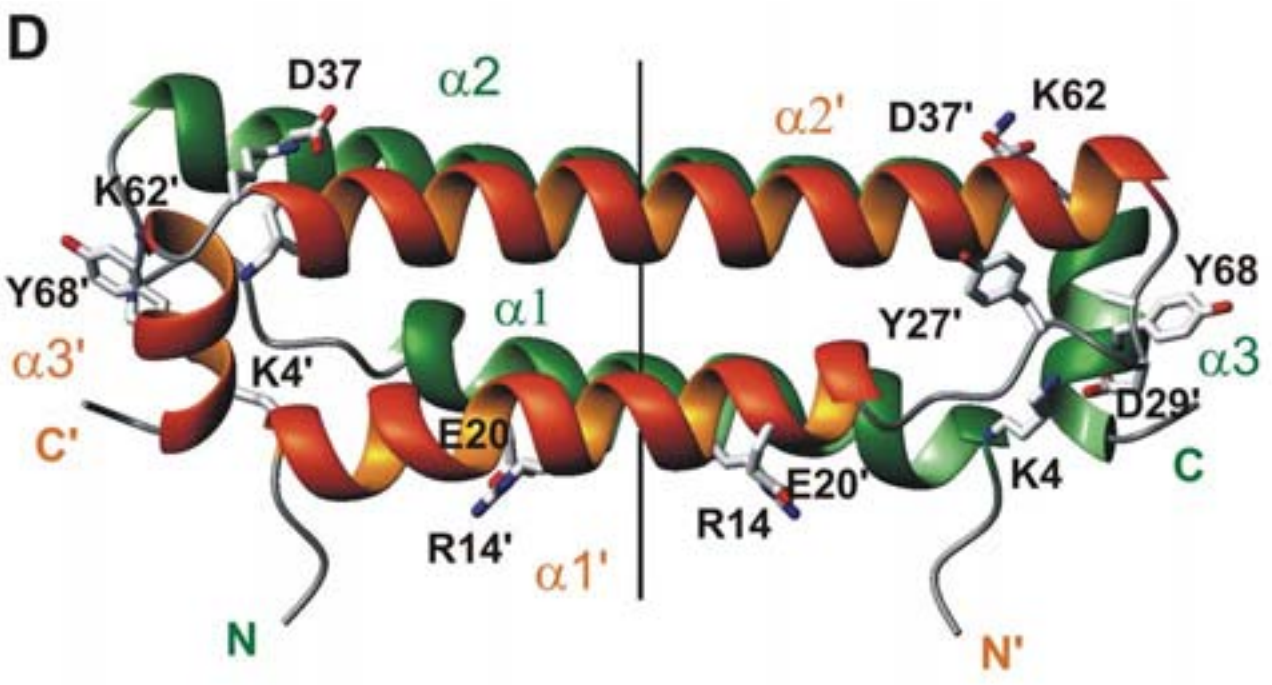
recording the NMR frequencies of the nuclei of individual atoms. The nuclei are influenced and can communicate with neighboring atoms within the molecule; from the measurements that they have taken, scientists can identify these atoms and plot their positions relative to each other.

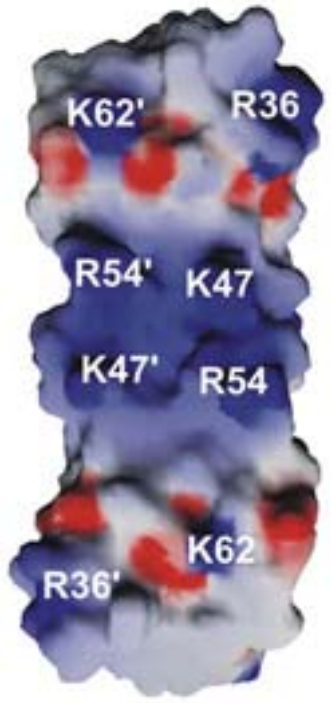
Andreas Lingel, a postdoc in Elisa's lab, along with Bernd Simon and Michael have used the method in the past to study other proteins involved in RNA recognition. What they saw in B2 was new.

FHV brings along a protein called B2 that can bind to double-stranded RNAs preventing cellular defenses from "seeing" the RNA and destroying it.

"Two copies of the protein link to each other to form a very elongated shape, something like a paper clip," Andreas says. "We haven't seen this type of fold before in molecules that bind to double-stranded RNA. The pair binds lengthwise onto an RNA and prevents other proteins from attaching themselves; they would need to do so to silence and destroy the RNA."

The shape adopted by the pair of B2s provides a very long surface to dock to an RNA of sufficient length, permitting it to recognize both double-stranded RNAs and other molecules, like small interfering RNAs. This gives it two ways to block cell defenses. As well as locking up the target, B2 can tie up RNAs that would otherwise bind to the RNA and call up molecules to eliminate it from the cell. ■

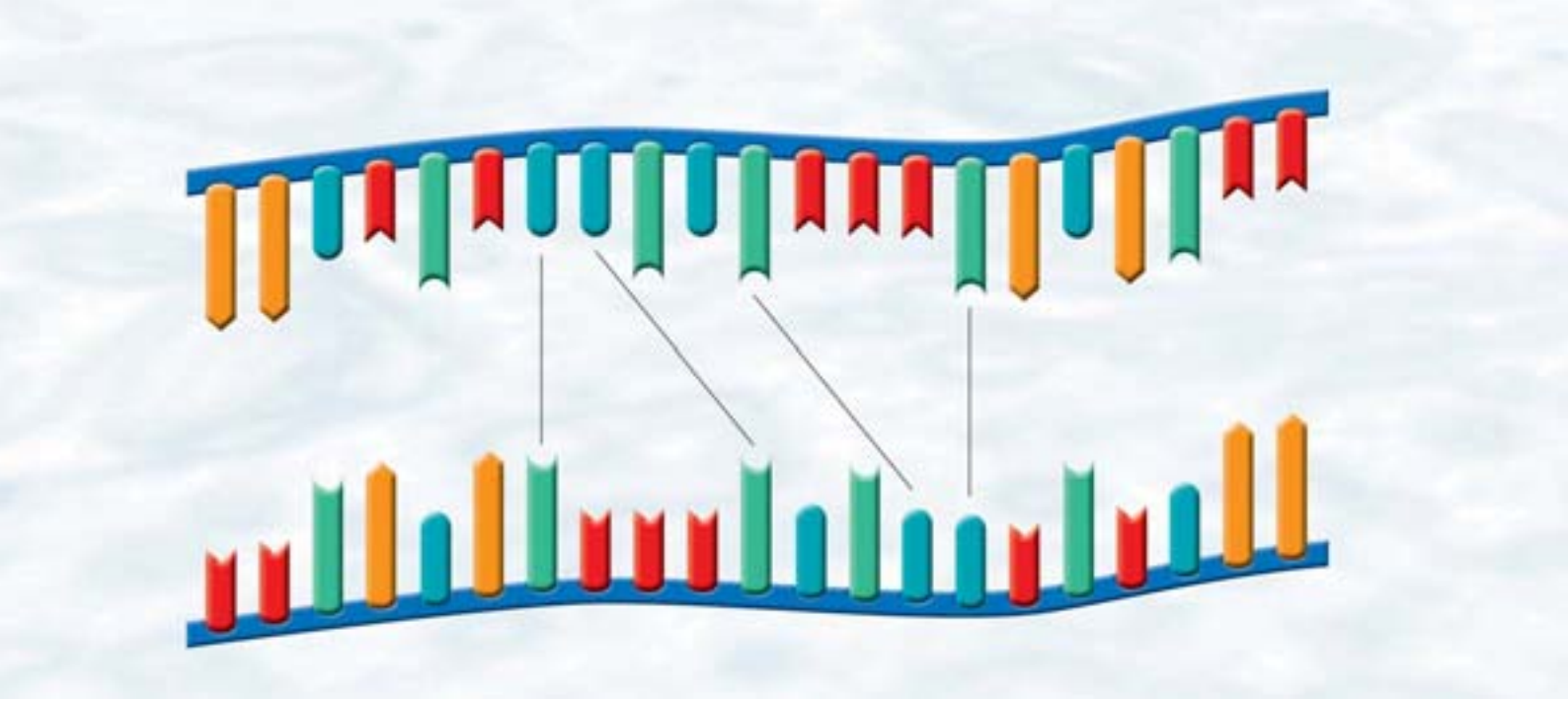




The structure of the B2 protein from the Flock house virus revealed that it has a unique way of binding to double-stranded RNA: two copies of B2 bind lengthwise, preventing other proteins from attaching themselves to the RNA.

Glauco Tocchini-Valentini, EMBL Council delegate and neighbor on the EMBL Monterotondo campus, is developing technology that will help EMBL researchers investigate cells at the level of RNA.

Off the beaten paths of genetic control



IT'S A BALMY SPRING EVENING in a small Italian town, with the scent of olives and the sea on the breeze, and Glauco Tocchini-Valentini is playing tour guide. As well as being one of Italy's most prominent scientists and a member of the EMBL Council, Glauco has an acute sense of history and culture, and during a walk through Ascoli Piceno, he points out the contributions of past epochs to the structure of the present-day city.

"When the Romans built roads, they were determined to make them as straight as possible," he says. "That's why the Corso Mazzini runs in a line from one end of town all the way to the other. But in the Middle Ages, the goal was to connect everything. So certain parts of town are like a labyrinth, with lots of small streets, a real network."

He sees a parallel to this in the way the cell operates. The information in the genome is contained in a linear sequence of bases – "The genome is Roman," he says. Proteins and cells, on the contrary are organized in complex networks, like the parts of towns built in the Middle Ages.

Glauco is Director of the Institute of Cell Biology within the Italian National Research Council, with a lab next door to the Monterotondo Outstation. His group has been working on new methods to move into the small and winding roads of the cell, and the result will be a new collaborative project with EMBL, dedicated to exploring a wide variety of biological questions in mice.

* * *

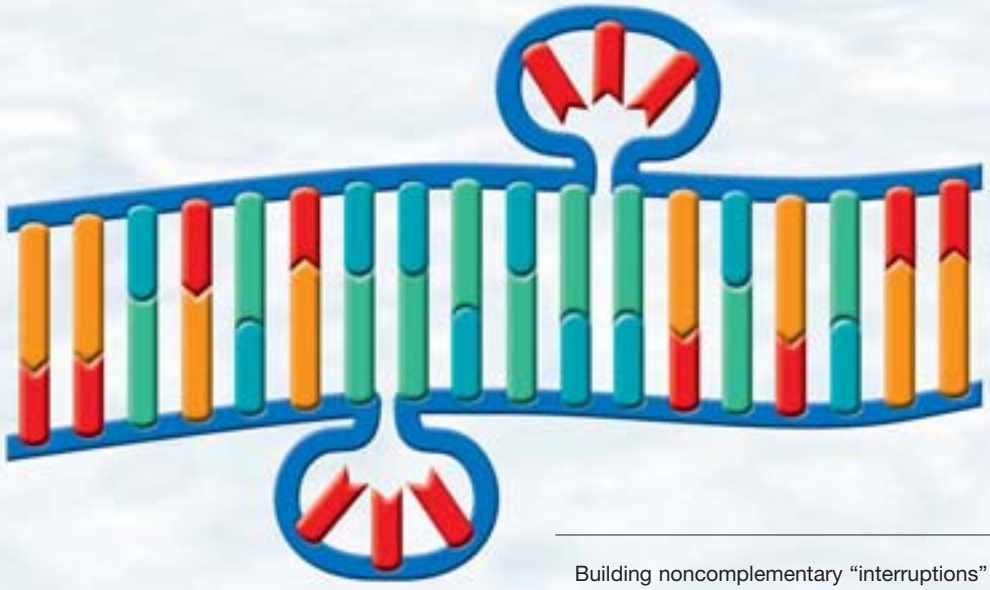
Over the last twenty years, scientists have become experts in modifying organisms' DNA to understand the roles that molecules play in cells, development, disease, and other biological processes. The most common techniques involve making direct changes in an organism's genome – removing genes, altering or adding them. Understanding how this affects cells and organisms, Glauco says, is like trying to understand a village by never straying from the straight Roman roads.

"When you eliminate a gene, it's hard to predict what's actually happening," he says. "Just as one example – a gene is located on one of the strands of the double-helix, but when you knock it out, you inevitably remove both strands. The second strand could contain information that shouldn't be lost – a second gene, or a sequence for a microRNA, or something else."

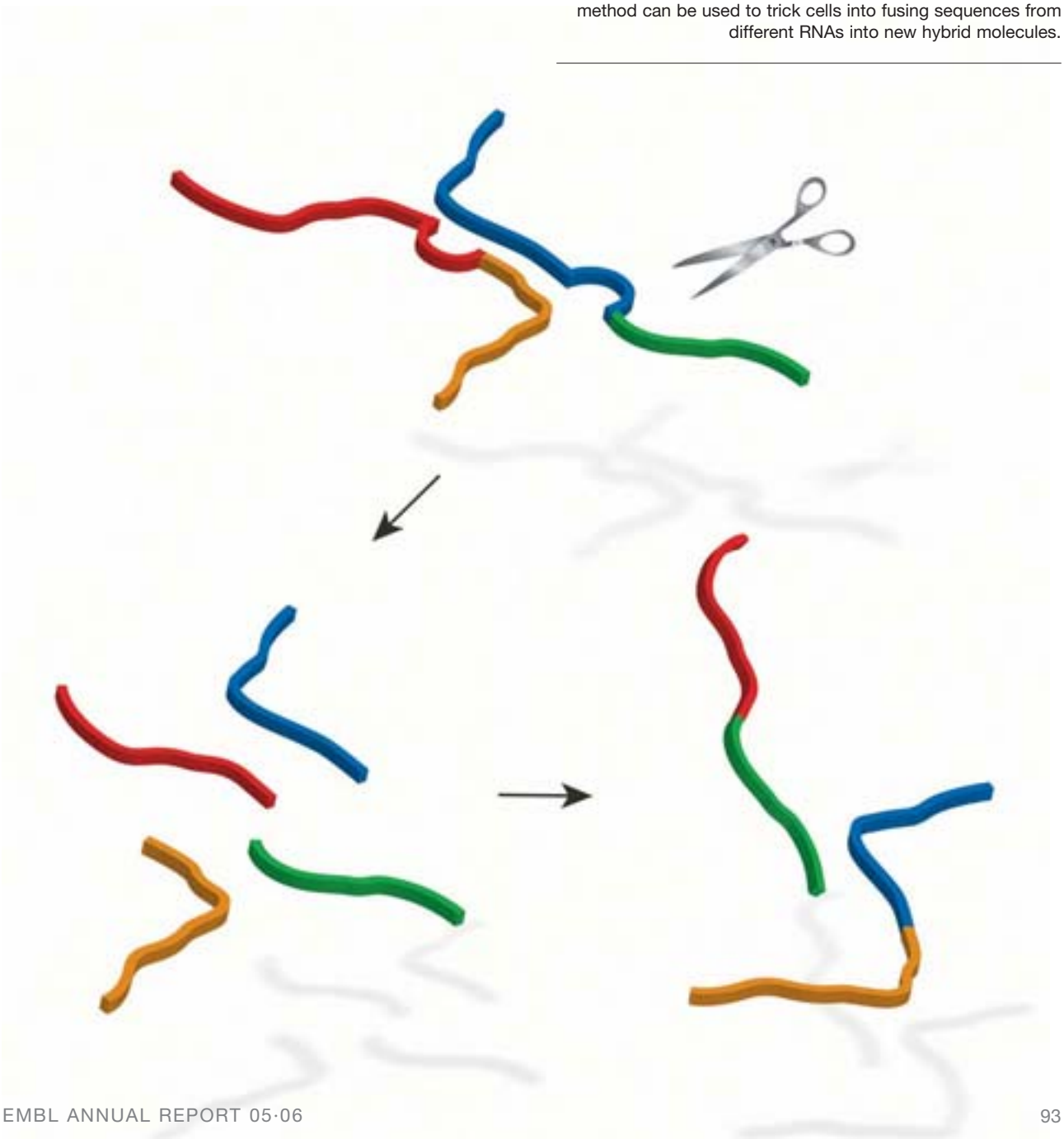
To get a sense of what is going on in the less linear, cluttered information alleyways of the cell, scientists have been developing other techniques which act at later stages in the information pathway – for example by interfering with RNA molecules. One approach is to insert artificial, *small interfering RNAs* (siRNAs) into cells which seek out and bind to a complementary molecule and prevent it from producing proteins. This technique was discovered seven years ago, and scientists quickly helped develop it so that it could be applied to the study of genes in insects and mammals.

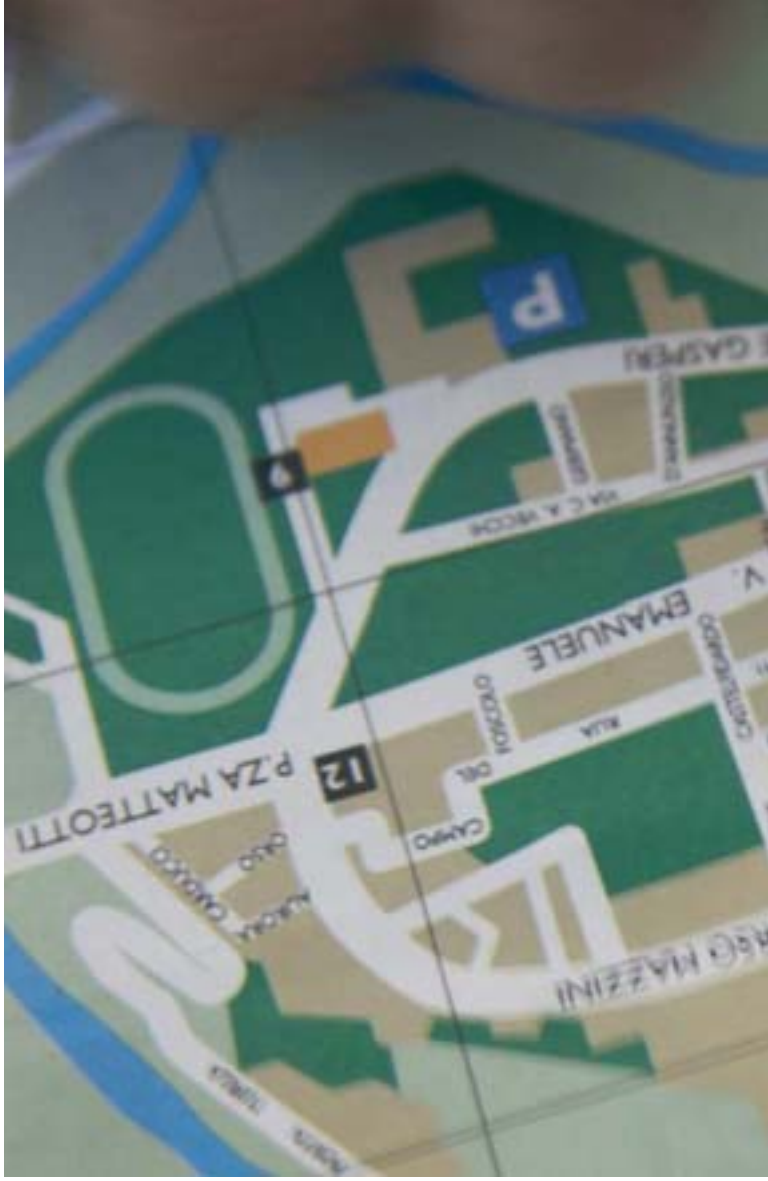
Even this, Glauco says, can have an effect that may be more generic than is desirable. "Unfortunately, a given siRNA can often target more messenger RNA types than the one for which it was designed. What we would really like is a technique that affects one specific molecule, in one context. So we need to create more methods that target RNAs or proteins."

Now Glauco and his group are developing new tools to engineer RNAs based on molecules from one-celled organisms called *Archaea*. The methods will allow scientists to shut down specific RNAs, to activate them in new



Building noncomplementary “interruptions” into genes causes two RNA strands to join in a structure called a Bulge-Helix-Bulge (above). This structure is recognized and removed by an enzyme taken from one-celled organisms called Archaea. Mouse cells then try to repair the molecule, gluing the broken ends together into a new RNA molecule without the bulge. This method can be used to trick cells into fusing sequences from different RNAs into new hybrid molecules.





ways, and to combine them to create entirely new proteins in organisms like mice.

The scientists' approach makes use of the fact that organisms both dismantle and sometimes glue together fragments of their own RNAs. Enzymes accomplish this by recognizing chemical signatures and structures in RNA molecules which tell them where to make cuts and what to join.

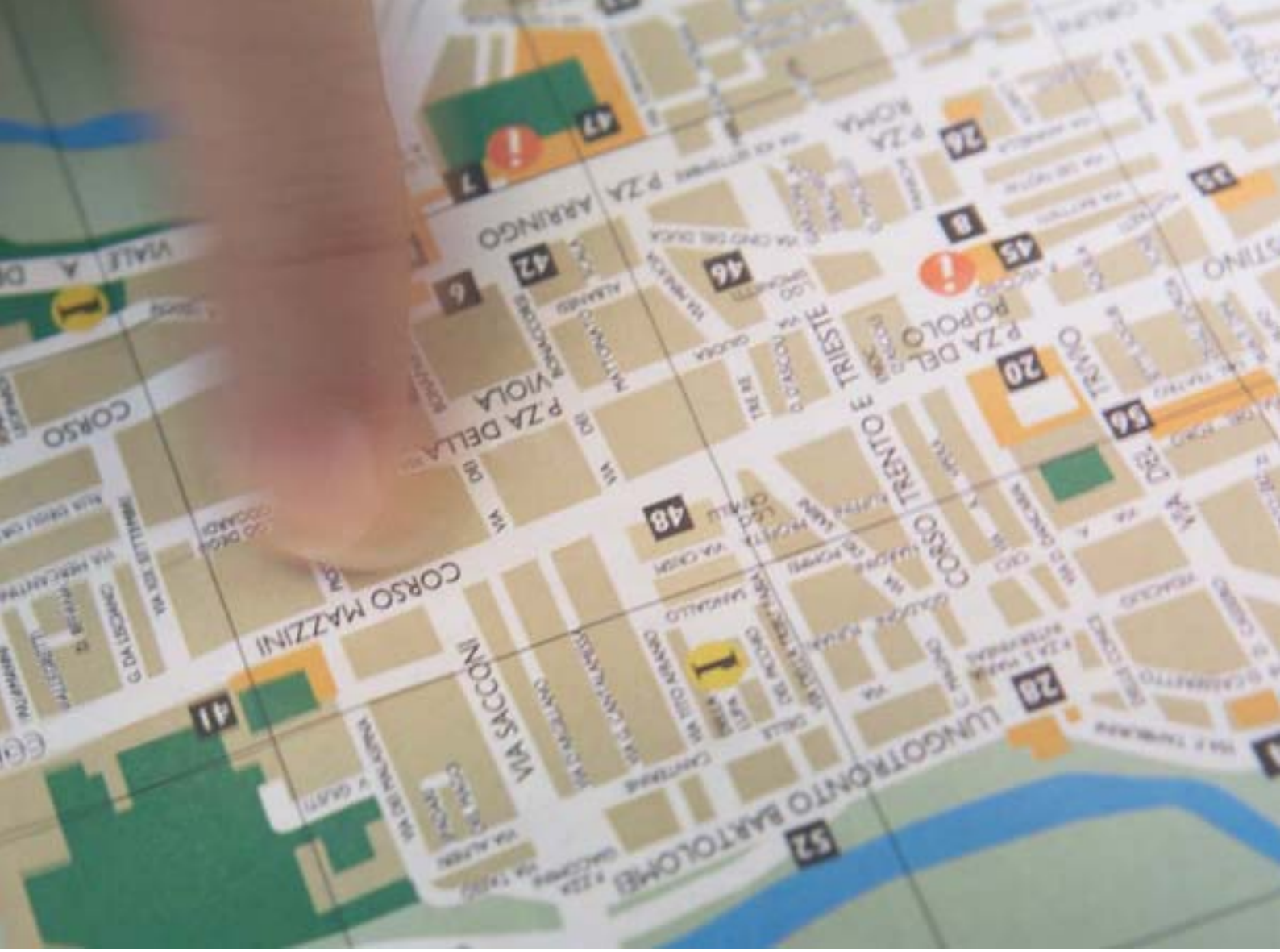
Species of *Archaea* have a cutting enzyme called a *tRNA endonuclease* which recognizes a very specific structure in RNAs called a *Bulge-Helix-Bulge*, or BHB. Chemical and structural studies revealed the details of BHB architecture. As the name implies, the structure consists of two bump-like shapes separated by a short helix. This is a small pattern which forms because two complementary strands of RNA bind to each other, making a helix (the way that two strands of DNA bind) – but in this case the match isn't perfect. Molecules that would otherwise make a perfect fit are interrupted at two places by extra nucleotides in one of the strands (see the image above). When they join, it is as if a piece of paper slides a bit while it is being glued to another, making wrinkles.

This can happen within a single molecule. Sometimes an RNA has a sequence which is complementary to another sequence within itself, and it folds over and binds. It can also happen between two separate RNAs which are almost complementary. The formation of a BHB doesn't depend on particular sequence; a wide variety of molecules with different codes can link in this way.

When the endonuclease from Archaea sees this structure, it makes a cut at each of the bumps. This produces fragments that are recognized by repair enzymes. They try to figure out what happens and glue the broken ends together again – making a new RNA that lacks the BHB. The result is a molecule that can sometimes be translated into proteins.

This endonuclease is found naturally only in one-celled Archaea – would it also cut RNAs in animals such as the mouse? "We were very interested in this question," Glauco says, "because we thought it might give us another kind of tool to manipulate RNAs very specifically."

The scientists did a test with the *green fluorescent protein* (GFP). GFP was originally derived from fluorescent jellyfish; it has now been transformed into a tool that can be inserted into the genomes of other animals, either as a tag on another molecule or as a probe to study processes like how RNAs are translated into proteins. If an organism successfully translates GFP, the protein can be observed under a fluorescent microscope.

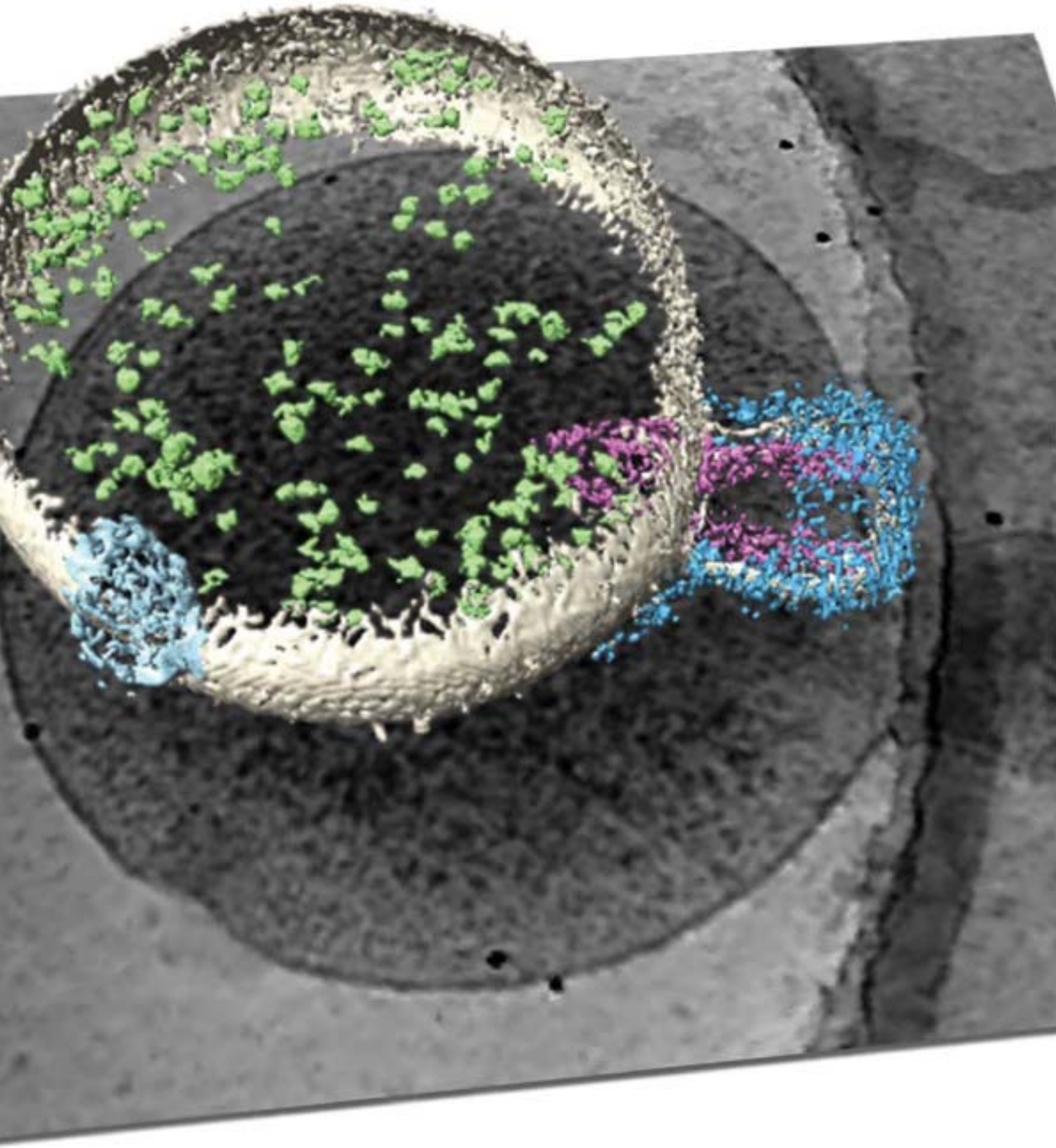


Glauco and his group put a specially engineered form of the GFP gene into mice. They had altered it by adding bits of sequence which would cause its RNA to fold into a BHB structure. These added bumps disrupted the machinery that translates the RNA into protein, so the mouse cells didn't produce any GFP. When the researchers added Archaeal endonuclease, however, the extra structure was sliced out and mouse enzymes repaired the RNA. Now it was translated, and they could see the GFP signal.

The next step was to discover whether the same strategy would cause a fusion of fragments from *different* RNAs. "To do this, you pick a target RNA," Glauco says. "You have to insert into the genome code that will create a second RNA, almost complementary to this target; again, it contains a few extra sequences necessary to form a BHB. The cells produce RNAs corresponding to both the new gene and the target and they bind to each other. If you add the endonuclease, it will break up the BHB, and once again repair mechanisms fuse the fragments. But this time you have joined pieces that used to belong to separate molecules."

Unlike RNA interference, Glauco says, such endonuclease "surgery" is extremely specific – it only affects preselected molecules. There are many potential uses. "For example, suppose that the target RNA and the new one belong to separate signaling pathways in the cell," he says. "Joining them in a single protein would force the two pathways to communicate with each other. This gives us a very precise tool to dissect and direct the transmission of information within the cell, which is crucial in a huge number of processes and diseases such as cancers."

The scientists have their sights set on a wide range of new projects. They call the new technology *Archaeaexpress*, and plan to put it to work in parallel with other types of genetic engineering. In Monterotondo they will create strains of mice for use in studies of behavior and disease, combining Archaeaexpress with "knock-out" technologies that shut down genes in only one tissue at a particular stage. This will take them into the alleyways and backroads that branch off the main road between gene and protein – subtle levels of control which play a decisive role in how life works. ■



The proteome

IN 1827, the French zoologist Henri Milne-Edwards claimed that our bodies are factories: not a single machine, but a collection of tiny machines, each with different functions. Although he had no idea of the scale of biological organization it would be applied to, his metaphor has persisted for nearly two centuries. Today's scientists speak of molecular machines, referring to a level deep within the cell: complexes of proteins and other molecules, tiny robots in a chemical world.

Milne-Edwards and his contemporaries were coming to believe in a mechanistic account of life: ultimately, organisms arise through physical and chemical principles working on matter. Although the first chapter of this report provides some examples of how molecules self-organize into complex structures, we are still far from a thorough machine-like description of life. Achieving that will require at least an exhaustive "parts list" of cells, a complete knowledge of their possible interactions, and an understanding of the rules of games within games, being played simultaneously at many levels within organisms. Theoretically, genome projects can provide the parts list – by uncovering what is encoded in organisms' DNA. But new findings reported in the second chapter of this report reveal that this goes far beyond proteins; a wealth of non-coding RNAs are also produced, and they actively participate in governing the cell.

Even without the complete catalog of cellular components, modern methods are revealing significant parts of the user's manual. We are learning when machines are assembled and some of the rules by which they operate. High-throughput approaches which look at organisms and processes from the top down promise to fill in many of the blanks. This section is a progress report on some of those efforts, with a particular focus on proteins and protein machines. ■

New techniques like cryo-EM tomography, used by the Frangakis group, are giving scientists a holistic overview of where molecules and complexes are active in cells.

Gitte Neubauer, Anne-Claude Gavin, Rob Russell and Peer Bork having fun in EMBL Heidelberg's mechanical workshop, which produces high-quality instruments used throughout the Laboratory.

Retooling the factory floor

A FEW YEARS AGO, EMBL team leader Giulio Superti-Furga was making the rounds of European and American investors, trying to attract interest in a new biotech start-up company. Cellzome's vision, he explained, was to understand the cell in a new light. "Over the last few decades, molecular biology has made great strides in deciphering the functions of single genes and proteins," he said. "But most proteins carry out their jobs in complexes that may contain dozens of molecules. Until now, our view of such 'molecular machines' has been restricted by limitations in technology."

But things were changing fast. The completion of genomes was providing a comprehensive parts list of the molecules that organisms could produce, and innovations developed largely at EMBL convinced Giulio and his colleagues that the time was ripe to make a high-throughput assault on analyzing their assembly into machines. The repetitive, multidisciplinary and large-scale nature of the task made it the sort of project that would be impossible to do at EMBL alone, but it could be carried out by a company.

The result has been a five-year collaboration which exemplifies, Giulio says, the relationship between industry and basic research. The project encouraged a smooth exchange of not only expertise, but also of personnel – Giulio stayed at EMBL as a visiting team leader while serving as the company's scientific vice-president. And early in 2006, Anne-Claude Gavin, who has had a leading role since the beginning, moved back to EMBL as a team leader.

The work has culminated in a rich catalog of the molecular complexes in Baker's yeast. "We picked yeast because of the manageable size of its genome and the fact that it is one of biology's most-studied organisms," Anne-Claude says. "It's also a *eukaryote* – on the same major evolutionary branch as our own cells – meaning that many of the general principles we learn can be applied to understanding human cells."

The result is an unparalleled view into what has been largely unexplored territory: a crucial stage of biological organization, a level at which the information contained in genes determines the behavior of cells and organisms. The molecular complexes found in the study, and the way that they are managed by cells, are beginning to give scientists new insights into themes ranging from the normal activity of cells, to the disruption caused by genetic diseases, to questions of animal evolution.

* * *

In the mid-nineteenth century, dramatic improvements in the light microscope enabled two German scientists to see that both plants and animals were composed of cells. As this occurred, chemists were beginning to learn how to break down organic substances into more basic elements. In 1901 Franz Hofmeister compared the cell to a factory, able to take in raw elements and convert them into the necessities of life; he even suggested that the subcompartments of cells that had been identified under the microscope might be responsible for specific types of conversions.

The factory analogy has persisted through a century of discoveries about the functions of molecules. Proteins were described as "worker molecules" and chemical processes as assembly lines. Unlike a car factory, however, where machines usually remain bolted to the floor and are only changed as new models come into fashion, the cell continually retools itself. Proteins are simultaneously workers and components of intricate robots that are continually assembled and taken apart; the same molecule can often be found in several machines.

The full extent of this flexible organization has only become clear through the recent study; previously, scientists had a very limited view of the machines and their components. "The situation was like coming into a factory and finding parts of single machines scattered across the floor," Anne-Claude says. "We know what some machines do, and a bit about how they operate, but there was really no view of the whole context."

Scientists had already started to piece together the construction of yeast machines based on single components, using a method called a *two-hybrid screen*. This matches every yeast protein to every other, like completely disassembling everything in a car factory and trying to fit pieces together one-by-one. The method has generated both a wealth of useful information and many false leads. It might be physically possible to insert a gearshift into an exhaust pipe, but that doesn't mean it ever happens in a functioning automobile. With 6500 parts to deal with – the approximate number of proteins encoded in the yeast genome – the pair-by-pair method provides a very limited view of complete machines, let alone the whole factory.

An alternative would be to start with complete machines and then analyze the molecules that compose them. But methods of extracting proteins from cells usually broke complexes apart. Then several years ago, Bertrand Séraphin's lab at EMBL invented a process called *tandem affinity purification* (TAP), a method which fishes single molecules from cells along with any machines attached to them, completely intact. The components of complexes could then be analyzed through *mass spectrometry* – a method that fragments proteins and "weighs" the pieces. Since each protein has a unique composition, mass spectrometry gives scientists measurements that can be matched by computer to the profile of a specific molecule. Significant developments in these techniques were also achieved at EMBL in the 1990s by Matthias Mann, Matthias Wilm and their colleagues.

Put together, TAP and mass spectrometry gave Cellzome the tools they needed to take on the entire yeast proteome. Thousands of experiments later, this has produced the first full scan of the genes of a eukaryotic cell, searching for molecular machines.

* * *

The study revealed 491 complexes, 257 of them wholly new. The rest were familiar from other research, but now virtually all of them were found to have new components.

"The situation was like coming into a factory and finding parts of single machines scattered across the floor. We have known what some machines do, and a bit about how they operate, but there was really no view of the whole context."

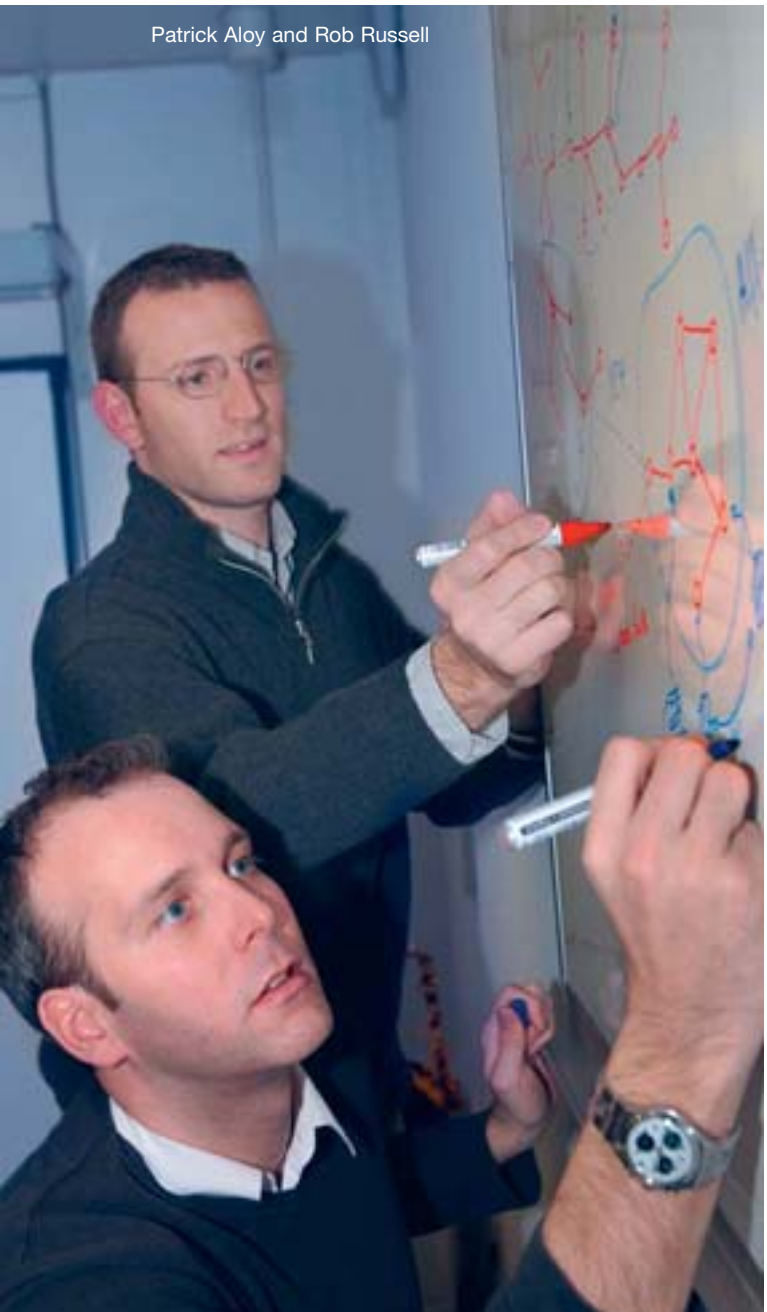
Is the list exhaustive? "We estimate there may about 300 more," Anne-Claude says. "Some complexes may appear only when particular conditions are used to grow the yeast, and others may not be discoverable with this method of extracting them." For example, it has been notoriously difficult to purify complexes attached the cell membrane. The researchers adapted their methods to accomplish this, and discovered 74 new complexes containing membrane proteins, but she is sure that many more remain to be found.

A parts list is only the beginning: the scientists also want to know where the complexes are stationed in the cell, what they do, and how they function. Sometimes these questions can be answered from the components alone. A complex with three proteins that respond to heat undoubtedly plays a role in helping the organism adapt to changes in temperature. Other complexes could now be linked to functions such as binding to DNA, or assisting in the processing of other molecules.

The information has also provided new insights into how the cell manages the incredibly complicated task of putting complexes together, and this says something important about the biology of yeast and other organisms. "Does a cell preassemble machines and have them on hand, or are they built from scratch when something happens?" Anne-Claude says. "In other words, how are the machines – and the factory as a whole – really managed? We didn't know, but now we can say a lot about this. But first we had to find new ways of understanding the data."

* * *

Road atlases often contain a chart showing the driving distances between cities. For example, 1150 kilometers separate Rome and Heidelberg; Heidelberg to Cambridge is 2045 kilometers (counting a transit by ferry). From such a chart and a bit more information, you could sort cities into regions, states, and countries. To understand the inner workings of protein complexes, says Patrick Aloy, scientists would like to have such a map of molecules.



Patrick Aloy and Rob Russell

Patrick is a member of Rob Russell's group, in which computational techniques are used to understand the inner workings of protein complexes. Combining information on protein shapes and functions with data on how they bind to other molecules has permitted the scientists to start drawing "technical diagrams" of machines. But sketchy knowledge has restricted the scope of these efforts.

"Imagine replacing a chart of distances with a table that tells you only yes or no, 'Can you get there from here?'" Patrick says. "From that information you wouldn't be able to draw a very meaningful map, but that type of study is what we've had so far. Well, now we've produced something more like the distance chart – each pair of proteins has a value which gives the likelihood that they are found together in purifications."

That information has now been converted into a map of the factory floor, complete with finished and partial machines, prefabricated parts, and snap-on modules.

"What you discover is that most complexes have a core set of components that are almost always found together and others that come and go," Rob says. "You can think of the cores as crucial, prefabricated parts of machines that are kept on hand, with temporary modules added on as the need arises."

The function of such modules, Anne-Claude says, may be to alter the job of the core machine, to link it to other things going on the factory, to switch it on or shut it down. "This has several very important effects. First, it gives the cell a way to carry out a large number of tasks with a limited number of basic machines. That gives it quite a bit of flexibility. Second, it means that in an emergency, the cell doesn't have to build all the machines it needs from scratch. It only has to produce a few really essential parts. The other side of that is that it may be relatively simple for the cell to control a quite sophisticated

machine, just by supplying or blocking the delivery of a critical piece."

There is an important connection to evolution because generally, certain types of machines and their basic components have been conserved over hundreds of millions of years as new species arose. Peer Bork's group has helped investigate this question.

"If you compare what goes on in yeast and our own cells, you find many of the same machines, using the same basic components to do the same things," Peer says. "The complexes reflect how evolution works – as variations on a theme. You don't find every species inventing a new way of doing things; instead, they're refining what they're up to by adding on specialized modules, or slightly changing the way the whole thing is regulated."

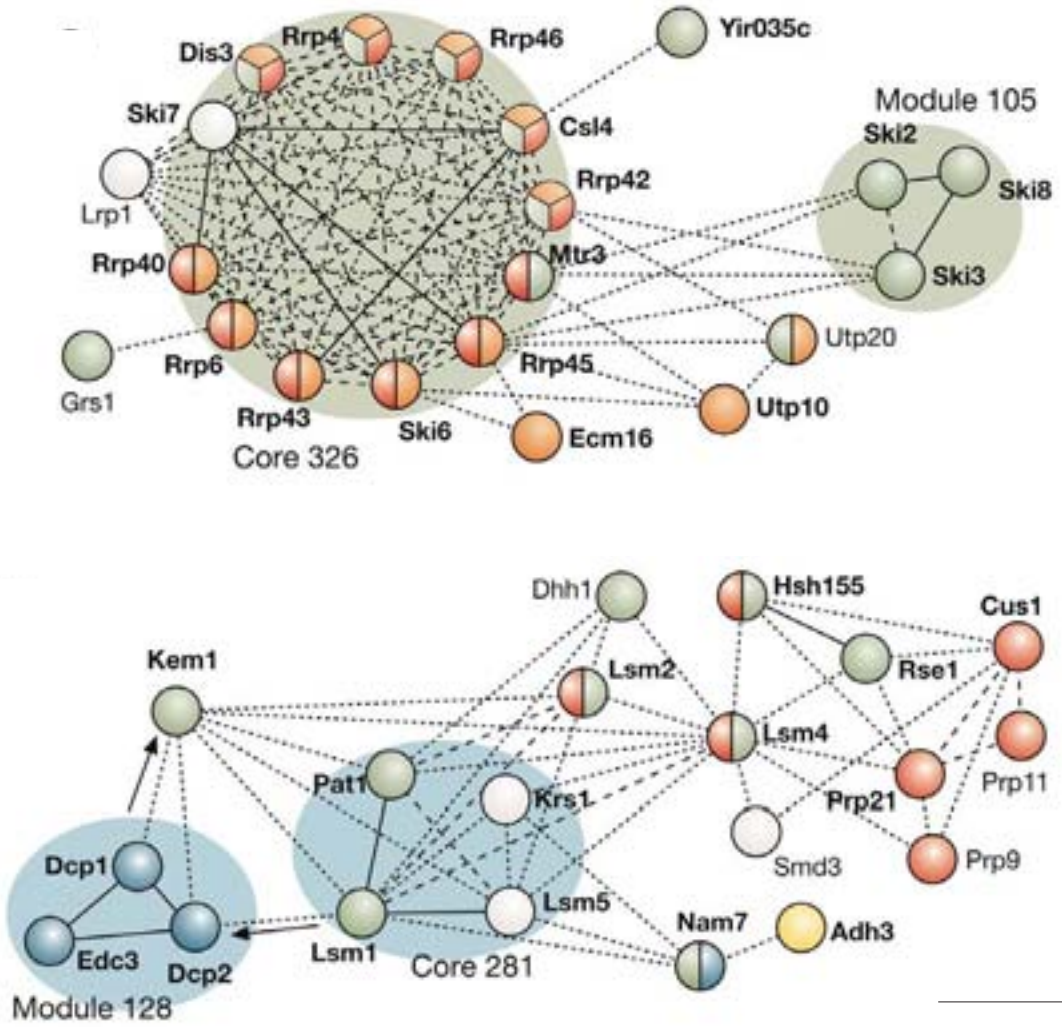
The study reveals a great deal about individual machines and how they work together. Yet much remains to be learned about their work in living cells – where many of them are located, and how many copies of each machine are at work at any given time. Structural information about the complexes is helping to answer some of these questions, because it gives scientists an idea about the overall shape of a complex. This means they can look for it under the microscope.

"Even with electron microscopy, protein complexes are fuzzy spots that are difficult or impossible to identify," Anne-Claude says. "But with good shape information, we might be able to put names to some of the shapes."

The groups, along with most other labs in EMBL's Structural and Computational Biology Unit, are working together to do this in a major EU-funded project called 3-D Repertoire, headed by Luis Serrano. Their goal is to combine structural information about complexes with imaging techniques to make a real plan of the factory floor – or to convert Patrick's driving distance tables into a satellite map.

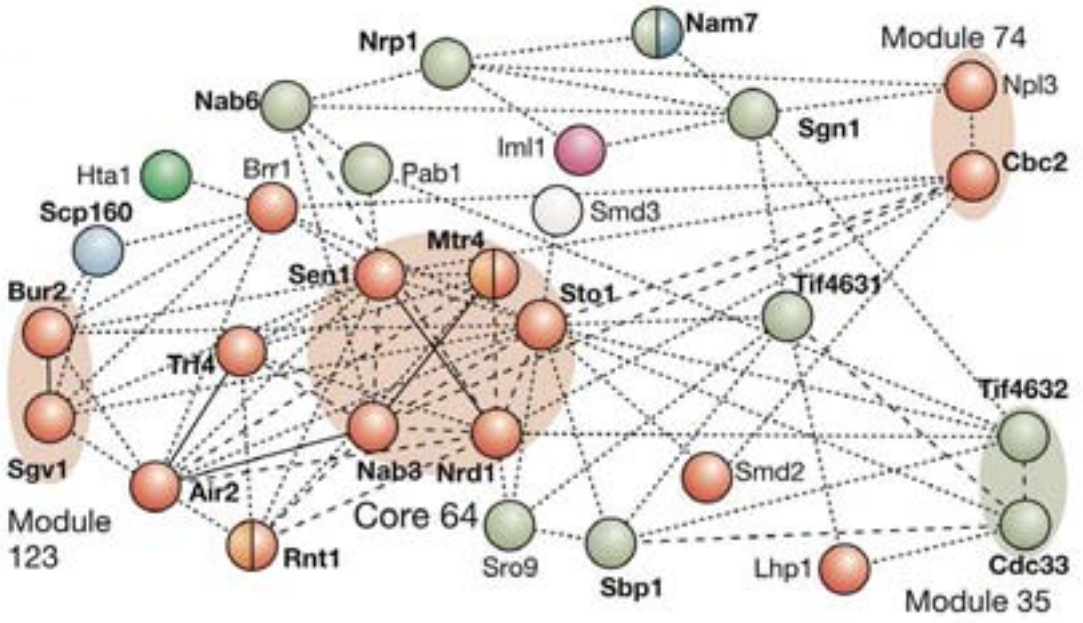
Giulio Superti-Furga has now moved to Vienna to head a new centre for molecular medicine as part of the Austrian Academy of Sciences. He says that what the study shows about the nature of machines brings things full-circle, to hopes of medical discoveries.

"Health reflects the operation of the entire cell in the context of the organism," he says. "The level of molecular machines is crucial in influencing that and keeping things in balance. What we have accomplished is changing what we think not only about how individual machines work and are regulated, but also about how they function together. That will obviously be crucial in guiding organisms from a diseased state to a healthy one." ■



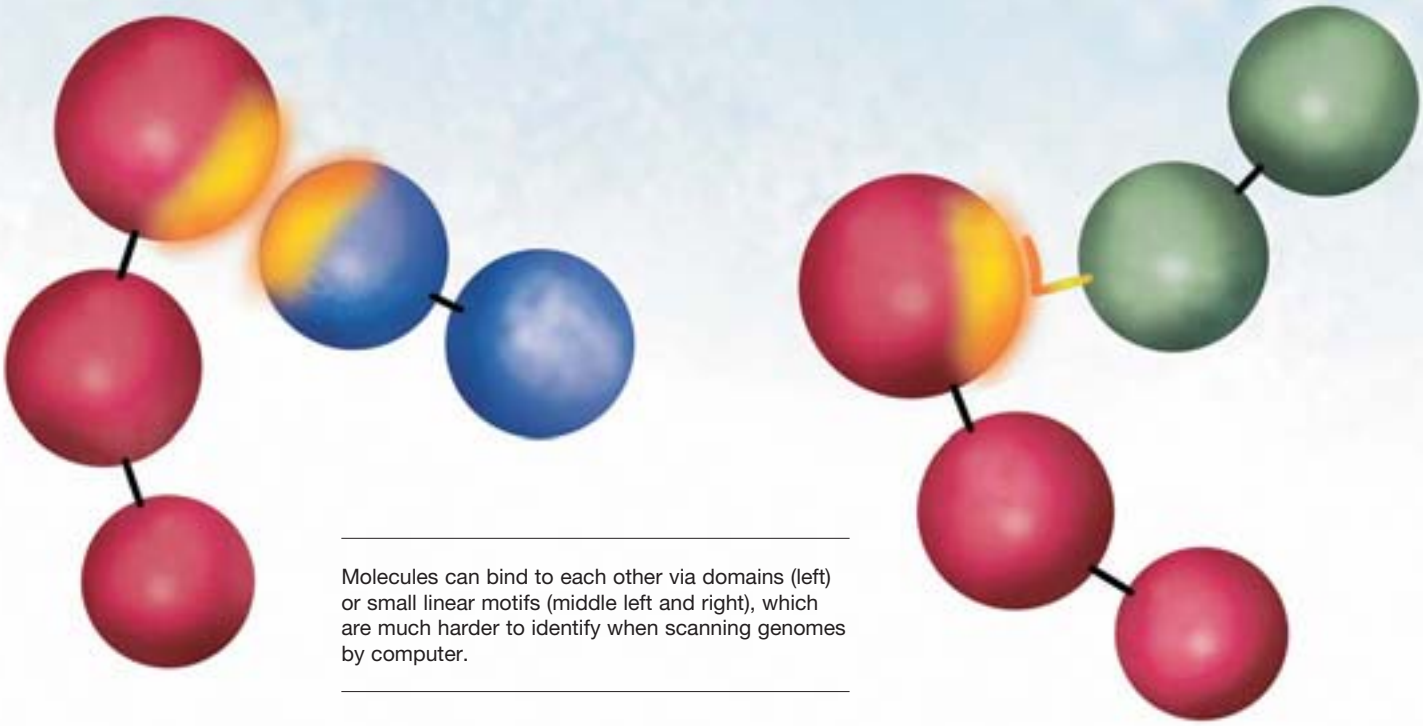
Left: Patrick Aloy

This page: some of the complexes and dynamic modules discovered in the yeast proteomics project.



Victor Neduva and Rob Russell

The cell's conjunctions



Molecules can bind to each other via domains (left) or small linear motifs (middle left and right), which are much harder to identify when scanning genomes by computer.

A RECENT GOOGLE SEARCH for the conjunction "and" returned 23 billion pages, probably every page written in English on the Internet. Searching genomes for specific sequences containing just a few letters gives similar results. Yet it would be helpful to find them: tiny protein sequences called *linear motifs* play a crucial role in many cell functions.

"Everything about the way the cell works depends on molecules recognizing each other – from the construction of cell machines, to communication networks, the transport and proper localization of molecules, the activation of genes – everything," says Rob Russell. "Often that recognition happens between the large surfaces of protein modules. But these modules also recognize linear motifs, and very few are known, because it has been difficult to identify the sequences involved."

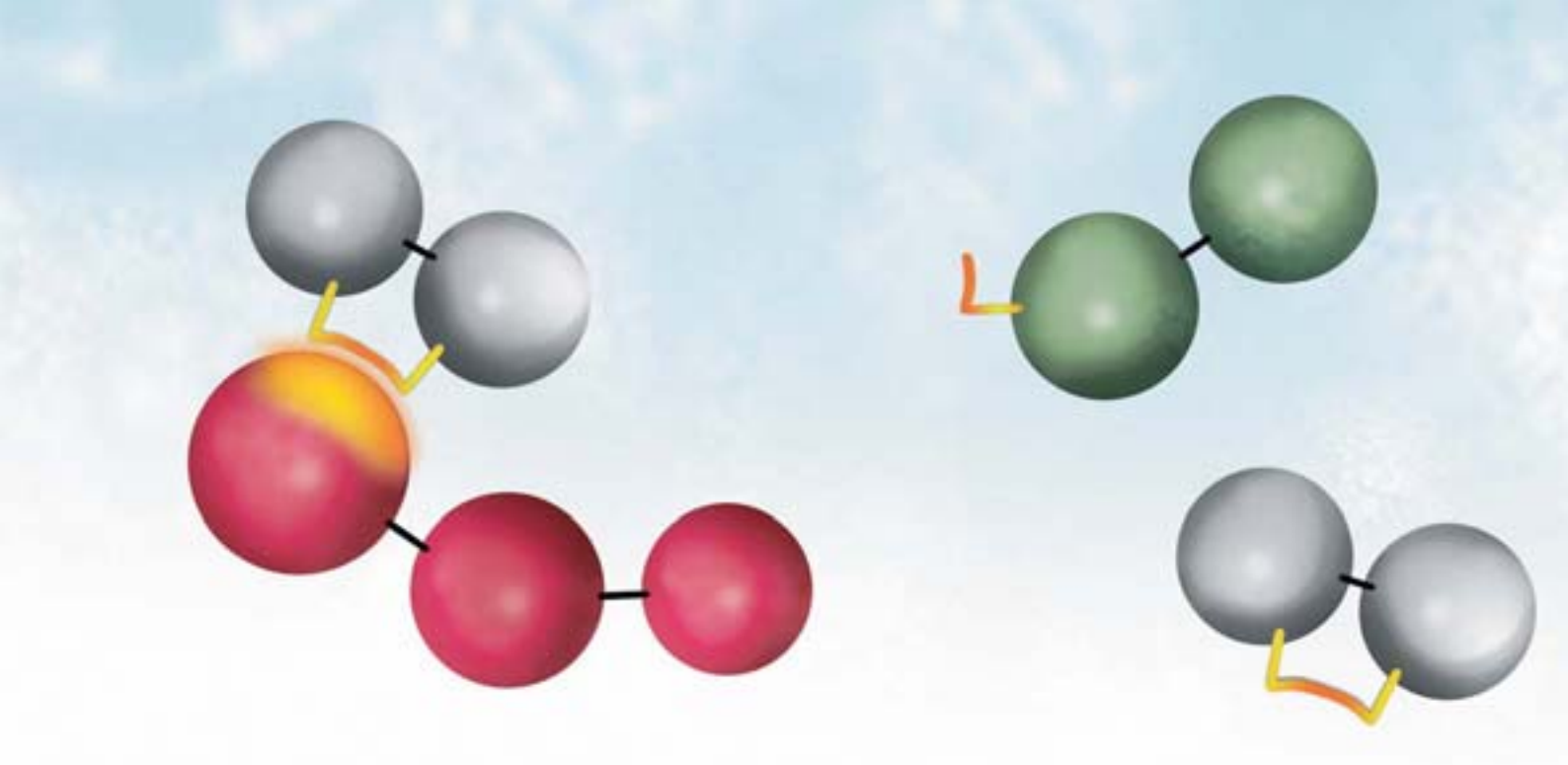
Large surfaces bind to each other because of complementary chemistries and shapes; usually several features are involved, and it's possible to detect them by looking at sequence and structural information. So Rob and his colleagues have developed computer algorithms to scan genomes for these features. Given one protein domain, they can often make predictions about its binding partners. Information from the yeast proteome project (previous story) is allowing them to improve their search methods.

Finding linear motifs is much more difficult, but PhD student Victor Neduva, Rob and other members of the group have now found a way to do so. "We began by assuming that if many partners bind to a common molecule, they must share features that permit them to do so," Victor says. "That should be either a domain or a linear motif. So we eliminated partnerships involving two domains and began looking for motifs."

Even with this smaller sample, it was a huge job to try to find tiny patterns in proteins that were hundreds of amino acid "letters" long. A linear motif might consist of just two or three strategically-placed letters, interrupted by "wild cards" – random bits of sequence. For example, a protein module called the SH3 domain, found widely in signaling molecules, recognizes the linear motif PxxP – two *proline* (P) amino acids separated by two random amino acids.

The scientists found other ways to narrow the search. "We began with sets of proteins known to interact, taken from studies of yeast, flies, worms, and humans," Victor says. "After eliminating domain binding, we threw out partners that were clearly related through evolution." This was important, he says, because related proteins share many so common sequences that everything might look like a motif. The next step was to focus on particular regions of the target proteins where linear motifs were likely to appear – they rarely occur in domains, for example. After eliminating these regions of the molecule, the scientists searched what remained for recurring patterns that were three-to-eight amino acids long.

"When we had a candidate, we calculated the chances that it was random versus the possibility of it being meaningful," Victor says. "A common motif *RxPxxP* occurs randomly in about one in 20 proteins. If you suddenly find it



in seven of nine proteins that you've examined, it is almost sure to be significant."

The method revealed several promising patterns – but were they real? Rob and Victor tested the results against a thoroughly studied sample set called the *Eukaryotic Linear Motif resource* (ELM), developed by EMBL group leader Toby Gibson. "The motifs listed in this database have been confirmed through experiments," Rob says. "So it provides a good benchmark to compare with our data."

Hundreds of motifs remain to be discovered, and this method can be used to explore other types of interactions involving short motifs, such as transcription factors' binding sites.

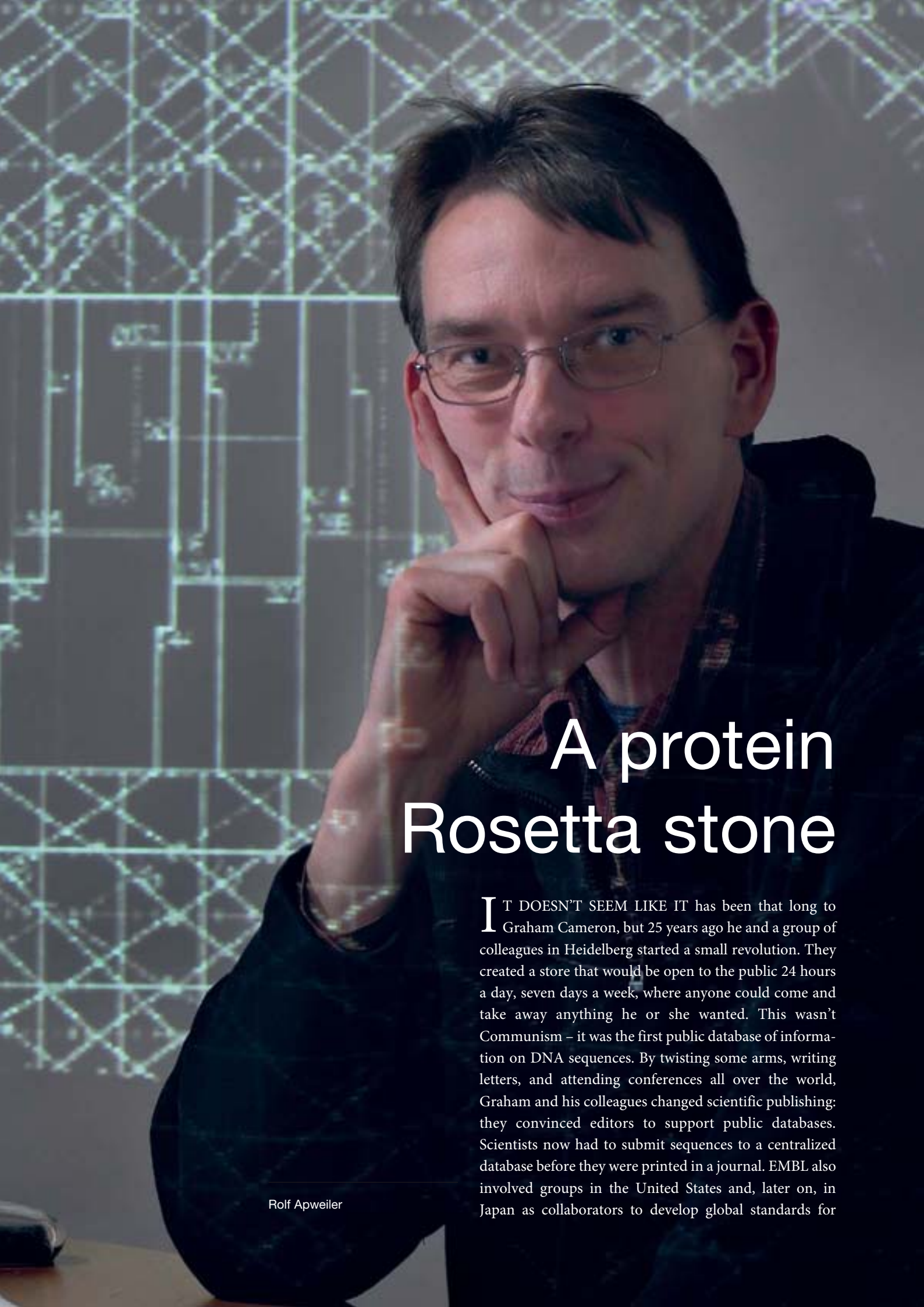
They also found a new motif in an important molecule used in all types of eukaryotic cells. *PP1* is an "interaction hub," Rob says; it acts as a sort of universal tool to strip *phosphate groups* from other molecules. It does so by working with many other "adaptor" proteins that fit it to specific targets. "PP1 was already known to contain the linear motif *RVxF*," Rob says. "We found a second linear motif, present in some species of PP1 that don't have RVxF, and this will help us explain its binding to even more proteins."

From ELM, 22 sets of interactions fit the criteria for Victor's and Rob's method. The program successfully found 14 of them, encouraging the scientists to take on a much larger set of data: information from genome-wide studies of protein binding. They discovered dozens of likely new binding motifs. They decided to test several of them in the laboratory.

"In most of these cases," Rob says, "we could show that the motif is responsible for, or contributes to, the molecule's ability to bind." For example, a motif *VxxxRxYS* binds to the protein *translin* from *Drosophila*. Little is understood about this molecule beyond the fact that it binds double-stranded DNA and RNA, and plays some role in the rearrangement of chromosomes. The new motif provides some hints about new interaction partners, and thus new insights into its function.

The scientists will continue to refine the method, using new data sets, such as the yeast proteome project (previous story). Rob believes that hundreds of motifs remain to be discovered. And the method can be used to explore other types of interactions involving short motifs, such as sites in DNA that transcription factors bind to, or microRNA targets.

"These patterns are very interesting from an evolutionary point of view," he says. "They are so short that they're likely to arise quite often through random mutations. A lot of molecules contain near-misses to the motifs, as if they're just waiting for a mutation to come along and turn them into functional sequences. They are powerful switches for nature to explore during the evolution of complex functions." ■



Rolf Apweiler

A protein Rosetta stone

IT DOESN'T SEEM LIKE IT has been that long to Graham Cameron, but 25 years ago he and a group of colleagues in Heidelberg started a small revolution. They created a store that would be open to the public 24 hours a day, seven days a week, where anyone could come and take away anything he or she wanted. This wasn't Communism – it was the first public database of information on DNA sequences. By twisting some arms, writing letters, and attending conferences all over the world, Graham and his colleagues changed scientific publishing: they convinced editors to support public databases. Scientists now had to submit sequences to a centralized database before they were printed in a journal. EMBL also involved groups in the United States and, later on, in Japan as collaborators to develop global standards for

storing, describing and exchanging the information. Thus the EMBL-Bank was born, and it threw open the flood-gates of biological data. The database was useful because it was universal, like a party that everyone goes to because everyone else is going.

EMBL-Bank was soon followed by SWISS-Prot, a database collecting everything that was known about proteins. SWISS-Prot has always been synonymous with two names: Amos Bairoch, the Swiss bioinformatician who invented it and personally approved every protein entry (a superhuman feat that he managed until quite recently), and Rolf Apweiler from EMBL, known for his skills in linking databases.

Amos and Rolf remain the principal investigators in the project, and neither of their names has changed in the meantime, but some of the others have: the database department of EMBL evolved into the EMBL-European Bioinformatics Institute in Hinxton, and SWISS-Prot has become part of the Universal Protein Resource (UniProt), an immense project supported by the US National Institutes of Health, the European Commission, the Swiss government and others. UniProt aims to link protein sequence data to everything that is known about the molecules – a huge challenge. In addition to the sheer amount of data, the task involves so many sources of information that members of the group sometimes feel they are trying to link everything in the universe to everything else.

The problem has increased by several orders of magnitude given today's high-throughput methods to analyze entire proteomes of cells. "We're experiencing the same type of boom in proteomics that happened with microarray experiments just a few years ago," Rolf says. "The literature was being flooded with an unbelievable amount of data about gene activity. To cope with that, Alvis Brazma and his team built the ArrayExpress database, and it was essential to develop at the same time a standardized language to describe results and the way the experiments are done."

Without such a language, scientists lack a "Rosetta stone" by which to compare their experiments. Hiding in millions of pieces of data created by a single experiment may be something that will help another scientist interpret his or her own work, but only if they can pull off a feat more difficult than translating hieroglyphics into ancient Greek. That problem was solved in the early nineteenth century by Jean-François Champollion, using the original Rosetta stone, and now Henning Hermjakob's Proteomics Services Team is helping to solve an analogous problem for proteomics data, with PRIDE, a new *Proteomics Identification Database*.

"A result from an experiment becomes meaningful if the right information is attached to it – telling what species it belongs to, what tissue it comes from, and often the location in the cell where it has been found," Rolf says. "There are all sorts of other features that may be described: the chemical changes it has undergone since it was translated, or the specific disease conditions under which it was discovered. If this information is the product of a mass spectrometry experiment, a huge amount of technical data needs to be attached to the results before anyone else can hope to reproduce it to confirm the results – too much information to include in a paper when researchers send off their results to a journal."

PRIDE was born out of a recognition of this problem, and like ArrayExpress, it is being developed alongside a standard way of describing the data that are used to populate it. Ideally, the result will be a way to capture data directly from a mass spectrometry machine, bundling it up with information describing the experiment, the sample, the technology used and the results, and forwarding it to PRIDE.

The standards are being tested on some huge sets of data submitted by two international projects running under the auspices of the Human Proteome Organisation (HUPO). The Plasma Project aims to capture a description of all of the proteins in plasma – the fluid that contains blood cells. The other is a profile of the proteins found in platelets – the fragments of white blood cells that

Graham Cameron (left) and Henning Hermjakob, facing off at EMBL-EBI

are responsible for clotting. Soon the database will contain a third set, from the HUPO Brain Proteome Project.

Researchers can query the database to discover whether a particular molecule has been found in serum, platelets or the brain, in a specific region in cells, or in association with a disease. "You can also be directed to all the experiments in which a specific protein was found," Rolf says. "From there you can follow links to explore what else is known about the molecule."

* * *

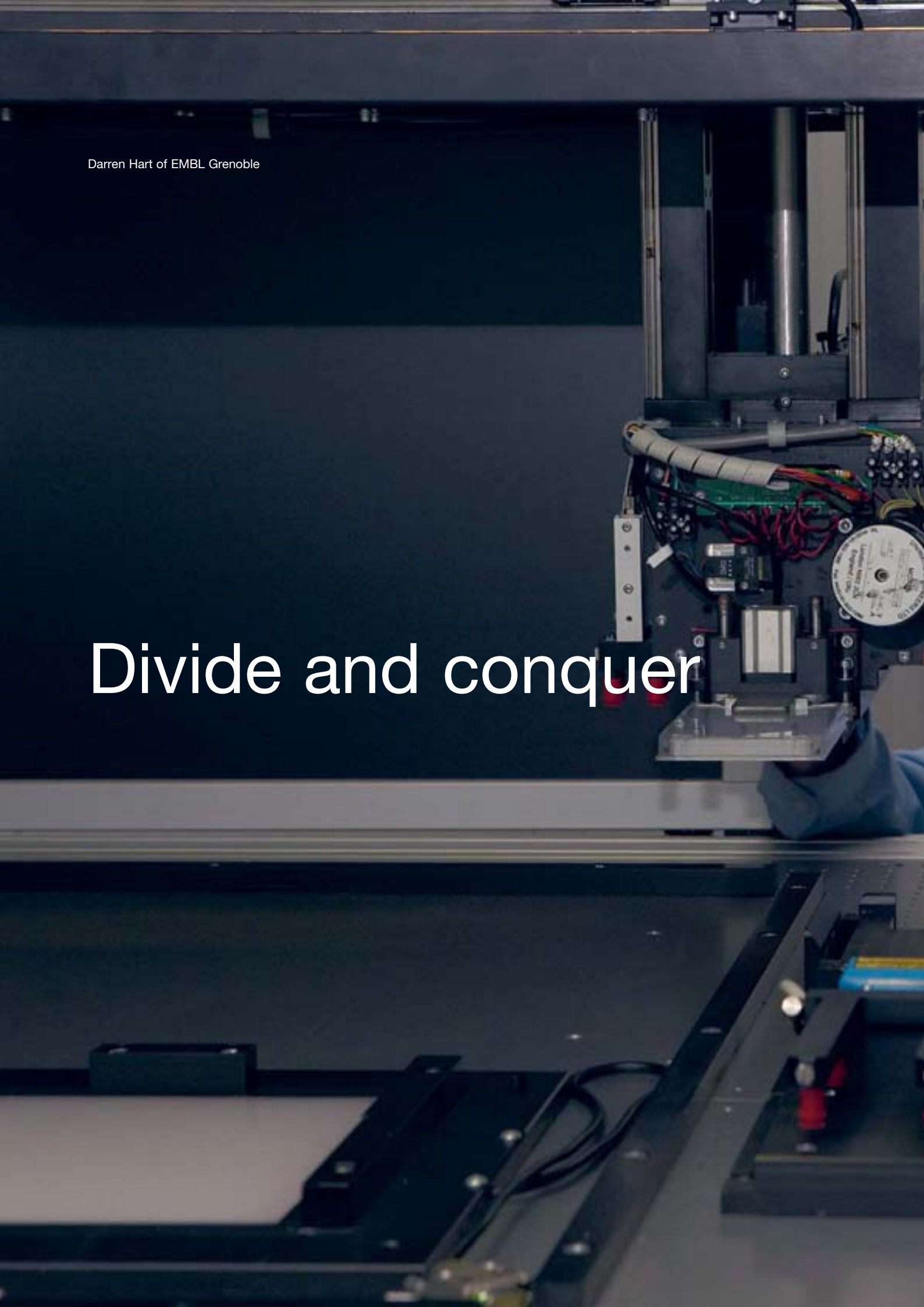
PRIDE is one component of a palette of tools that Rolf and his colleagues are developing, aiming to give researchers as much information as possible about the proteins they are studying. EMBL-EBI's dream of a "one-stop shop" for molecular information is moving ahead – yet it is a constant challenge, because scientists keep inventing methods that deliver massive piles of wares to the door, without coming in to arrange them on the shelves – often without even saying what they are.

The closest thing at present to a universal information desk for proteins is another project overseen by Rolf, a part of UniProt called the Knowledgebase. This combines the depth of knowledge about proteins captured in Swiss-Prot with the breadth of information generated by computer-based predictions about proteins' functions based on analyses of genomes and comparisons between species. Any information that has been derived purely by computer is checked by a curator, either in Rolf's group at EMBL-EBI, Amos Bairoch's group at the Swiss Institute of Bioinformatics, or the Protein Information Resource group at Georgetown University in the United States.

A Knowledgebase entry links to information about a protein's function, its structure (if known), the molecules it interacts with, any alterations it undergoes after it is translated... "It's not an endless list, but it's quite long," Rolf says. "It's a central hub for information derived from more than 60 databases which are themselves cross-referenced."

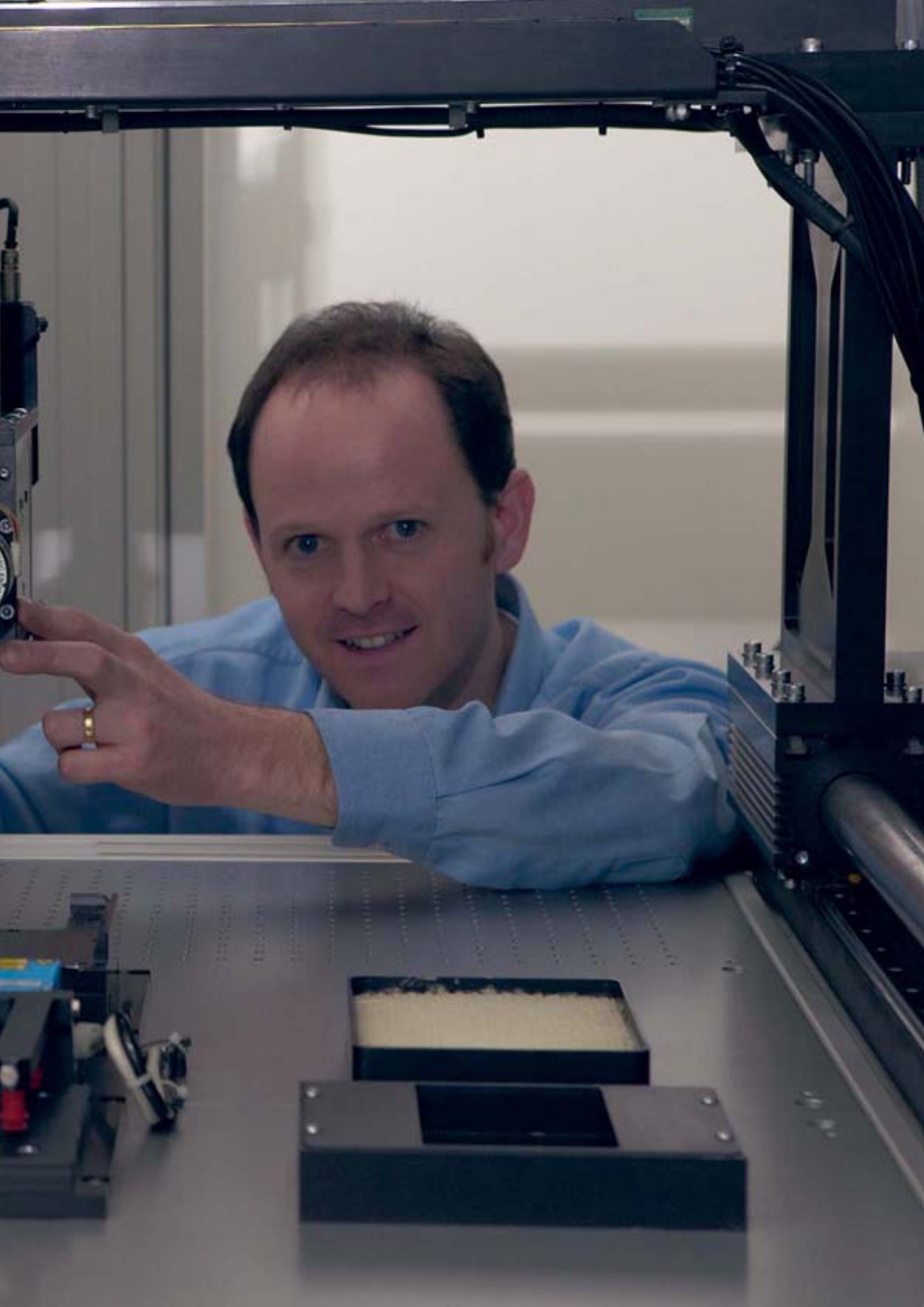
He calls up a scheme of the connections on his screen. In spite of the fact that there are far too many criss-crossed lines for the layman to make any sense of, it's obvious that there is a logic and an order to things, and if Rolf had the time, he could explain them all. Still, it's not the kind of knot you would want to have to untie.

The chart reminds me, in fact, of some other tremendously complex diagrams I have seen, of the biochemical reactions that take place in cells. Maybe it shouldn't be surprising that the networks needed to understand life might have to become nearly as complex as life itself. ■



Darren Hart of EMBL Grenoble

Divide and conquer



ESPRIT instrumentation (left) is housed in the new CIBB building in Grenoble (lower right)

IN THE LATE NINETEENTH CENTURY, chemists were embroiled in a debate every bit as heated as the controversy over evolution. *Materialists* believed that life obeyed the same chemical and physical rules as non-living matter; living systems ought to be explainable by understanding their components and the rules by which they operated. *Vitalists*, on the other hand, held that an extra force – likely supernatural – was necessary to animate non-living matter.

Molecular biology is the heir of materialism, and the goal of structural biology is to expose the physical and chemical properties of proteins and other molecules, in hopes of understanding how biological processes function, and how they break down during diseases.

Today the most common technique used to reveal protein structures is X-ray crystallography, carried out at synchrotron beamlines like those operated by EMBL teams in Hamburg and Grenoble. Groups at the Outstations and in Heidelberg are involved in several Europe-wide initiatives to improve and speed up the process of obtaining molecular structures. They have helped standardize methods and technology, vastly streamlining the process of obtaining highly purified proteins and other molecules, transforming them into crystals, and analyzing their structures. The ultimate goal is to understand complete systems by solving a large number of structures. One focus in Hamburg is the tuberculosis bacteria, and Grenoble groups are working on the components of several viruses.

Last year's report described the structural "pipeline" that has been established in Hamburg, and this year saw the integration of several facilities into a common building in Grenoble. The Carl-Ivar-Brändén Building (CIBB) was inaugurated on the site in January, 2006, to host common facilities for groups from the EMBL, ESRF, ILL and the University of Grenoble. Groups from these institutions have been carrying out collaborative projects within the Partnership for Structural Biology (PSB) for several years.

But for many scientists participating in high-throughput structural genomics projects, the focus is changing. These collaborations have often aimed to obtain as many protein structures as possible – for example, projects have been launched to structurally characterize large proportions of the human proteome – and the quest for numbers means less attention is paid to individual candidates. Those that are troublesome may simply be shelved in the pursuit of quantity.

"Indeed, the structural genomics pipeline appears to run reasonably efficiently for the 'low hanging fruit,'" says Darren Hart, of the Grenoble Outstation. "But problems arise when applied to more challenging targets, including many eukaryotic and viral proteins, complexes and membrane proteins – often the hot proteins we are most interested in. A significant challenge is producing and purifying proteins at a yield and quality appropriate for structural studies. Notably, during the early phase of a project, the design of gene constructs for soluble protein expression in *Escherichia coli* is unreliable. When it works, it works well, but it often proves difficult or even impossible."

Three years ago, Darren developed a new method for generating successful constructs when conventional

approaches fail. Now his approach has become one of the steps in the Grenoble crystallography pipeline.

* * *

There are different ways to create variations of a protein that may improve how it is produced by bacteria. Mutations can work quite well with small proteins or single domains of a molecule. "But there are other situations where it is more effective to take a protein and make simpler versions by shortening it," Darren says. "This is an appropriate strategy if the molecule contains several domains, or if it doesn't fold properly when produced in bacteria. Truncating may also help produce molecules that exists as part of a complex in which the other partners stabilize its structure."

Usually scientists have produced these fragments after inspecting a molecule's sequence and trying to predict where its domains might be. Sometimes this "rational" method works, Darren says, but not always. "Another approach is to emulate evolution – generate random genetic diversity and see what emerges as the fittest construct. In this way, dependence on tricky rational prediction is avoided. We use enzymes to chop up genes randomly in the test tube and make tens of thousands of constructs at a time. The burden is then shifted to identifying clones that make a useful version of the molecule in *E. coli*."

"Another approach is to emulate evolution – generate random genetic diversity and see what emerges as the fittest construct."

For this he uses high-throughput screening robotics developed originally for genomics projects. "Only about one in a thousand of these constructs produces protein, but it does so pretty well," he says.

When in France, find an acronym that works well in French – so Darren has called the technology *ESPRIT* (for the Expression of Soluble Protein by Random Incremental Truncation). It was developed as part of SPINE (Structural Proteomics in Europe) and patented by EMBLEM. ESPRIT is now being used in collaborations with EMBL groups in Grenoble, Heidelberg and Hamburg as well as other European labs and a pharmaceutical company. Recent successes include high-level expression of several proteins from the influenza virus that had seemed impossible to purify and crystallize. One of these has resulted in a structure determination. ■

Santosh Panjikar at
EMBL Hamburg

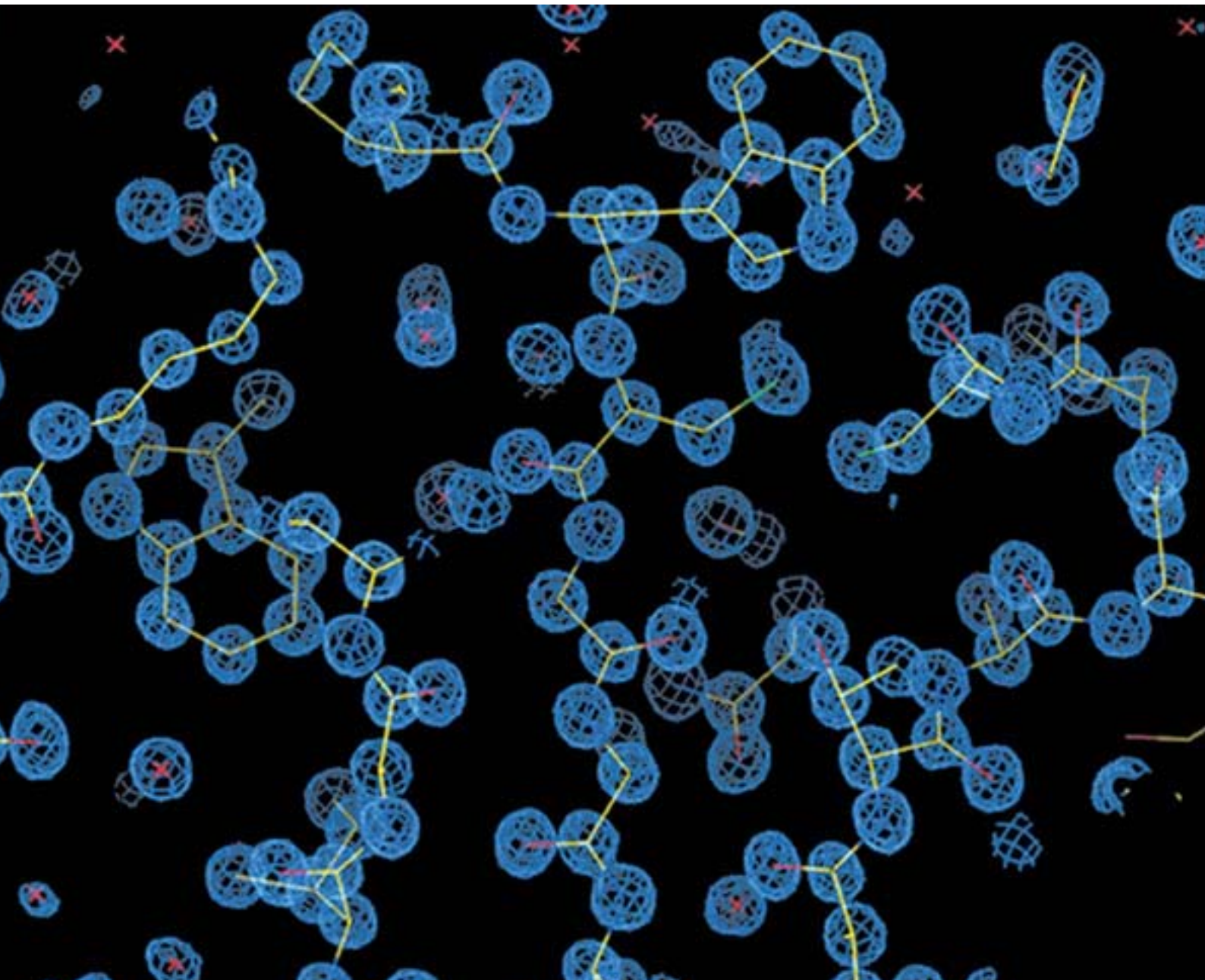
Instant structures

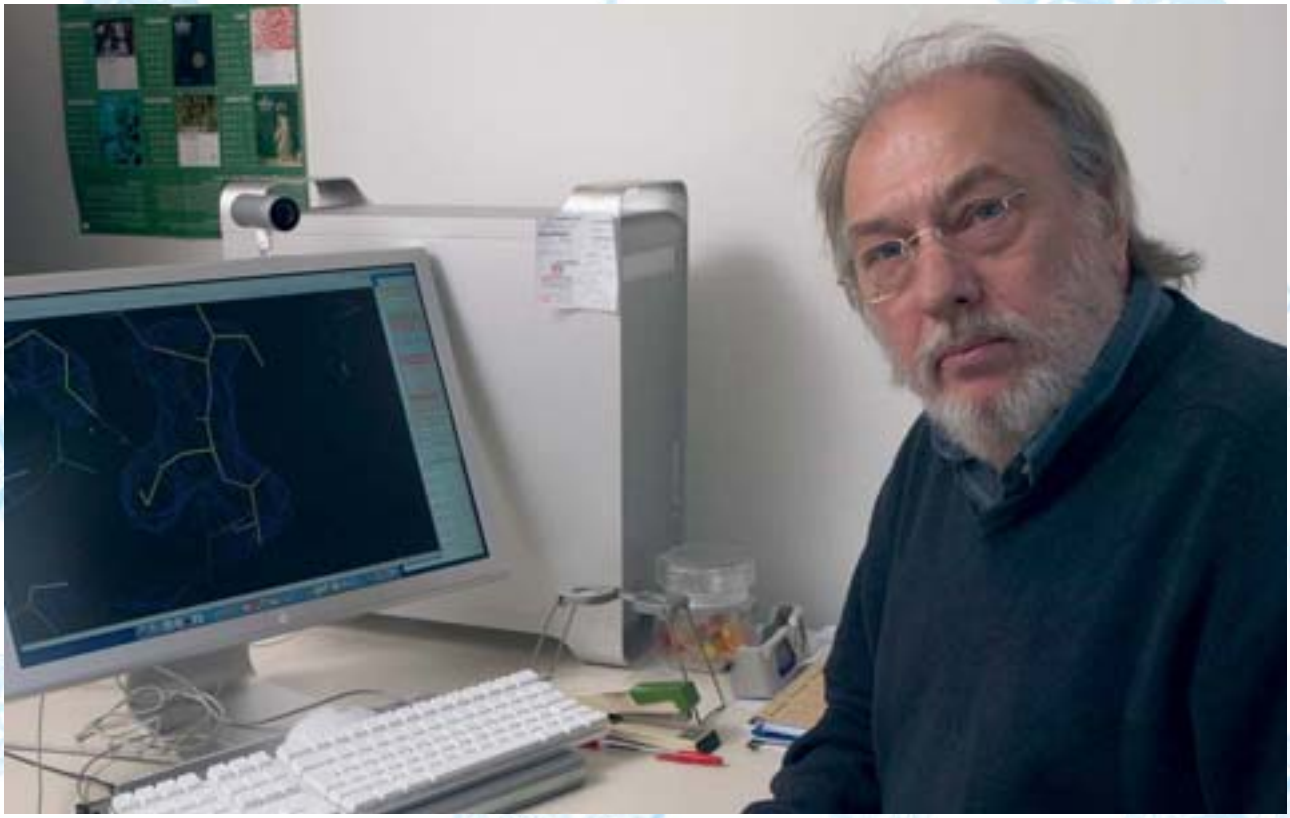
WHEN AN X-RAY HITS A PROTEIN CRYSTAL, 99% of the beam blasts straight through. Nine-tenths of the remaining one per cent are absorbed by atoms, and the last bit – one-thousandth of the original radiation – strikes electrons and is deflected. Caught on a detector, like those on the beamlines at the Hamburg Outstation, it forms a diffraction pattern which can be reinterpreted into a three-dimensional structure of what's in the crystal.

Doing so requires experience and expertise. In the 1950s, piecing together the first protein structure required years of work, carried out by John Kendrew's huge team of human "calculators". Things have improved a lot: with the help of software packages like Victor Lamzin's ARP/wARP, an expert can now solve a structure in a few hours. That's still not fast enough, says Paul Tucker of Hamburg, in the days of high-throughput structural genomics. So Santosh Panjikar and other members of Paul's team have been trying to turn some of their expertise into code and load it into the computer. The result is a new software platform called *Auto-Rickshaw*.

"This system emulates the decision-making and judgment that a crystallographer brings to the handling of data," Santosh says. "There are many different software packages available to analyze structures, and typically at each phase in the development of a structure we step in to make decisions about which is best to use. Another task is to adjust the parameters of each of these programs to fit the data. That takes time."

Usually this means that the interpretation of a structure takes place after an experiment has been carried out. With the development of high-throughput "pipelines", aimed at obtaining as many structures as possible, this creates a problem. A scientist may not know whether the quality of the data is high enough to obtain a structure; if it isn't, the experiment may have to be repeated. Ideally, results could be determined immediately – at least a preliminary judgment of the quality of the structure.





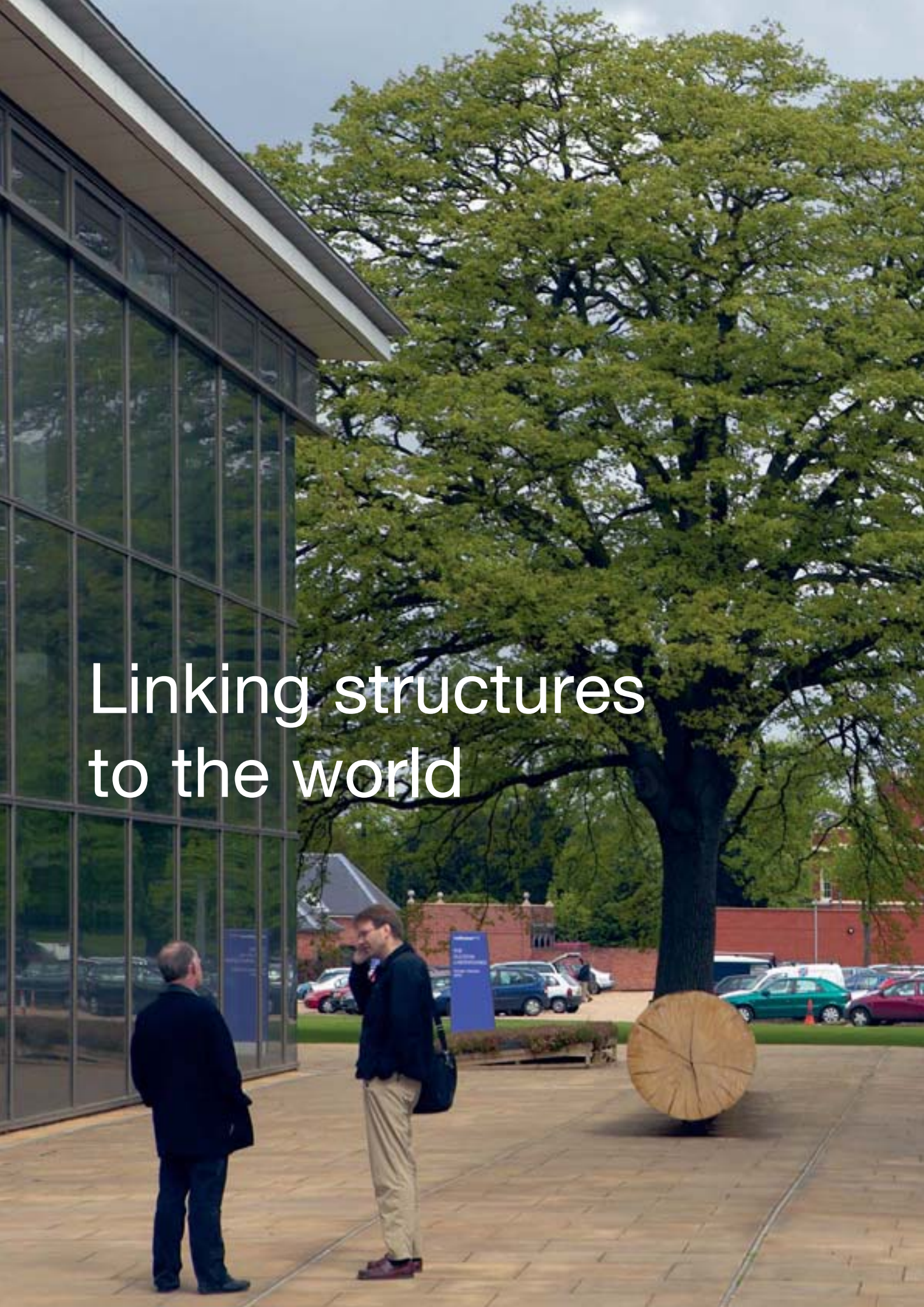
An electron density map (left), not to be confused with EMBL Hamburg's Paul Tucker (above), although he does possess quite a lot of electrons.

Santosh has sped up the beginning of the analysis – now launched as soon as data starts being collected – and automated some crucial decision-making steps. "X-rays interact with electrons, so you see a lot of interactions with atoms that have a lot of electrons, the so-called 'heavy atoms'," Santosh says. "The first step in an analysis is to find the positions of these atoms. That helps us to calculate the electron density and determine the atomic structure using a combination of mathematics and knowledge of the chemical makeup of the protein."

The main goal of Auto-Rickshaw is to obtain this map very quickly and propose a partial structure to confirm that the experiment has been successful. It does so by examining the data and using decision-making routines to choose which type of software to use for each experimental phase. The scientist can watch its progress on a web browser and intervene if needed. If everything goes smoothly, and the data is of sufficient quality, the experiment can be stopped with the confidence that a structure can be produced. If not, the program can suggest various things to try.

Santosh and his colleagues have developed two versions of Auto-Rickshaw: a version for use directly at the beamlines, and another which can be used to create more refined structures from the data at a later stage. The second version relies on the ARP/wARP structure-building program designed by Victor Lamzin and his group.

The platform was tried out on 14 test and 11 real experiments. In each case, scientists obtained a useable electron density map. Now Auto-Rickshaw has been installed on computers at the Hamburg beamlines, giving users and staff a way to validate the results of their experiments as quickly as possible. Santosh and other members of Paul Tucker's team continue to make improvements, adding on other programs that can work in parallel and improving the decision-making process. Santosh aims to reduce the amount of input data that is required and make the platform into a "self-learning" system that can profit from its own experience. ■



Linking structures to the world

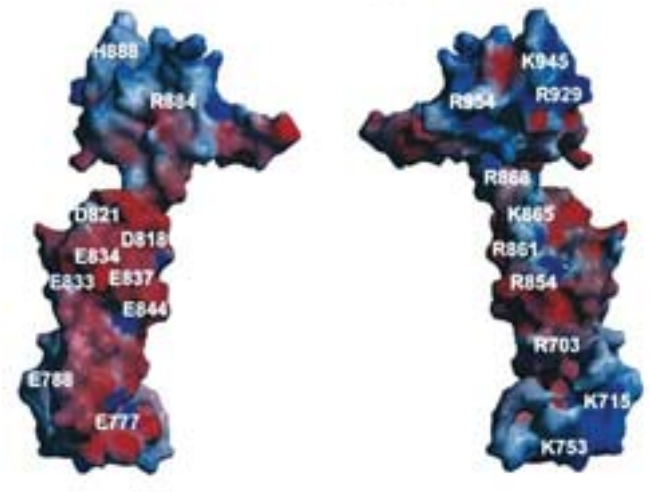
Kim Henrick (left) talks to
Heikki Lehväslaiho at EMBL-EBI

THEY LOOK LIKE WHEELS, spikes, tongs or multi-colored balls of yarn that your cat has played with for far too long. They have heads, tails, and fingers. They engage in digestion, communication, walking, packing and towing. And they have a portrait gallery: the Macromolecular Structure Database (MSD). Kim Henrick's MSD team at EMBL-EBI is one of three pillars supporting the worldwide Protein Data Bank (wwPDB), a collaboration of groups in the USA, Europe and Japan. Their goal is to gather, maintain and distribute the world's collection of large molecular structures.

The MSD is far more than a portrait gallery – it provides a means for scientists to get to know the characters behind the portraits – to discover the physics and chemistry behind what a protein does, why a mutant form does something different, and how that behavior might be affected by a drug. The MSD is growing by leaps and bounds. Kim's team and their collaborators have developed tools to make it as easy as possible for scientists to submit their structures to the collection, while keeping an eye on the quality of the submissions. In the process, they add value to the data.

"What you get from a structure doesn't always match what a researcher would like to know," Kim says. "Automated methods of structure determination usually provide a snapshot of a single protein chain, but in the cell proteins usually work in pairs or larger groups. One thing we've done is to develop tools to make predictions about the complexes that proteins form. This allows researchers to visualize protein structures as they might look in real life."

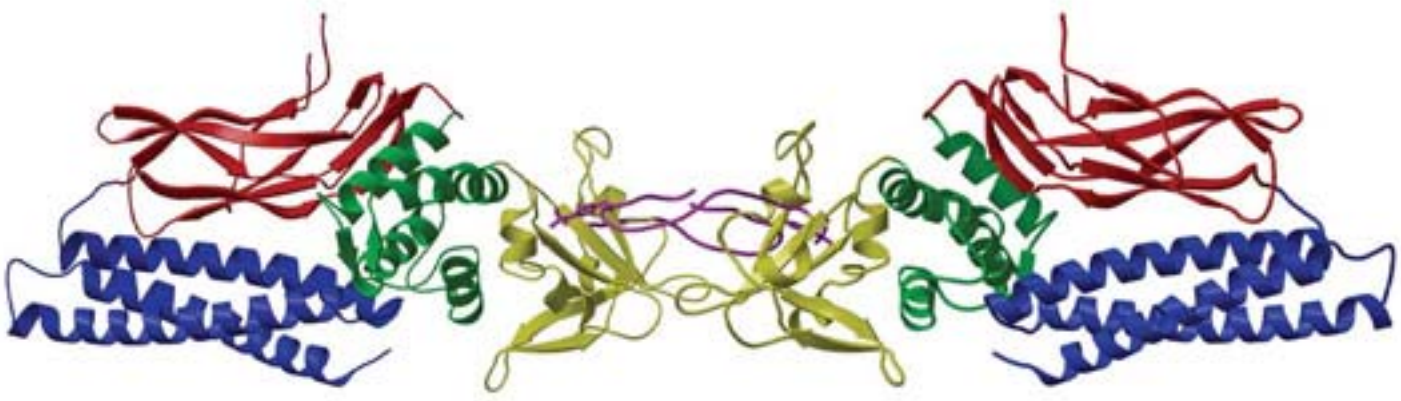
In 2005 Eugene Krissinel and other members of the group introduced a new tool called MSDpisa; pisa stands for *protein interfaces, surfaces and assemblies*. "This allows users to explore the structure and chemistry of protein surfaces, to study interfaces and predict how a given molecule will bind to a complex," Kim says. "We've calculated some of these things for structures already in the archive, and users can call it up; alternatively, they can upload their own structures and scan their structures with MSDpisa."

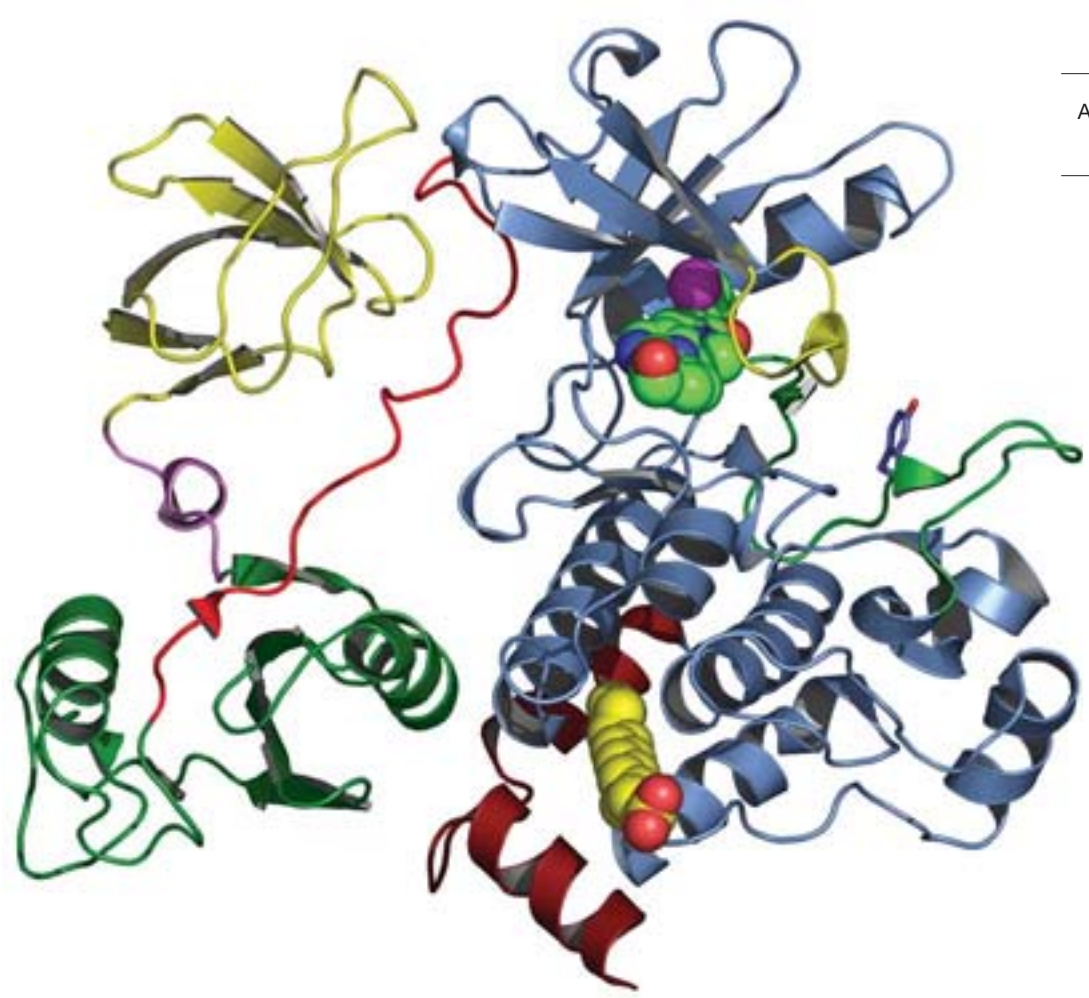


By doing so, they can learn about the biological role of the interface – for example, it may work as a landing platform for a neurotransmitter, or it may have "fingers" to slide into the grooves of the DNA helix.

"One thing we've done is to develop tools to make predictions about the complexes that proteins form."

Another task of Kim's group is to link a structure to information contained in other databases, such as UniProt (see page 106). Such connections may help users find information about diverse forms of the same molecule – for example, structures of mutant forms – that can give insights into its functions.

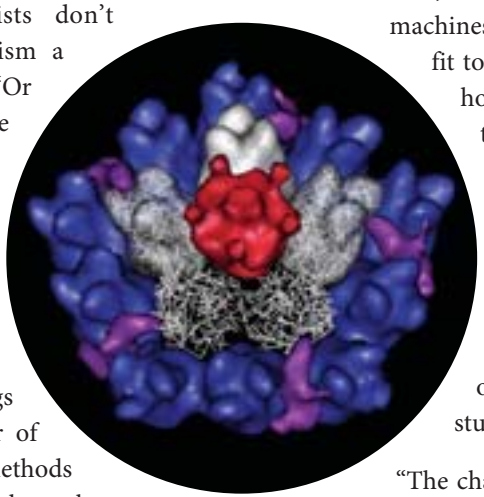




A gallery of structures from the EMSD

Comparing variations in the surfaces may explain why a protein no longer snaps into a molecular machine. Sometimes the links answer far more trivial questions.

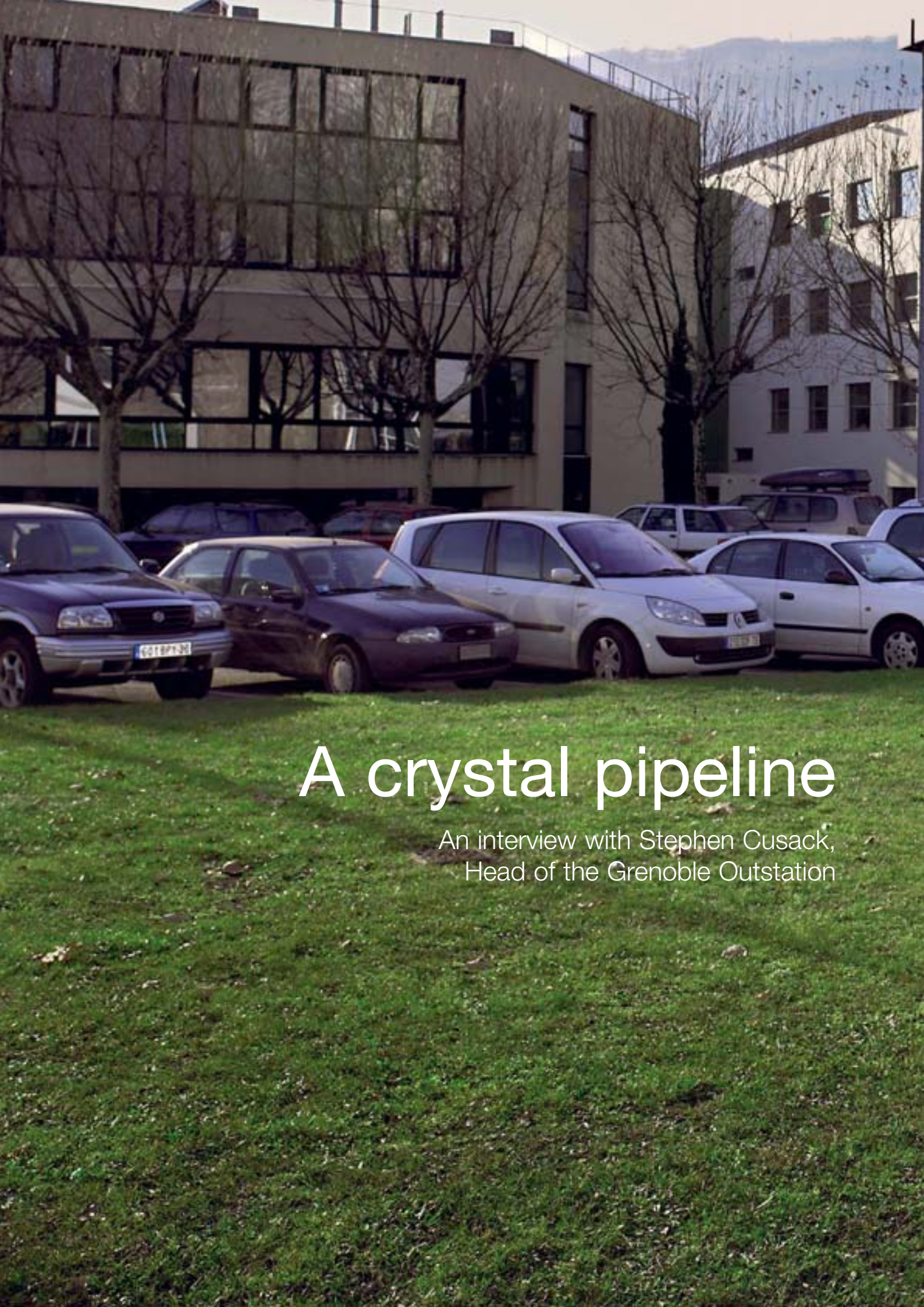
"Amazingly enough, experimentalists don't always clearly indicate what organism a molecule has come from," Kim says. "Or they may provide a Latin name for the organism, and have misspelled it. That obviously will cause problems for a database."



Kim's team helps curate the entries, checking them for accuracy, comparing new structures to molecules already on hand, and cleaning up things like misspelled Latin. As the number of submissions grows, thanks to better methods for obtaining structures and high-throughput structural genomics projects, the team has developed algorithms to automate the process as much as possible.

Recently the database has been making room for lower-resolution data from sources other than protein crystallography, such as electron microscope studies of protein complexes, or small-angle scattering experiments (see story on page 20). Integrating this is a challenge because it is image data, rather than the standard forms of vector data – plots of points – that have been used to represent protein structures. But images provide the link to the level of the cell, revealing not only what molecular machines are made of, and how their pieces fit together, but where they operate and how they change as they carry out their functions. Projects like the yeast proteome study of Anne-Claude Gavin and her colleagues (page 98) have produced vast amounts of information that need to be linked to structural data; the same has to be done with fuzzy objects revealed by microscope studies.

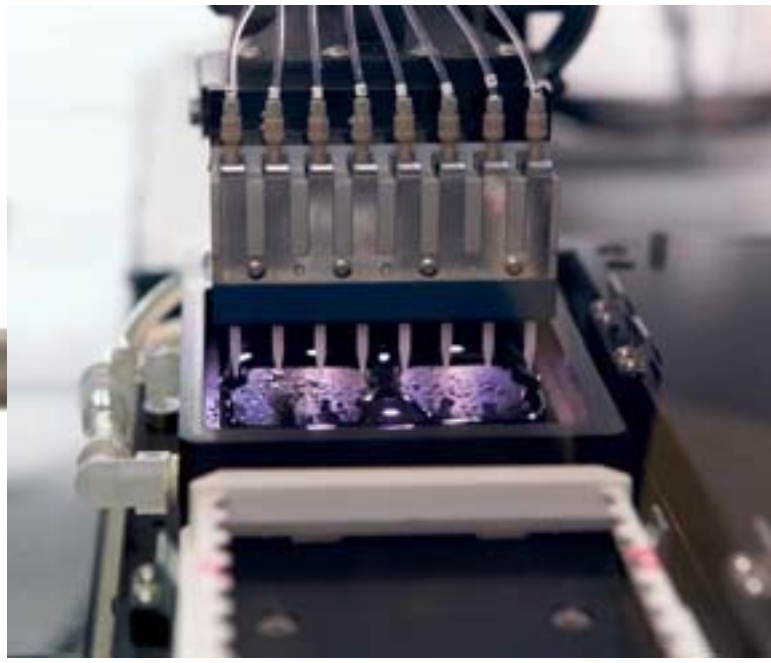
"The challenge over the next few years will be to integrate structures of single molecules with their roles in dynamic machines and at the level of the cell," Kim says. "This will mean bringing images and other new types of information into the MSD and linking the database more closely with other types of biological information. For example, we are investigating MSD structures with information on chemical compounds, such as drug libraries, and adding links to structures of alternative splice forms of proteins." ■



A crystal pipeline

An interview with Stephen Cusack,
Head of the Grenoble Outstation

Stephen Cusack and Rob Ruigrok in front of the new CISB in Grenoble



Steps in the crystallization pipeline: Josan Marquez and Darren Hart in front of ESPRIT robotics; high-throughput nanovolume crystallization; Florent Cipriani and Franck Felisaz developing automation in the loading and handling of crystals on beamlines.

What are the components of the pipeline that have been assembled in Grenoble, particularly in connection to the new facility?

In Grenoble we have never really aimed at having an A (gene) to Z (three-dimensional structure) automated pipeline for structure determination in the style of big structural genomics projects. For a start we do not have the resources to do this and secondly it is not appropriate to the kind of projects we want to do, where we focus on a few important proteins and not a long list of targets, many of which will fall by the wayside. Thus our approach has been to use automation and high-throughput methods selectively where they really bring something useful and combine them with more traditional, labour intensive methods of structural biology. Moreover, in collaboration with our PSB partners we have tried to make available the whole panoply of technical platforms required for today's structural biology, building on the complementary skills of the partners. In particular the excellent nuclear magnetic resonance, cryo-electron microscopy and mass-spectroscopy facilities of the Institut de Biologie Structural (IBS) are now extensively used by EMBL scientists. The three areas that we have developed most at EMBL, within the context of the PSB and SPINE (see below), are high throughput screening for soluble constructs of otherwise insoluble proteins (ESPRIT platform of Darren Hart), high-throughput nanovolume crystallisation (Josan Marquez) and automated data collection on the ESRF synchrotron beamlines (Florent Cipriani, Raimond Ravelli, the ESRF and the MRC-France).

What's new or extraordinary about single parts or the way they have been brought together?

The ESPRIT platform developed by Darren Hart is a truly novel development that is attracting a lot of attention. Everybody is faced with the problem of not being able to express interesting proteins in *E. coli* in soluble form, particularly multi-domain eukaryotic proteins, and there can be many reasons for that. However, many people find that with the right construct it can work. ESPRIT is a systematic approach to finding the right construct and indeed most targets that have been tried yield soluble fragments, which gives a foothold onto the problem. A good example is our joint project on the subunits of the influenza virus polymerase. It does not mean that all these fragments crystallise of course; that would be too easy!

The high-throughput crystallisation platform is based around a Cartesian nanodrop dispenser and a Robodesign storage and imaging system. It has been an incredible success, thanks to the technically proficient and user-friendly service set up by Josan Marquez and his team. During the last year more than 170 researchers

from the five PSB-associated institutes (EMBL, Institut de Virologie Moléculaire et Structurale, IBS, ILL and ESRF) benefited from access to our high-throughput crystallisation platform. The throughput is stabilizing at close to 400.000 experiments (i.e. individual drops) per year. A large part of the current work is focused on upgrading the data management system and developing a new web-based interface giving access to the results of crystallisation experiments. This interface, which is now fully functional, will soon be shared with the Hamburg crystallisation facility. During this time we have also formalized our collaboration with the Protein Management Information Project (PIMS) whose aim is to create Europe–wide standards for data storage and retrieval through all the experimental levels in structural biology and to which we will be contributing our expertise and some of the software we develop for our platform.

Concerning automation of frozen crystal sample handling and data collection at the ESRF beamlines, EMBL scientists and engineers have been at the forefront of an impressive series of new developments including sophisticated and highly accurate automated diffractometers, especially suited to microcrystals coupled to a frozen crystal sample changer (this combination is now installed on all seven ESRF MX beamlines and the UK-EMBL CRG BM14), the European SPINE standard for mounting frozen crystals, automatic crystal centreing software using UV imaging and a mini-kappa goniometer. In collaboration with the ESRF and the UK operated MRC-France and with the support of EU integrated projects SPINE and BIOXHIT, all these developments are being incorporated into a global system linking the hardware to an expert system for automated data collection (DNA) and a tracking database (the Information System for Protein Crystallography Beamlines, ISPyB). Overall this will mean a revolution in how synchrotron beamlines are used but because these are extremely complex integrated systems it is taking time to get things working reliably.

Perhaps the unique thing about the Grenoble setup is that the PSB partners and the IVMS have built a new building,

the CIBB, adjacent and connected to EMBL, which houses members from all institutes as well as some of the platforms.

What will this particular setup allow people to do that hasn't been easy before?

Three quick examples. The ESPRIT platform has allowed identification of an independently folded domain of influenza virus polymerase PB2 subunit, resulting in the first structure of a fragment of this enzyme. The structure was actually done using NMR by Jean-Pierre Simorre of the IBS, as part of the PSB collaboration. The domain contains residues implicated in the transmission of influenza viruses between avian and mammalian hosts. Concerning the high-throughput crystallisation platform, there are several examples whereby the much reduced sample quantities required for a full screen have now resulted in crystal hits for very scarce proteins. Regarding automated data collection at the ESRF there have been several full-scale trials involving UK groups who have entered sample information into the ISPyB database from their own labs, sent pucks full of frozen crystals to the ESRF, which have then been automatically screened on beamline BM14 using the sample changer and DNA expert system. Fifty samples can be screened in 2.5 hours. The best samples are then lined up for proper data collection. The ability to rapidly screen a lot of samples is very important for small, not very good crystals (as is usually the case!) of very interesting proteins and complexes to find the rare one that diffracts better than average.

Who will it serve?

The different PSB platforms serve different communities depending on their capacity, state of development and available resources. Thus ESPRIT is open on a restricted collaborative basis, the high-throughput crystallisation platform is open to PSB members and associates as a service and the ESRF beamlines are of course open to anyone who applies and is approved by the selection committee. In the long run we hope to open more and more platforms to external users but this will require more resources.

What grants do you play a major role in and what are their goals?

A major driving force behind the development of high-throughput methods for structural biology in Europe was the SPINE project, one of only three Framework Programme 5 Integrated Projects, which started in October 2002. The SPINE project was Europe's belated answer to highly financed structural genomics projects established in the USA and Japan. But the special situation in Europe forced us to establish a more distributed, democratic mode of operation with no centralized protein production or crystallisation facilities. SPINE focused firstly on technical developments of high-throughput (with an emphasis then on making them available to the community at large), and secondly, right from the start, on high-value targets related to human health rather than low-hanging fruit from bacterial genomes. I coordinated the Grenoble node of SPINE,

About the CISB

The Centre for Integrated Structural Biology (CISB) is a unique infrastructure run jointly by the Partnership for Structural Biology (PSB): EMBL, ESRF (European Synchrotron Radiation Facility), ILL (Institut Laué-Langevin), the IBS (Institut de Biologie Structurale), and the Institut de Virologie Moléculaire et Structurale (IVMS).

The Carl-Ivar Brändén Building (CIBB) is the new CISB laboratory complex on the Polygone Scientifique Campus in Grenoble. The building was named after Carl-Ivar Brändén, the former Director of Research at the ESRF, and houses a complete pipeline for carrying out highthroughput structural biology.

which was the major integrative force to establish the collaborative scientific programme of the PSB. Thus our SPINE tasks included co-ordination of WP6: high-throughput synchrotron facilities, which provided the impetus for the SPINE frozen sample standard to be created and provided resources to the EMBL and ESRF for the sample changer and automated data collection projects. The development of both ESPRIT and the high-throughput crystallisation platform was also stimulated by SPINE. The Grenoble node also contributed to the target-orientated work packages related to human health, notably proteins from pathogens such Epstein-Barr virus (from the IVMS), kinesins involved in cell cycle and cancer (IBS) and neurobiology and immunology proteins (EMBL).

How will funding for these activities continue once the current external support expires?

EU Framework Programme 6 was rather generous to large-scale structural biology projects and in Grenoble we benefit from most of them including BIOXHIT, 3D-Repertoire and, as recently announced, the successor to SPINE, SPINE2-COMPLEXES, which will begin in July 2006 for four years. In addition the PSB-IVMS project (together called the CISB, Centre for Integrated Structural Biology) itself was supported by an FP6 grant under the Construction of New Infrastructures programme, which funds 10% of the new investment, including the CIBB and new equipment, up to 1.73 million Euros. ■

CISB platforms for protein and crystal production

P1: RoBioMol: high-throughput cloning and protein expression screening (IBS)

P2: ESPRIT: Finding soluble domains in proteins (EMBL)

P3: Protein characterisation: NMR, Mass-Spec, CD, ITC (IBS, EMBL)

P4: Isotope labeling for neutrons and NMR (ILL, EMBL, IBS)

P5: High-throughput nanovolume crystallisation (EMBL)

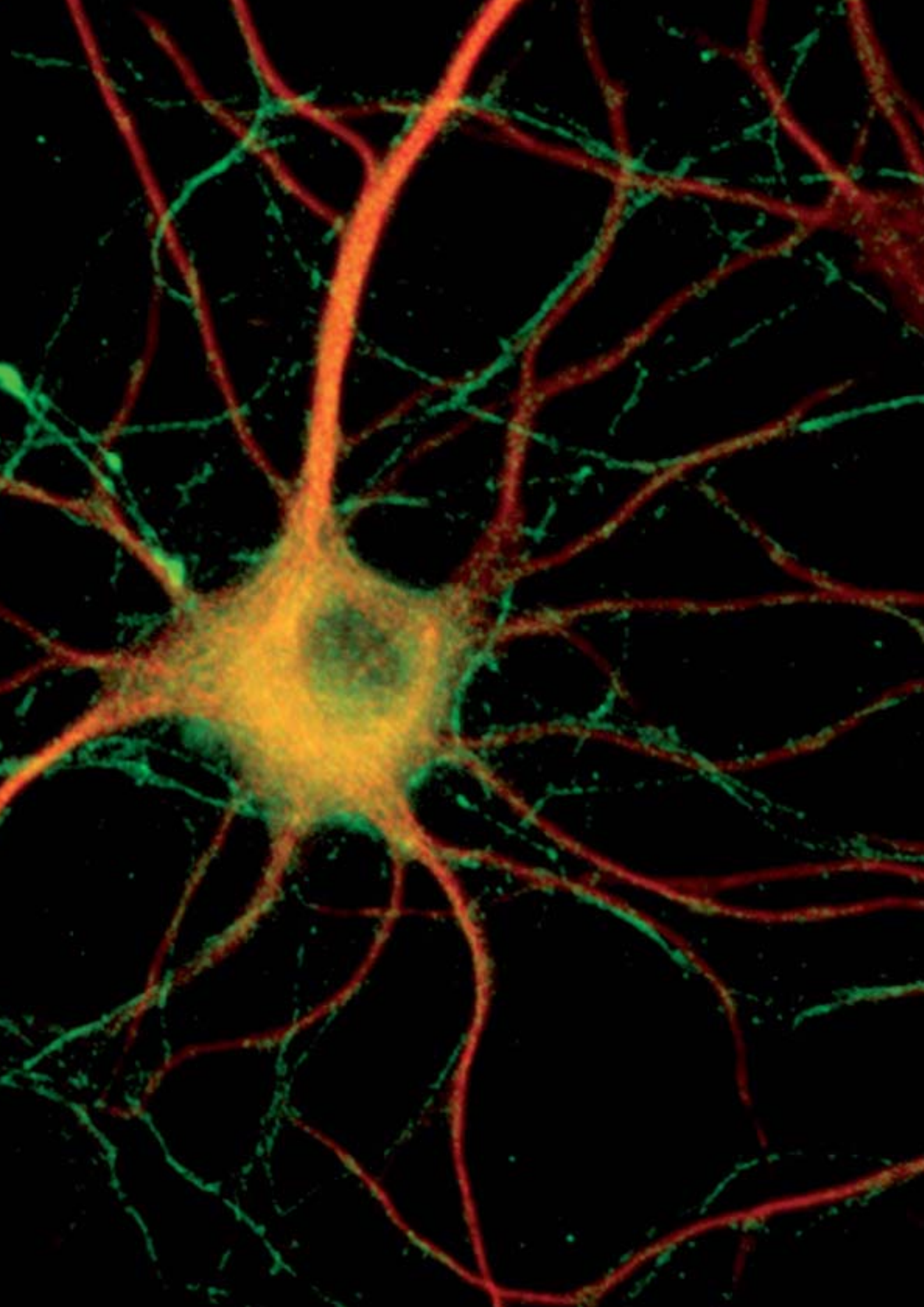
Structure determination

P6: Automated synchrotron X-ray data collection (ESRF, EMBL)

P7: Neutron crystallography and SANS (ILL)

P8: Electron microscopy (IBS, IVMS)

P9: Nuclear Magnetic Resonance, 600MHz + 800MHz (IBS)



Brains, models and systems

IN 1583, A JESUIT PRIEST NAMED MATTEO RICCI arrived in China to set up a Catholic mission. Along with European wonders like clocks, prisms and musical instruments, he brought along a map of the globe – astounding to a people who believed that the fifteen provinces of China comprised nearly the entire world – and an ingenious method for memorizing vast amounts of information. This system involved creating a mental image of a vast palace with hundreds of rooms. Things to be learned were placed in the rooms, and when they had to be recalled, the task was to imagine moving from room to room, opening doors and retrieving them.

Learning and memory somehow alter the structure of the brain, but its networks of tissues, cells and molecules are so complex that it is still impossible to find the physical basis of Ricci's rooms and the objects they contain. Scientists have begun to grasp how neurotransmitters and other molecules alter cell structure and behavior. But the system does not work purely from the bottom up. The behavior of molecules permits and directs high-level activity, such as thinking, but thinking likewise reorganizes the behavior of molecules.

Grasping the interplay of levels in this very complex system will require models which integrate mathematics and experimental science, genetics and psychology, computer science, imaging, and linguistics. But these fields have arisen through different impulses, to meet different needs, and they do not have a common language. In the 1980s, "network modelers" hoped to simulate large-scale behavior like learning and memory while ignoring some of the basic physical characteristics of synapses. While this caused biologists to grumble, it had an important function: pointing out the need to find common ground.

Other areas of biology are confronting the same issue. One story in this chapter describes experiments that aim to create something like an artificial organism in the test tube and simulated in the computer. The scientists limited themselves to trying to simulate one phenomenon: how an embryo's body develops a basic, segmented structure. Although the end product doesn't look much like a real animal, there is a much stronger resemblance at the molecular level. Genes have to behave in a particular way to create tissues and boundaries. There have been many hypotheses about how this happens, and the project has confirmed some of them.

Even when this problem is pared down to its most basic form – an investigation of how a few genes regulate each other – such processes are so complicated that they can only be approached through models, in a highly interdisciplinary way. At the moment, there is no guidebook to show how this should be done; it is being invented as we go along. Using human intelligence to investigate itself may not necessarily be a paradox, but it will certainly force scientists into new theoretical territories. As scientists build models, they must come to grips with fundamental questions about models themselves. ■

Hold that thought

IF OUR BRAINS NEEDED as much time to "unwrap" an idea as it takes to open a package, then reading a single sentence of this book might take longer than the age of the universe. Each of our thoughts requires the packing, transport and unwrapping of billions of molecules called *neurotransmitters*. One nerve cell envelopes them in membranes; then they are released and picked up by a neighboring cell. This happens at an interface between cells called the *synapse*, and the result is a signal, an impression, an idea. It happens so quickly that an impulse can travel an amazing distance in a fraction of a second.

Turning molecules into thoughts or instructions to the body, such as a command to press a key on a piano, requires amplifying the signal. A single cell may transmit an impulse to hundreds of neighbors, which pass it along to hundreds more. At the same time this activity has to be kept under control by tuning down signals. Otherwise chemical information would spread through the brain like a runaway wildfire, causing a massive electrical storm. There would be no way for the brain to control pain; muscles would lock into cramps; fine-tuning the nervous system in order to learn a piece by Chopin (or to unlearn mistakes) would be impossible.

How well a signal spreads is partly controlled by atoms of calcium drifting in the space between cells; these alter the receptivity of neurons by opening and closing channels in cell membranes, changing the release and uptake of neurotransmitters. This is related to other controls of the flow of stimulation, for example when cells snatch back some of the neurotransmitters they have released. They are retrieved in a process called *endocytosis*, which requires the cooperation of many molecules.

Winfried Weissenhorn's group at the Grenoble Outstation has now gotten a handle on a protein involved in these processes. What they've discovered may provide some insights into how the immense job of packing and shuttling neurotransmitters is managed by cells.

* * *

"Ejecting neurotransmitters from the tiny membrane packages called *vesicles* and retrieving the vesicles go hand in hand," Winfried says. "Cells often experience bursts of activity in which they secrete massive numbers of neurotransmitters. All of these molecules are first packed up into vesicles in the interior of the cell, and are then shipped off to the outlying regions for release. If this were an entirely one-way process, cells would soon run out of packing material. So endocytosis has an essential function: it brings back membranes for recycling."

How cells handle membranes and vesicles are fundamental processes involved in nearly everything the cell does, and these processes are central themes of Winfried's group. He's interested in neurotransmitters because they pose some unique "engineering" problems. Membranes are made of fat molecules; they are rubbery and can be bent. Neurotransmitter vesicles are small and have to be bent particularly sharply.

Several years ago, scientists discovered a protein called *endophilin* in neurons that seemed to be required for the retrieval of vesicles that have undergone endocytosis at the synapse. Cells that lacked the protein had problems with a certain type of packing used in endocytosis.

"When material outside has to be brought in, the cell membrane buckles inwards and creates a pocket-like structure," Winfried says. "This should pinch off to form a vesicle, but without endophilin, that usually doesn't happen."

He looks around for something to use to demonstrate. On the lab bench is a latex glove; Winfried blows into it, inflating it like a balloon. He grips the surface in his hand and pinches it to form a bubble. If this small segment separated from the glove and floated away, it would be something like a vesicle. "In the cell there would be proteins on its surface that tell it where to go," Winfried says. "As if it traveled to a glove over on the next bench, fused onto it, and whatever was inside the vesicle would be deposited into the new glove."

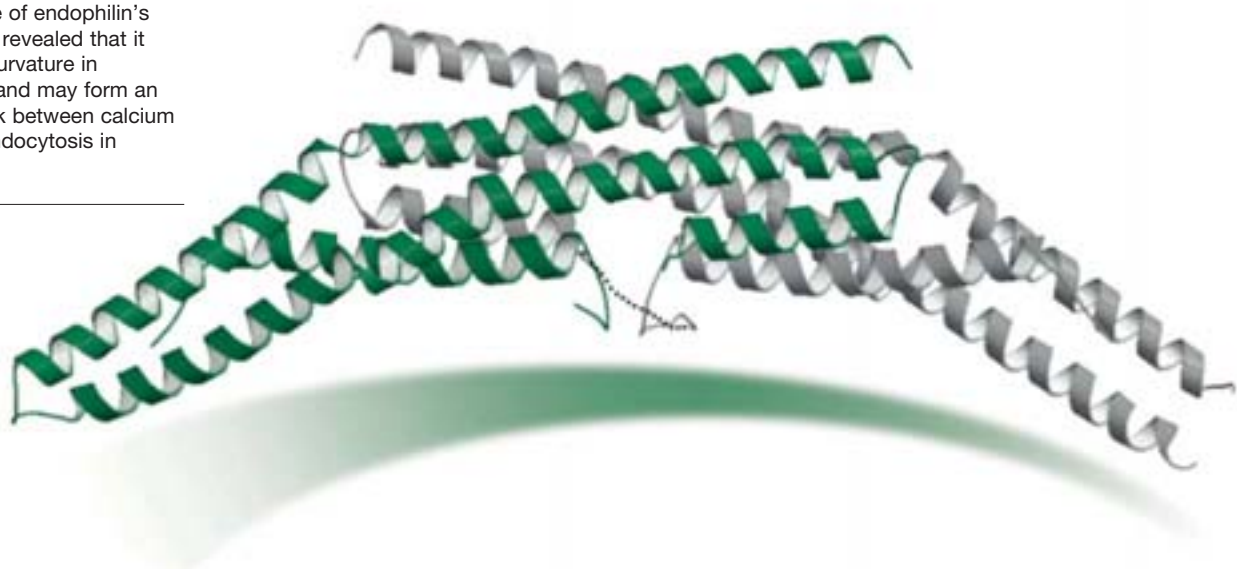
Scientists have been looking into the role that endophilin plays in this process. An analysis of the molecule's *amino acids* – the chemical building blocks that make up proteins – revealed a familiar pattern. The amino-acid code suggested that it folds into a pattern called a *BAR domain*, already known from other proteins. The banana-like shape of BAR modules allows them to dock onto membranes, possibly shaping them. This could happen via the concave side; by teaming up with other membrane-bending molecules, including the proteins *dynamin* and *amiphiphysin* (which also has a BAR structure), it helps to shape a vesicle. Linked together, the proteins use their BAR domains like a pincer, clamping down onto a flat membrane and squeezing, the way your fingers might press on the edges of a playing card until it bends. BAR domains may also simply act as sensors for curved membranes, or do both .

In the test tube, endophilin has this effect: it pinches round membrane structures (liposomes) into tube-like shapes. Although that's an *in vitro* artifact, Winfried says, it demonstrates that BAR domains have the potential to bend membranes.

"We know something about BAR domains in other proteins, but this module in endophilin has some unique features," Winfried says. "If we were to understand its activity, we needed to get a look at the three-dimensional structure of the protein and to see how similar its building plan is to the other domains."

Doing so would require obtaining highly purified samples of endophilin to eventually obtain protein crystals. As is often the case, it was impossible to crystallize the entire protein. Endophilin had to be chopped down to its main functional module, the BAR domain, leaving off some sequences which help it link up with partners. This region did form crystals, and analysis at the ESRF synchrotron showed that the module had some unique and enlightening features.

Winfried explains what the scientists saw by lacing his fingers together and making a cup of his hands. "Most BAR domains join up in twos, like my two hands, and form this sort of bowed structure," he says. "If you press something soft into it, it will be molded into a curve. But endophilin's BAR region has something extra that we could only recognize when we had the structural picture." Keeping his hands interlaced, he slightly raises two of his fingers so that they stand up. "There is an extra domain that sticks out into the cup, like my fingers. You can see



The structure of endophilin's BAR domain revealed that it may sense curvature in membranes and may form an important link between calcium levels and endocytosis in neurons.

Winfried's lab works on membrane and viral proteins.

what would happen if you put something soft into this new shape – it won't bend, or it won't bend as much."

Winfried believes that the finger-like domains may be flexible. It's possible that they can be moved out of the way, which means that endophilin might be able to shape membranes into more than one form. "It suggests that we might need to rethink how BAR domains bend membranes or sense membranes' curvature."

In the cell, specific lipids may play an important role in endophilin's activity. In a liquid environment, fatty lipids cluster together in a structure called a *micelle* – the way oil forms a droplet in water. In the presence of endophilin, the micelles have a distinctive size. This function was discovered in a collaboration between Winfried and Dmitri Svergun's group in Hamburg, using a technique called SAXS (see page 20).

In the meantime, the new structural picture from the crystal has revealed something else. "Communication between nerve cells is controlled by calcium, which regulates not only the release of neurotransmitters, but also endocytosis," Winfried says. "Different levels of calcium influence endocytosis, but so far, no molecules are known which *directly* coordinate the two processes."

Could a single protein establish such a link? It would need to have specific qualities, Winfried says. It should somehow be involved in endocytosis, and there would probably have to be positions in its structure where calcium could be plugged in; when that happened, you would expect to see changes in its activity. "The structure of endophilin shows multiple sites where calcium can bind, and the BAR domain should let it link to membranes," Winfried says. "That gives us some hints which can now be tested experimentally. In addition to anything it might be doing to vesicles, endophilin might have an additional role as a coordinator of the two processes." ■

Things fall apart

NEURONS ARE SOCIAL CREATURES; they must constantly be linked to their neighbors and the bloodstream to survive. Several years ago scientists discovered small proteins called *trophic factors* which stabilize connections between cells; they help wandering axons find and link themselves to precise targets, which may be far away in the body. In the absence of such factors, the cell dies in a process called *apoptosis*. This is a normal process – in fact, it's the fate of about half of the motor neurons that make up the spinal cord as an embryo develops. Just as a gardner has to prune to keep plants from forming dense thickets, neurons need to be pruned in the brain. Cells undergo explosive growth as this organ develops, forming trillions of connections; too many connections would lead to chaos.

Trophic factors and other proteins link to receptors on the surface of neurons, setting off signaling cascades within that alter the activity of genes and determine whether the cell lives or dies.

When the flow of blood is interrupted, as in a stroke, neurons are deprived of oxygen and other vital substances. The cells aren't simply starving, says Manolis Pasparakis of the Monterotondo Outstation. In a collaboration with scientists at the Universities of Ulm and Heidelberg, his group has recently shown that stroke damage doesn't inevitably lead to neuronal death. "In normal development, signals are the key to both survival and death," he says. "They may also be the key to a rescue."

* * *

Rossana De Lorenzi

Postdoc Rossana De Lorenzi and other members of Manolis' group are tackling these questions with an approach typical of Monterotondo groups: exploring the biology of the mouse in search of the causes of human diseases, and hopefully directions for new therapies. Their work has focused on a cellular communication pathway involving the signaling molecule *NFkB*. "This protein has been linked to inflammations and cell death," Manolis says. "In many tissues it is known to prevent cell death – which is important as an embryo develops; on the other hand, it can also promote the survival of malignant cancer cells. We know it is active in the brain, but there have been conflicting reports about what it does."

Manolis Pasparakis

Right: IKK proteins participate in a complex that activates the molecule NFkB. In the aftermath of a stroke, this pathway is active and leads to cell death. Blocking IKK shuts down this type of NFkB signaling, giving cells much better prospects for survival.

The lab has developed a strain of mouse which permits a precise control of NFkB's activity through two other molecules, called *IKK1* and *IKK2*. These two proteins form a complex which permits or blocks NFkB's activation. "It has been difficult to study their role in brain diseases using traditional genetic methods that remove the genes completely," Manolis says. "Mice lacking either IKK1 or IKK2 don't survive into adulthood."

The lab works with special strains of mice in which IKK2 is absent only from the brain. In the current study, the scientists were able to work with additional mouse models, developed by Bernd Baumann and Thomas Wirth at the University of Ulm, which allow the reversible repression or activation of the IKK complex in neurons. Neither seems to affect the development of the animals in any significant way.

These changes have an important influence, however, on the brain in the aftermath of a stroke. Markus Schwaninger's group at the University of Heidelberg had been studying a stroke-like condition in mice, a model that can be used to investigate new therapies. The researchers had noticed that the IKK complex is unusually active in stroke-damaged tissue. Given its relationship to NFkB, and this molecule's known role in promoting cell death, it was logical to assume that the signaling pathway might be triggering irreversible damage in brain tissue.

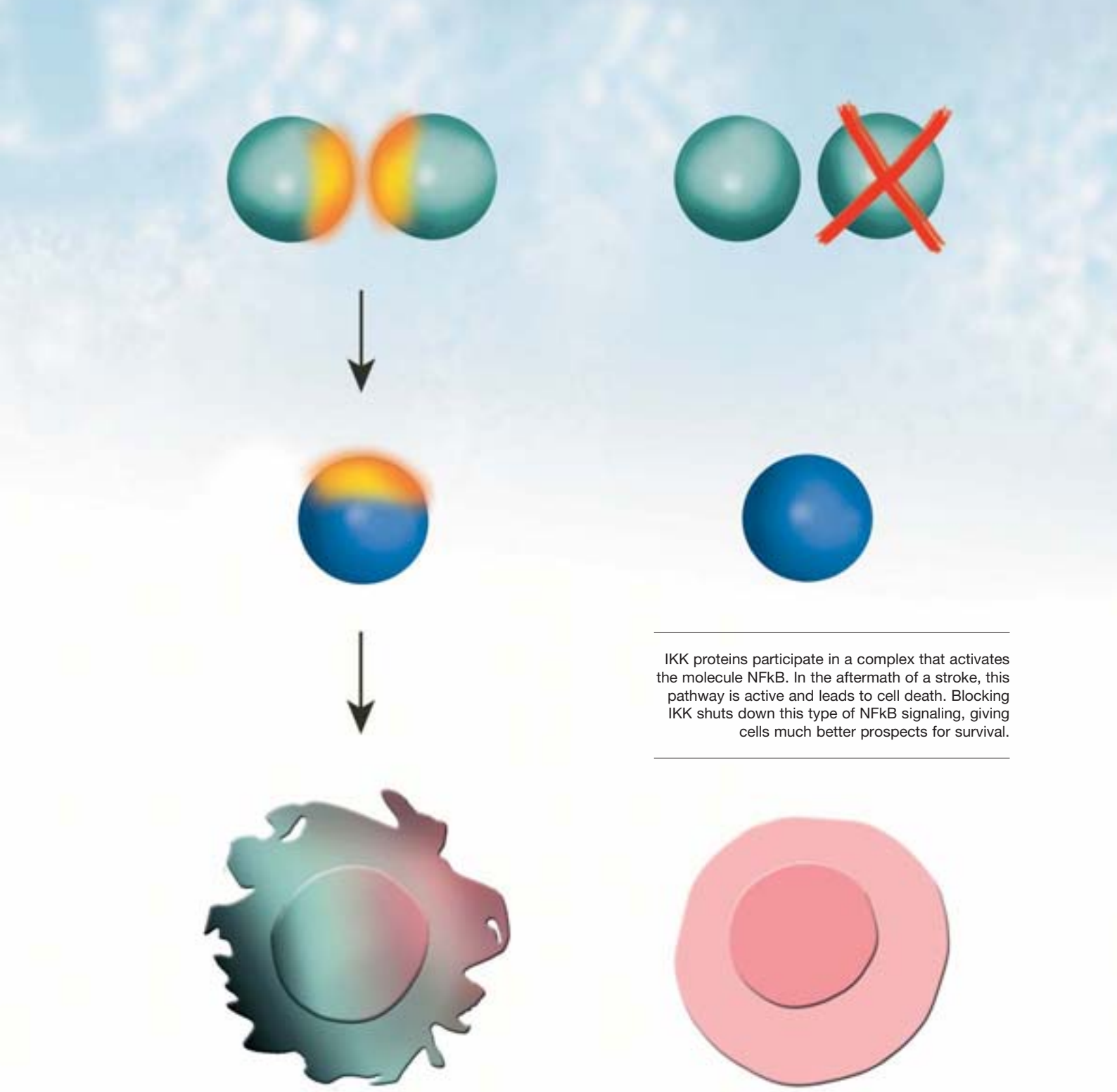
In the aftermath of a stroke, the mice with the hyperactive version of IKK and too much signaling suffered even more damage than usual. But when IKK2 was absent, or its activity was blocked, damage was much less than usual. Instead of dying, nerves were even recovering their functions. And the effects seemed to be long-term; cells were alive several days later.

"This immediately suggested a possible therapy," Manolis says. "When we blocked IKK genetically, by inactivating it in the mouse brain, it could no longer activate cell-death programs after a stroke. Blocking it with a drug ought to have the same effects."

Such a drug was widely available – aspirin is known to shut down IKK very effectively. When aspirin was injected into brain tissue at the onset of a stroke, it reduced the size of the damaged area. But this wasn't a clear enough answer – aspirin has other protective effects on neurons, and it might be helping cells recover for other reasons.

Recently, Manolis says, scientists have discovered another very potent inhibitor, a small molecular compound that seems to affect only IKK. Injected into the brain after a stroke, this also led to an impressive recovery of damaged cells; they were still alive and functioning after two weeks. "Especially important, when thinking of therapies, is that the treatment still worked when it was administered several hours after the stroke," Manolis says. "It often takes a few hours for victims to reach the hospital."

Will the same approach one day work for human patients? Several studies show that drugs which block NFkB, which lies downstream of IKK and is normally activated by it, can offer some protection to damaged neurons. But NFkB stands at a crossroads between other types of signaling, Manolis says, and shutting it down



IKK proteins participate in a complex that activates the molecule NFkB. In the aftermath of a stroke, this pathway is active and leads to cell death. Blocking IKK shuts down this type of NFkB signaling, giving cells much better prospects for survival.

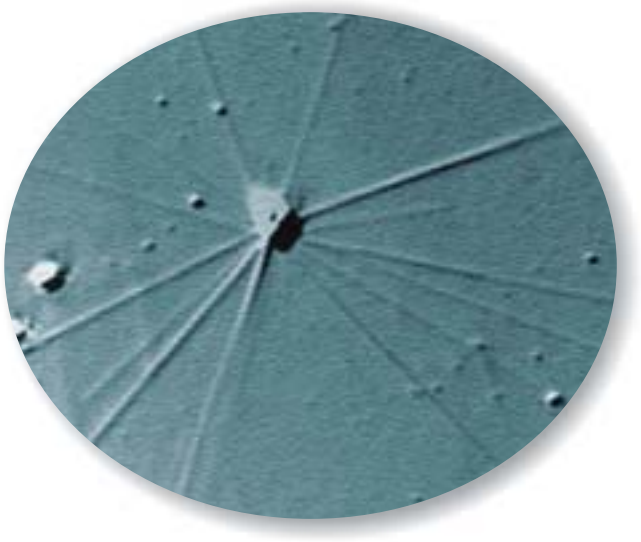
may interrupt other messages that need to get through to genes.

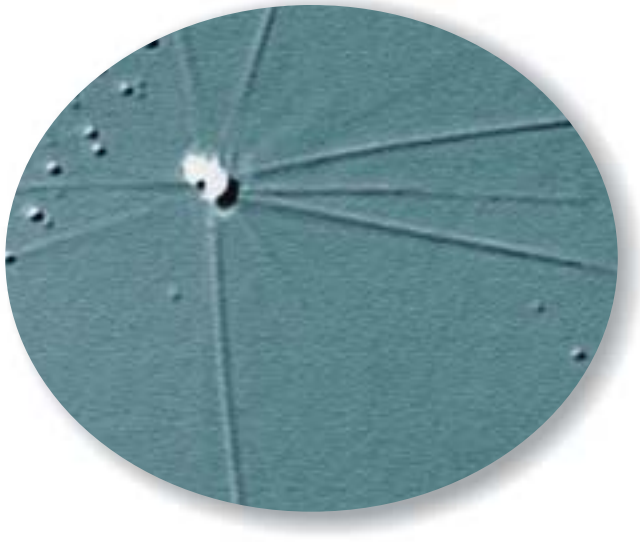
Blocking IKK alone, on the other hand, seems to be more powerful, probably because it interrupts a specific pathway that is especially destructive in strokes without interfering with other signals that aid in survival. The scientists are starting to be able to distinguish between these different information routes. They compared the activity of genes in healthy brain cells with those that had been damaged in strokes, and then with activity in cells that had been treated with the inhibitor. This has helped them focus on a few genes that seem to be particularly important.

The compound used in this study doesn't offer an immediate therapy in its current form, because it doesn't cross the blood-brain barrier. So there isn't yet an effective way to deliver it to damaged tissues. But Manolis is optimistic. " The first step is to find a precise target whose behavior has a major effect on the disease," he says. "We think we've accomplished that. There are several other inhibitors of IKK in development, and some of them may have access to the brain. " ■

Thomas Surrey. If a microtubule in a long axon had the diameter of this pipe, it would stretch from Heidelberg to Milan.

The origins of disorder





IMAGINE STRETCHING A PLASTIC PIPELINE from Heidelberg to Milan and trying to keep it intact. Those are the dimensions and engineering problems faced by some neurons, says Thomas Surrey, but despite its length, the axon of a nerve cell isn't all that fragile. It is supported by a system of stiff *microtubules* which also serve as tracks along which molecules are moved inwards towards the main cell body or outwards to the extremities. This activity is crucial to the axon's function; the cargoes help it broadcast signals to other neurons. Neighboring cells receive signals through branch- and twig-like dendrites, which also depend on microtubules.

A failure in either of the microtubule's functions – delivery or structure – can lead to the death of the cell. Thus this scaffold is one thing that researchers have investigated in search of the causes of neurodegenerative diseases. Some of these conditions have been linked to problems in transporting molecules through axons or in the stability of microtubules. If microtubules break down, the molecules that they are transporting might be lost before they arrive at their destinations. These could well include factors that need to be released and picked up by other cells, like trophic factors and neurotransmitters, which ensure mutual survival and normal activity of neurons.

What happens to the cytoskeleton during the onset of Parkinson's disease isn't clear, but there are good reasons to suspect a connection to the cytoskeleton. As the illness develops, neurons degenerate, and contacts between different parts of the brain are lost. Cells fill with clumps of proteins called *Lewy bodies*, and a major constituent of these are *tubulin* proteins – the subunits of microtubules. Graziella Cappelletti, from the University of Milan, was trying to figure out why. Thomas became involved in the project when she came to Heidelberg to work as a visitor in his lab. The scientists worked with cultures of a type of cell called PC12, which is similar to the neurons that produce the neurotransmitter dopamine and that are destroyed in Parkinson's disease. They added a drug called MPP+, a neurotoxin which kills this type of cell and causes Parkinson-like symptoms in humans. Put together, the two elements serve as a test-tube model for the disease.

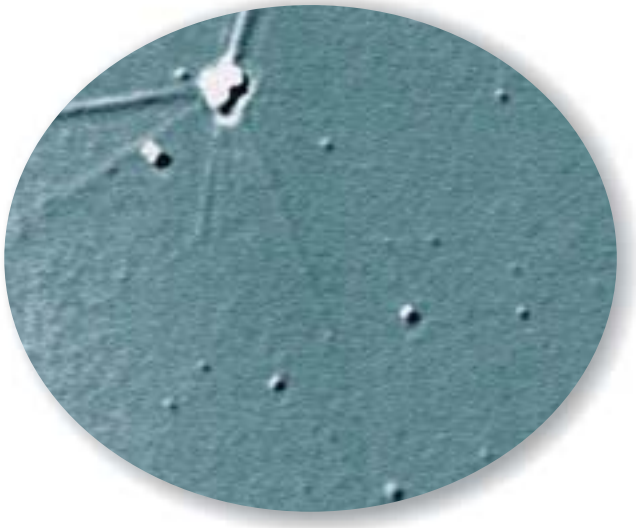
"With the introduction of the neurotoxin, we saw an increase in the pool of free tubulin molecules which were no longer being assembled into microtubules," Graziella says. "This is exactly what happens in cells when there are factors around that destabilize the structures. Microtubules are generally very dynamic: the tips are continually being built and disassembled, but the rate at which this happens has to be carefully controlled, and it changes under different circumstances. Otherwise it's impossible to create long, durable structures like those that support the axon."

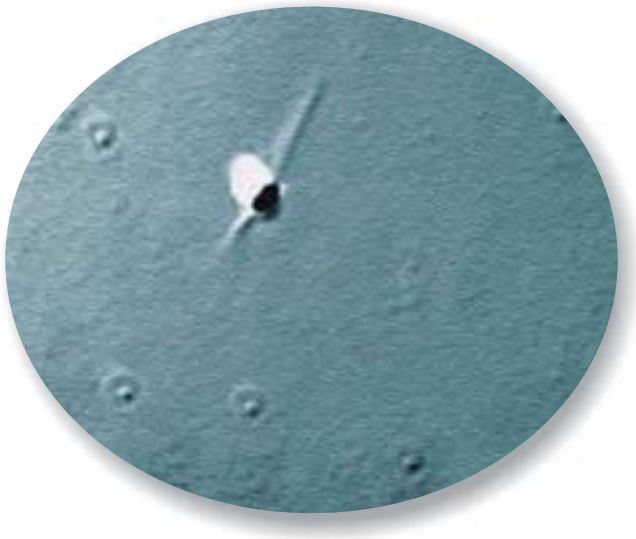
The scientists used time-lapse video microscopy to watch the effects of the drug on the formation of microtubules. They found that adding MPP+ decreased the number and size that were created. A close look at individual microtubules revealed that they went through alternating phases of lengthening and shortening, a process known as dynamic instability. The phases could be measured very precisely – the more MPP+ that was present, the more quickly microtubules were broken down.

There were two likely reasons: either MPP+ was interfering with molecules that help assemble microtubules, or it was binding directly to sites on the structures, preventing new subunits from being plugged in. The scientists performed further experiments to see which scenario was the case.

"Tubulin forms pairs that are then plugged onto the growing end of a microtubule," Graziella says. "MPP+ doesn't bind strongly to the pairs, but it does bind to tubulin proteins when they are already attached to the microtubule. It does so in an unusual way – at high concentrations there might be a copy of MPP+ bound to nearly every tubulin pair."

Parkinson's disease disrupts a specific area of the brain by breaking crucial connections between neuronal cells that produce the neurotransmitter dopamine and those that are stimulated by it. Ideally, scientists would like to track this destruction back to events within the cell. A major factor could be the breakdown of microtubules, Thomas says; the similarity between the disease and how the neurotoxin affects the brain shows that this may be the case. Evidence from other studies shows that microtubules malfunction under the influence of mutations that cause Parkinson's disease and the devestating effects of some neurotoxins. This makes a strong case for looking at the surface of these structures when searching for the causes of the disorders and for cures. ■

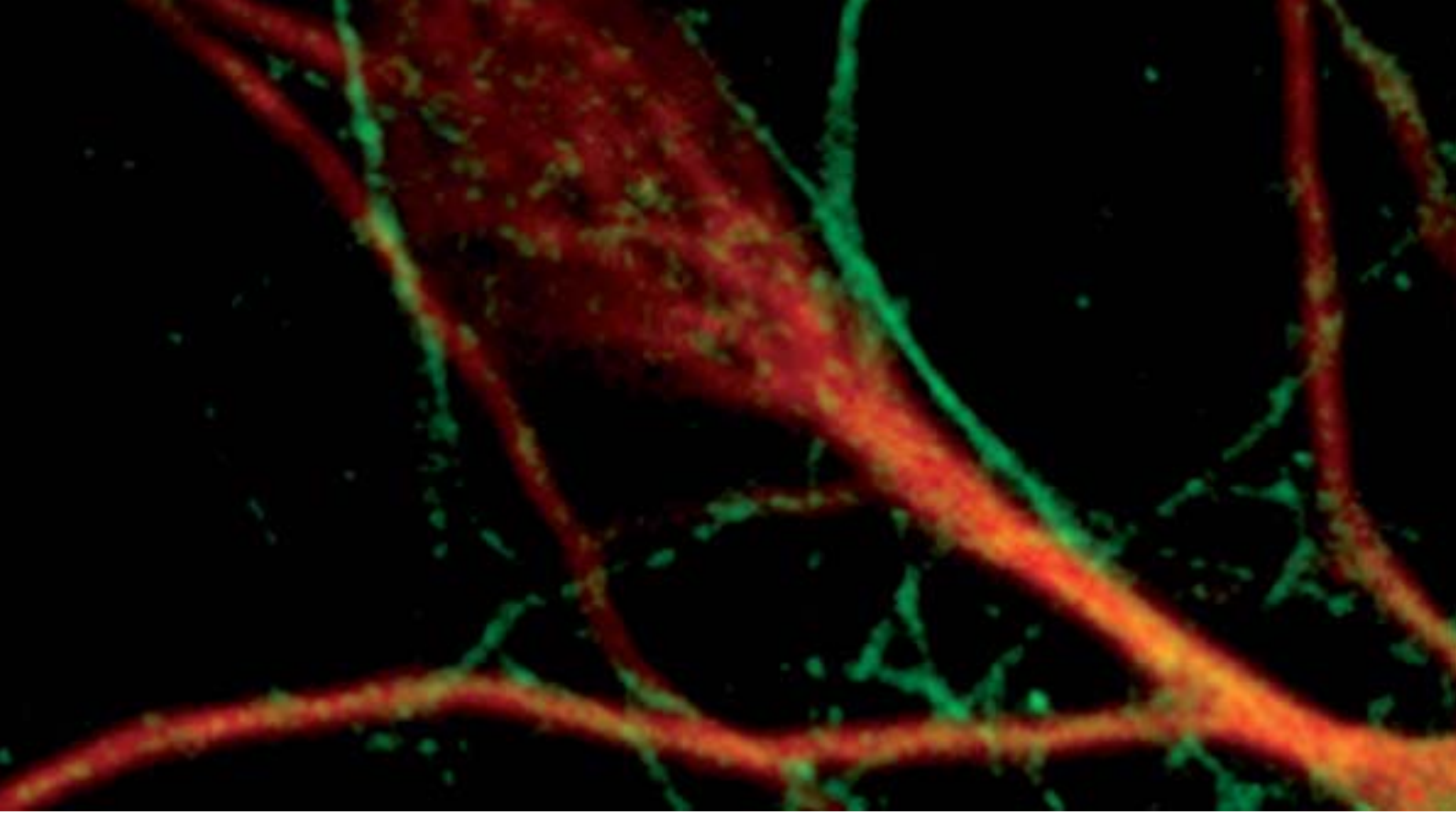




MPP+ binds to the surfaces of microtubules and causes them to break down, changing their length and behavior.


7821
Dealic

The operators of the brain



IN THE EARLY DAYS OF THE TELEPHONE, operators sat in front of huge switchboards, receiving a call on one line and physically connecting it to the line of the person the caller wanted to talk to. In the brain, neurons act a bit like operators: a single cell can transfer "calls" coming from many different directions. On an even smaller scale, within each cell, molecules also act as operators. Single proteins can receive multiple signals and route them in different directions. These levels are connected: the information pathways in a cell help determine the behavior of the neuron, and thus the networks in which it participates.

This might not seem surprising in our age of networks, where the behavior of the vast Internet ultimately depends on what is going on in the circuitry of lots of single computers. But in contrast to logic boards, network routers and firewalls, the brain doesn't come equipped with schematic diagrams of its operations, and the way it works reflects millions of years of evolution in very diverse organisms.

"It would be good to have such a plan," says Nicolas Le Novère of EMBL-EBI, "because it would give us a way to look for treatments when things go wrong in the brain, for example in diseases, or addictions. But figuring out the connections is very difficult, because it ranges from the behavior of molecules within single neurons to the function of entire networks of cells."

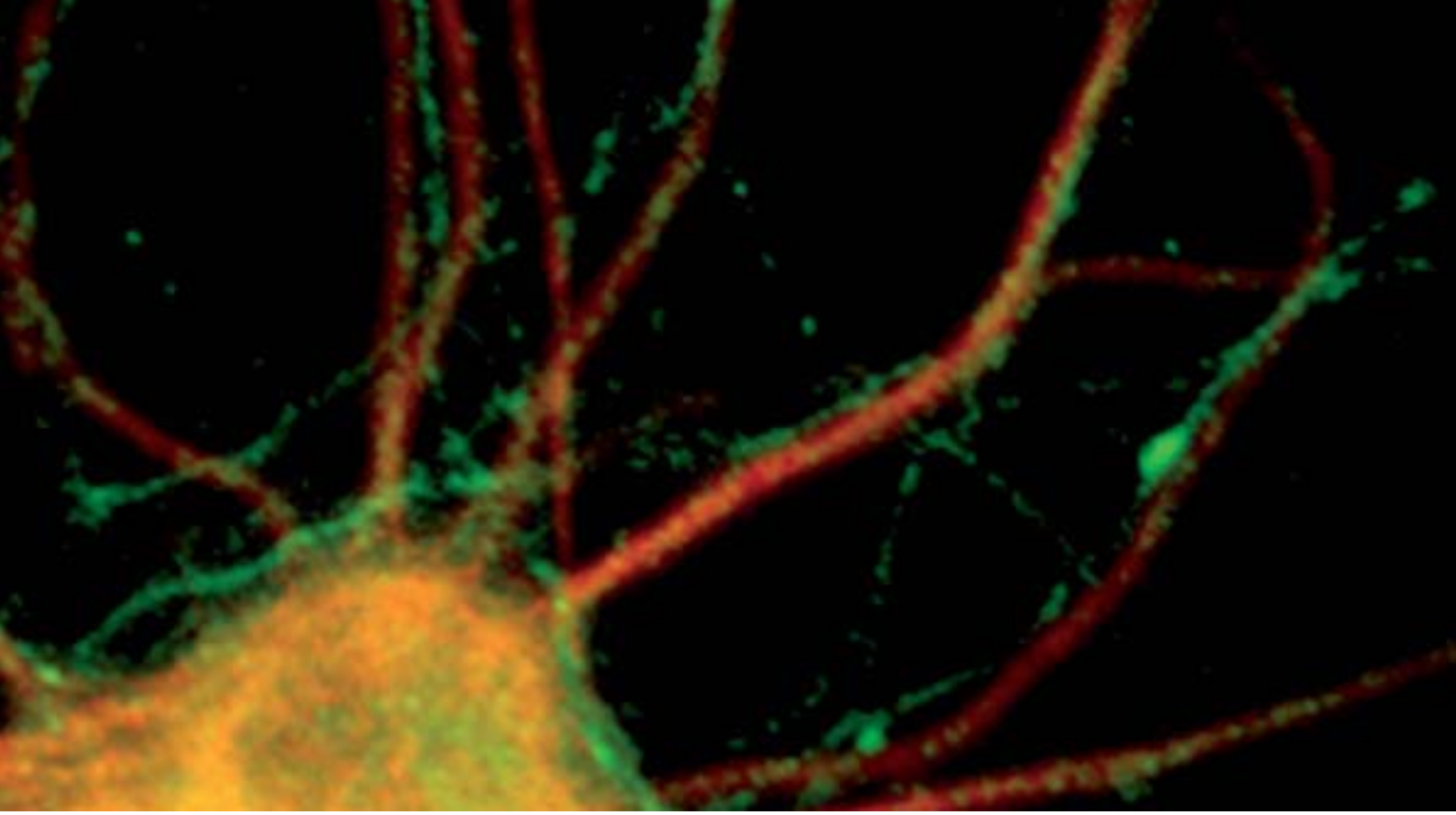
Every question about the function of the brain is inevitably a question about such networks and their adaptation. Molecules permit brain activity; conversely, what the brain does has an effect on its structure. So many components are involved in this self-organizing behavior that it will be impossible to understand without sophisticated models. In many cases, the models that are needed don't yet exist, so Nicolas and his colleagues are helping build them. Of those that exist, it's hard to tell which one to use for what purpose. So to understand the language of the brain, Nicolas' lab is designing another language to describe models themselves.

* * *

One project of Nicolas' group has been to model the activities of a protein called *DARPP-32*, which has been linked to substance abuse. The protein is produced by neurons of a brain structure called the *basal ganglia*, where it is known to help to decode chemical signals received by the cell. Drugs alter how these cells communicate with each other, and ultimately the structure of the brain.

Like an efficient switchboard operator, DARPP-32 can receive a wide variety of signals from different directions and route them to their proper recipients. The reason lies in its structure: DARPP-32 contains many different amino acid capable of being "tagged" by chemical messages.

"One of the most common systems of cell communication involves molecules that add phosphate groups to, or strip them from particular amino acids in a protein," Nicolas says. "DARPP-32 has several sites which can be *phosphorylated* in this way. Since each modified site has a different structure and chemistry, it has its own partner



molecules that recognize it, and that accounts for the overall switchboard-like activity of DARPP-32."

Scientists have discovered that messages involving the neurotransmitter glutamate and calcium ions pass signals through DARPP-32. The same is true for neurotransmitters such as *dopamine* which regulate the production of *cAMP*, a small signaling molecule involved in tuning the receptiveness of neurons by helping to open and close channels in the membrane. "DARPP-32 seems to be such an important operator that it might be useful in understanding addiction or other brain dysfunctions," Nicolas says. "But first we have to understand in-depth how it works."

That means tracing all the "calls" that come through the protein switchboard – following communication pathways back through the many molecules that transmit a calcium or cAMP signal into the cell, and establishing all the links both upstream and downstream of DARPP-32. "Each of *those* molecules has its own activators and repressors which have to be figured into the system," Nicolas says. "Maybe you're beginning to get an idea of how complex this is. Think of finding a chessboard that somebody has abandoned in the middle of a game, and trying to figure out all the moves leading up to this situation, and all the moves that might come later. But this chessboard has thousands of pieces, and they don't always follow the same rules."

At any one point in time, DARPP-32 might be tagged with one group of phosphates, or none, or several. Each of those states arises from a different set of complex events, involving signals received by the cell and the context of the other molecules that phosphorylate DARPP-32 and strip phosphates off again. Nicolas and his colleagues started to model all of these behaviors using simulation software, permitting them to reproduce chemical reactions *in silico*. To firmly base the models on experimental findings, Nicolas established a collaboration with Jean-Antoine Girault, in Paris, who started to work on DARPP-32 in the 1980s in the group of Nobel prize-winner Paul Greengard.

"Filtering through all of this information, carrying out simulations with the software and comparing the results to experimental evidence showed us some interesting things," Nicolas says. "We were able to show that another molecule called CDK5, which acts on DARPP-32 phosphorylation, is a key regulator of the sensitivity of neurons to dopamine. We were also able to show that other candidate regulations – old likely suspects – actually have little effect on the system."

The analysis also shows that DARPP-32 plays an even more central role than scientists had thought. It not only detects signals, but helps makes sense of complex situations. Nicolas calls it a "coincidence detector" whose various states help neurons make a coordinated response to complex situations. In a sense, it's listening in on the calls, and helping cells to decide what to do based on the information passing through the circuits.

* * *

Nicolas Le Novère of EMBL-EBI

Changes in information pathways alter the cell, as if a switchboard could learn and rewire itself as calls came through. These changes involve physical rearrangements of components such as receptor proteins in the cell membrane. Adding more receptors or clustering them closely makes the cell more sensitive to a signal, like adding antennae to improve reception on a radio.

"These receptors are embedded in complex molecular machines whose composition alters depending on what they need to do," Nicolas says. "They're also linked to very dynamic elements of the cytoskeleton. It's another problem that is now recognized as being very complex, and we'll only be able to cope with the interactions of all the elements through modeling."

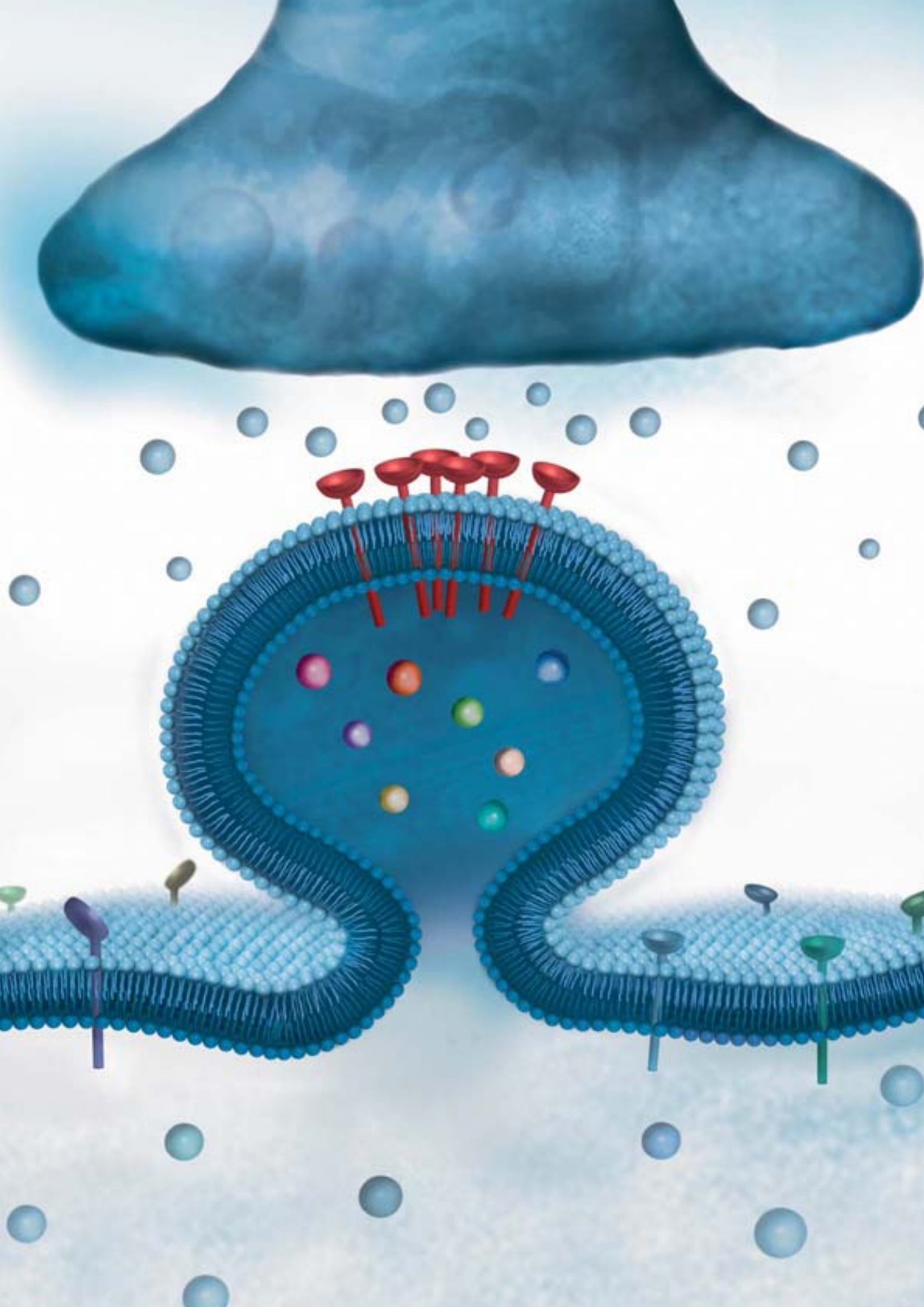
His group is now looking at a region of neurons called the *dendritic spine*, a bulb-shaped compartment responsible for decoding signals coming from neighboring cells. Nicolas is interested both because it undergoes significant changes through activity – many scientists believe that changes in the spine are the physical basis of learning and memory – and because its contents are isolated from the rest of the cell. This means that a finite number of elements are involved when it restructures itself, making things easier when it comes to modeling.

In trying to tackle this problem, Nicolas and his colleagues have come across others. "Many of the most interesting molecules involved in neuronal signaling are membrane proteins," he says. "Truly understanding their activity means having a very clear picture of their structure and how it changes. But it has been very challenging to obtain structures of membrane proteins; they are typically very difficult or impossible to crystallize." The group maintains the computing infrastructure for a project called E-MeP, a research platform that aims to solve the bottlenecks in obtaining structures for 300 proteins embedded in the membrane and complexes associated it. Most of these are involved in neurotransmission, and many of them have been linked to devastating diseases.

Trying to grasp complex processes has brought other challenges. "Modeling requires various types of software – simulators and what we call 'design environments' – which may exist or may have to be invented," Nicolas says. "A good model should be accessible to all sorts of users who use it to understand processes, refine it and adapt it to new situations. But there are obstacles to this type of sharing. There has been no standard way to describe models so that they can be operated under different types of software, and no centralized collection of models with good descriptions of how they work and what they are for."

So Nicolas' group has been participating in initiatives to establish the BioModels Database and a standard language to describe its contents, called the Systems Biology Markup Language (SBML). "We hope this will become a major resource that will allow scientists to store, search and retrieve published mathematical models of biological interest," he says. "They are annotated and linked to relevant data, such as publications, databases of compounds and pathways, and standardized languages by which things are described." ■

Nicolas is taking a systems approach to the study of dendritic spines, small structures in neurons whose contents are rather isolated from the rest of the cell, and which undergo fascinating changes as the neuron communicates with its neighbors.



Valeria Carola, Giovanni Frazzetto, Cornelius Gross and Simone Santarelli of EMBL Monterotondo, with a not-so-anxious mouse named Volker.

Baby mice and the fear of flying

FOR PARENTS, the question of nature versus nurture is more than an abstract philosophical issue – it's a daily struggle. Anyone who has raised more than one child knows that some things seem to come with the package: children have a character from the day they are born, often markedly different from their siblings', and certain aspects of their behavior are difficult or impossible to change. At the same time, we know that childhood experiences have a profound impact on our lives, and psychiatrists' couches everywhere are full as people try to come to grips with their early years.

Scientifically speaking, says Cornelius Gross, it's hard to peel apart the genetic and environmental factors that contribute to mental illnesses which appear in adulthood. "It would be useful to be able to do that," he says, "because conditions like depression and anxiety incapacitate a huge number of people every year. We might be able to recognize susceptible children and deal with them in a healthier way."

Cornelius and his group in Monterotondo are using strains of mice to take a look at genetic and environmental aspects of behavior and mental illness. Just like people, he says, mice respond to differences in rearing. They develop personalities, and some of them suffer from anxiety. It may seem like a stretch to extrapolate from mouse in a plastic experimental box to a human's distress upon boarding an airplane, but if Cornelius is right, these animals provide valuable clues about the connection between genes, early childhood experiences, and the quality of our lives as adults.

* * *

Over the past few years, scientists have discovered that a common alteration in a gene seems to link adverse childhood experiences to depression in adults. The gene encodes a protein called *5-HTT*, which is involved in communication between neurons in the brain. Normally it helps neurons take up *serotonin*, a neurotransmitter that plays a role in emotions and moods.

"5-HTT helps transport serotonin into the cell, and the altered form that has been discovered in patients leaves a lot of serotonin floating around between brain cells," Cornelius says. "Imaging techniques have shown that this can make brain circuits susceptible to long-term changes when there is a lot of stress early in life."

But he adds that there are so many variables in the life of a mouse – or a human – that it has been hard to estimate the importance of parenting in the development of anxiety. With a lot of work, he thought, it might be possible to control some of these variables in the mouse.

"First, we needed offspring that might develop symptoms of anxiety but also might not," he says. "Mice, like humans, have two copies of each gene. Animals with two altered forms of 5-HTT are likely to develop anxiety no matter what type of care they get from their mother, and animals in which both copies are normal tend to be robust and don't develop anxiety. So we used offspring that had one copy of each form of the gene, reasoning that they might go either way."

It was equally important to control the genetic background of the mothers; different strains of mice have different rearing behaviors. A normal female mouse is attentive to her offspring, grooming them and helping them during nursing by arching her back, making it easy for the pups to find her nipples. But other strains are less attentive, and some of this mothering behavior has also been linked to genetic factors.

"Some genetic profiles yield mothers that exhibit a lot of caring behavior; other profiles produce mothers that barely seem to notice they have offspring. But this is surely not the only difference in the animals, and we didn't want even more factors entering into the environment. So we needed to find genetically identical mothers that differed only in their behavior toward the young."

Cornelius and his colleagues had developed such a strain, using fathers and mothers from two different strains, called C57 and BALB. C57 mothers groom their young and show typical "caring" behavior. A mother with a BALB background is considerably less attentive. Crossing a BALB female with a C57 male yields offspring with the same genetic makeup as crossing a C57 female with a BALB male, but the offspring receive different types of care. And if they are raised by an inattentive BALB mother, they go on to care less for their own young than mice with a C57 mother, showing that some behavior can be environmentally "transmitted". Such crossed pairs provided Cornelius and his lab with the mothers they needed. They were crossed with fathers that had the 5-HTT mutation, giving some of the offspring one copy of the altered gene.

"Now we had genetically identical mothers that showed different caring behavior and identical offspring that ought to be susceptible to different environments," he says. "All we had to do was see if any of the mice grew up to become anxious."

* * *

In the three years since he established his lab in Monterotondo, Cornelius has accumulated a diverse team. Valeria Carola is trained as a psychotherapist; Olivier Mirabeau came to EMBL with a degree in applied

Monterotondo – the way the mice see it.

mathematics. Another member of the group is Giovanni Frazzetto, who splits his time between Monterotondo and the BIOS Centre at the London School of Economics run by Stephen Rose. Giovanni received a fellowship from the Branco Weiss 'Society in Science Fellowship Programme' in Switzerland to combine science with a socially relevant theme – to investigate the impact that advances in neuroscience are having on individuals and society.

All of these types of expertise have been crucial in unraveling the complex biological and environmental determinants of behavior. In this case, Cornelius says, Valeria's expertise in statistics was crucial to the project. "There are several standard tests to look for symptoms of 'anxiety' in mice," he says. "One involves putting them into an open area; anxious mice immediately run to the edge – they don't like to be out in the open. There are also differences in sniffing behavior and what we call 'risk-assessment behavior.'"

The lab carefully monitored how mice with one copy of the altered 5-HTT gene – raised in different caring environments – performed on the tests. The results were odd. "They were all over the place on the chart," Cornelius says. "On some tests one group seemed anxious and the other didn't; on other tests they were reversed. Clearly there was something going on, but it was hard to pick out a pattern."

Valeria subjected the results to a statistical procedure called *principle component analysis*. This method helps determine whether different types of variables can be clustered into meaningful groups. Suddenly a pattern emerged.

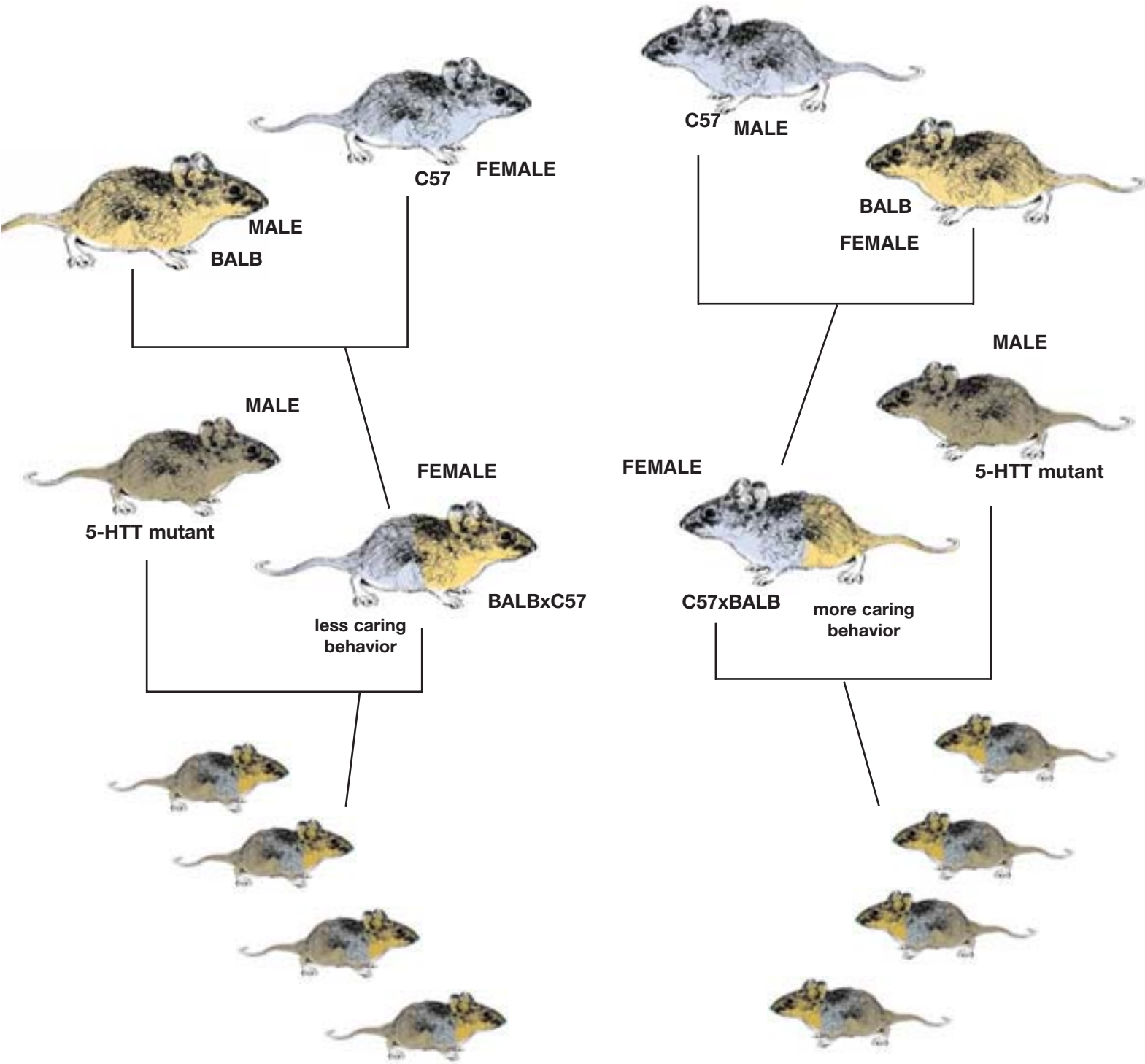
"Behavior in our tests can be divided into two major categories," Cornelius says. "One type can generally be called avoidance and risk-assessment behavior. The other has to do with exploration and risk-taking. When we considered each of these as separate classes of behavior, rather than simply lumping them all together in a general type we thought of as anxiety, we found strong correlations between the kind of rearing the animals received from their mothers and how they performed."

Mice with low maternal care are more likely to avoid possible threats; they also perform more risk-assessment when investigating an environment. Cornelius says this can be regarded as a specific behavioral strategy that may be applicable as models for human behavior.

What is the difference in maternal care doing to the brain?

"You have to remember that we're looking at genetically identical mice," Cornelius says. "We think we've shown that the two versions of the gene make them susceptible to this environmental influence. In infants, differences in care change the amounts of free serotonin in the brain – neurotransmitter that is released but isn't being taken up again. Our first thought was that this difference continues into adulthood – but that didn't turn out to be the case. Instead, we think the change in early serotonin levels affects something else, something fundamental about the brain circuitry, which is carried into adulthood. We'll now search for what that might be." ■

A lineage of anxiety – or not



To study environmental effects of anxiety, Cornelius and his colleagues needed mothers of an identical genetic makeup but which showed different levels of care for their young, and newborn mice that were susceptible to, but not strongly predisposed towards, adult anxiety. Females with a mixture of BALB and C57 parents received different types of care as infants and thus cared differently for their young. Infant mice with one defective copy of the 5-HTT gene proved susceptible to developing some types of anxiety as adults.

An organigram for muscle

Janus Jakobsen, Thomas Sandmann, Michal Karzynski and Eileen Furlong

THOMAS HUNT MORGAN, the famous pioneer of animal genetics, ran afoul of his biochemist contemproraries when he stated: "At the level at which the genetic experiments lie, it does not make the slightest difference whether the gene is a hypothetical unit or whether it is a material particle." Morgan and his lab had found an ideal organism to work with, the fruit fly, and they were steadily discovering mutations that affected eyes, wings, and other parts of its body. If a characteristic could be inherited according to the principles of Gregor Mendel, there was a gene behind it, and it didn't matter whether genes were made of nucleic acid, proteins, or green cheese.

What genes are made of, and what the proteins they encode do, matter very much to today's developmental biologists. "The development of an embryo depends on complex patterns of gene activity in space and time," says Eileen Furlong. "Understanding the regulation of these

patterns will require dissecting the components of networks that transcribe the genes, locating precise regulatory sequences near the genes that control how they are transcribed, and understanding how these sequences act together to give rise to complex pattern of expression."

Eileen has been studying these questions in a family of molecules called *Mef2*, which are known to have a powerful influence on the development of muscle. Vertebrates have four versions of Mef2, which are produced in overlapping waves as the muscle tissue of embryos develop. Flies only have one version; it, too, appears in all muscle types. She decided to use flies as a model system to try to understand, in detail, the behavior of this molecule. It ought to be easier to understand one molecule than four, she reasoned.

* * *

Transcription factors often work by pairing up with other molecules, and are often bound to DNA as members of large protein complexes. Their presence attracts the machinery that transcribes a nearby gene into RNA, which can then be translated into proteins. Although the overall impact of transcription factors may be obvious, scientists often have no idea precisely which genes they affect.

"Most evidence that we have regarding Mef2 has been indirect," Eileen says. "Genetic studies show that muscle cells no longer differentiate if this protein is missing. If a mutation in another gene has a similar effect, it may be because Mef2 can no longer directly activate it – or there may be another, less direct connection between the two molecules."

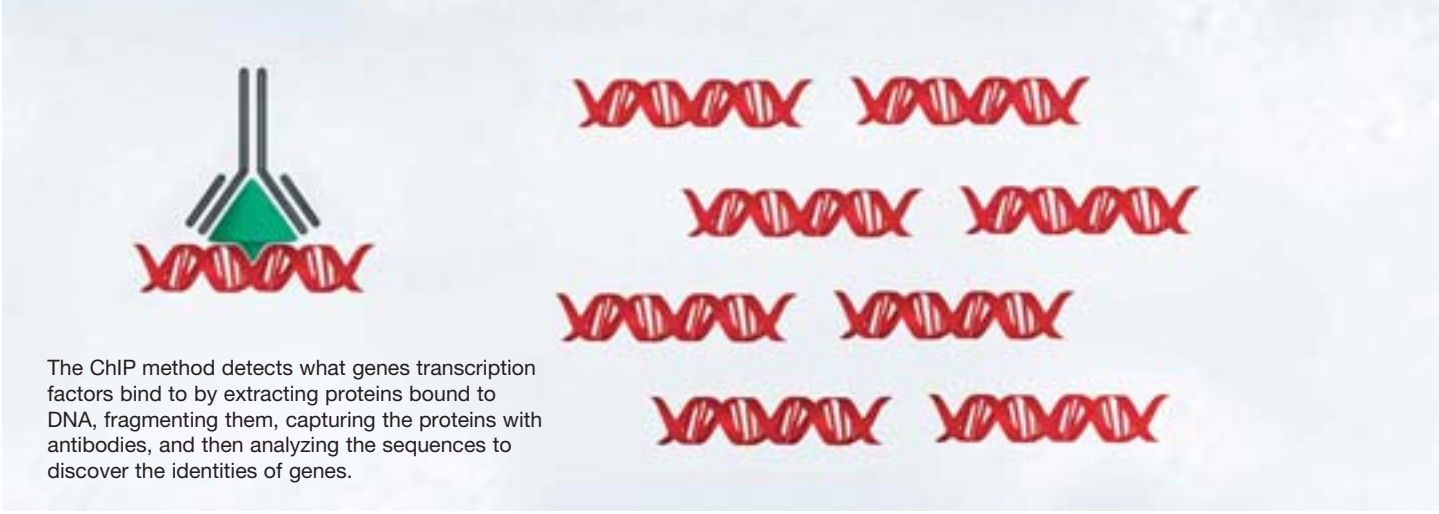
Bioinformatics approaches can be used to try to ferret out common features of the genes that Mef2 can regulate. These predictions can then be tested in experiments. Transcription factors usually work in groups and bind to DNA sequences called *cis regulartory modules* (CRMs). These sequences can be a large distance away from their target genes, making it very difficult to predict what may act as a CRM. Even if a sequence recognized by a transcription factor is clearly present in a CRM, scientists usually have no idea if and when it is active.

Recently scientists have developed a powerful new method called *ChIP*, for *chromatin immunoprecipitation*, which directly detects the DNA sequences that transcription factors bind to. "The beauty of this technique is that it identifies where the transcription factor binds in living cells – in our case embryos – rather than in a test tube," Eileen says. ChIP extracts DNA from cells; the genetic material is cut into fragments with a method that doesn't interfere with any proteins that might be bound to it. Scientists lock onto these proteins with antibodies, which act like fishhooks. The DNA fragments attached to the transcription factor are analyzed and compared to the sequence of the genome to discover what genes have been "caught".

Other groups have used the method to examine Mef2 binding, but experimental limitations meant that only a fraction of the genome was covered. "We wanted to get a complete reading, and we wanted to do it at different stages of development," Eileen says. "Mef2 is produced throughout the lifespan of muscle cells – from the very first steps of differentiation in the early embryo all the way to adult tissue. So we needed a readout of the activity of genes it might be affecting over time."

Postdoc Ioannis Amarantos from Eileen's lab worked with Jos de Graaf from the GeneCore Facility and members of Eric Johnson's lab at the University of Oregon to create a tiling array (see page 76) spanning over half of the fly genome. This provided a global picture of where Mef2 could bind. They repeated the experiment at several stages of embryonic development.

The scientists discovered that Mef2 attaches to 670 distinct regions of the genome. "Among these sequences were CRMs that Mef2 was already known to bind to, in connection with eight particular genes," Eileen says. "In several of these cases, we found that it could bind in more than one place near the gene. Other sites lay near genes that have been connected to the Mef2 network."



The ChIP method detects what genes transcription factors bind to by extracting proteins bound to DNA, fragmenting them, capturing the proteins with antibodies, and then analyzing the sequences to discover the identities of genes.

Knowing what CRMs Mef2 can bind to is a good start, she says, but when, during the many stages of development, is it actually doing so? And what effect does Mef2 have on the target genes when it binds? Previously, Mef2 was thought to play a key role in the later stages of muscle differentiation, but the scientists knew that it also appeared as cells took their first steps along the path of differentiation. The ChIP method gave snapshots of its behavior at different times.

Thomas Sandmann, a PhD student in Eileen's lab, discovered that Mef2 seems to work in three main phases – each time, locking onto different CRMs. It occupies 50% of the sites at all phases of muscle development. "That's what you would expect," Thomas says. "But we were excited to see two other types of behavior. It binds to another 21% of the CRMs when the fly embryo is 4 to 6 hours old, and then it leaves again. And 32% of the CRMs are only occupied during the later phases of development."

The transcription factor is produced continually throughout the lifetime of the cells – why doesn't it bind to all the CRMs, all the time? Eileen says this is likely due to the presence or absence of other molecules that help control its effects on specific genes. For example, many of the genes activated during the early phase depend on a co-factor called *Twist*, and Eileen and her colleagues discovered that Twist and Mef2 bind on the same sites during the 4-6 hour phase of development. Later Twist isn't produced. There also seem to be differences in the CRMs activated at later stages. Such genes tend to be accompanied by multiple binding sites.

* * *

What does Mef2 do when it binds to a CRM? "It might just be sitting there," Eileen says. "We needed to see in which cases it was having an effect on the genes expression." Another type of "chip" (DNA chips, page 72) helped provide an answer. In one type of experiment, Eileen and her colleagues surveyed the entire fly genome, comparing the difference in gene activity between normal muscle and cells with a mutant, non-functional form of Mef2. They were particularly interested in the behavior of genes near the Mef2 target CRMs.

The scientists discovered that the presence of Mef2 seems to have a direct effect on a high percentage of these genes. Even making very conservative judgments, 218 genes seemed to be directly under the control of Mef2.

Another study compared the times at which Mef2 binds to the activity of the genes, based on the three phases discovered in the ChIP study. "Mef2 acts as a trigger," Eileen says. "Almost as soon as it binds to a CRM, something happens to a nearby gene."

The scientists also added Mef2 to other types of cells which don't normally produce it, and discovered that these cells were now switching on genes that usually only appear in muscle.

Eileen says the current study covers about half the fly genome; the rest should reveal many more Mef2 targets. "It's likely that the transcription factor regulates as many as 1000 genes during development," she says. "Given the huge number and diversity of molecules that these genes encode, and the fact that regulation takes place over the entire course of muscle development, it's likely that Mef2 acts in a broad way to help regulate the entire development of muscle."

A list of targets and CRMs is only the beginning; the next step is to develop a picture of how Mef2 functions within a larger network of transcription factors. "We are at an exciting stage at which we now have the ability to compare multiple transcription factors binding to CRMs in overlapping temporal patterns," Eileen says. This type of data will help unlock the combinatorial code that regulates precise temporal and spatial expression during development. With about 1000 genes potentially involved, Eileen says, unraveling the details is likely to keep the group busy for a while. ■

Mark Isalan and Luis Serrano

Artificial bodies

GEOGRAPHY HAS ALWAYS PLAYED a crucial role in the development of life. An earthquake or a flood can create natural barriers that divide the population of a species into groups; they no longer mate and thus take separate evolutionary paths. In human history, the development of races, cultures and nations have largely followed lines set by oceans, rivers and mountains.

A growing embryo is also defined by geography. In the early phases of an organism's life these aren't physical obstacles like skin, or the walls of blood vessels, but hills and valleys and slopes created by concentrations of molecules. Cells at particular locations in the body secrete proteins which spread and affect other cells in different ways, depending on their concentrations. This leads to differentiation between tissues and structural borders; easily recognizable examples are the digits of fingers, or the veins that separate "panes" in the wings of flies.

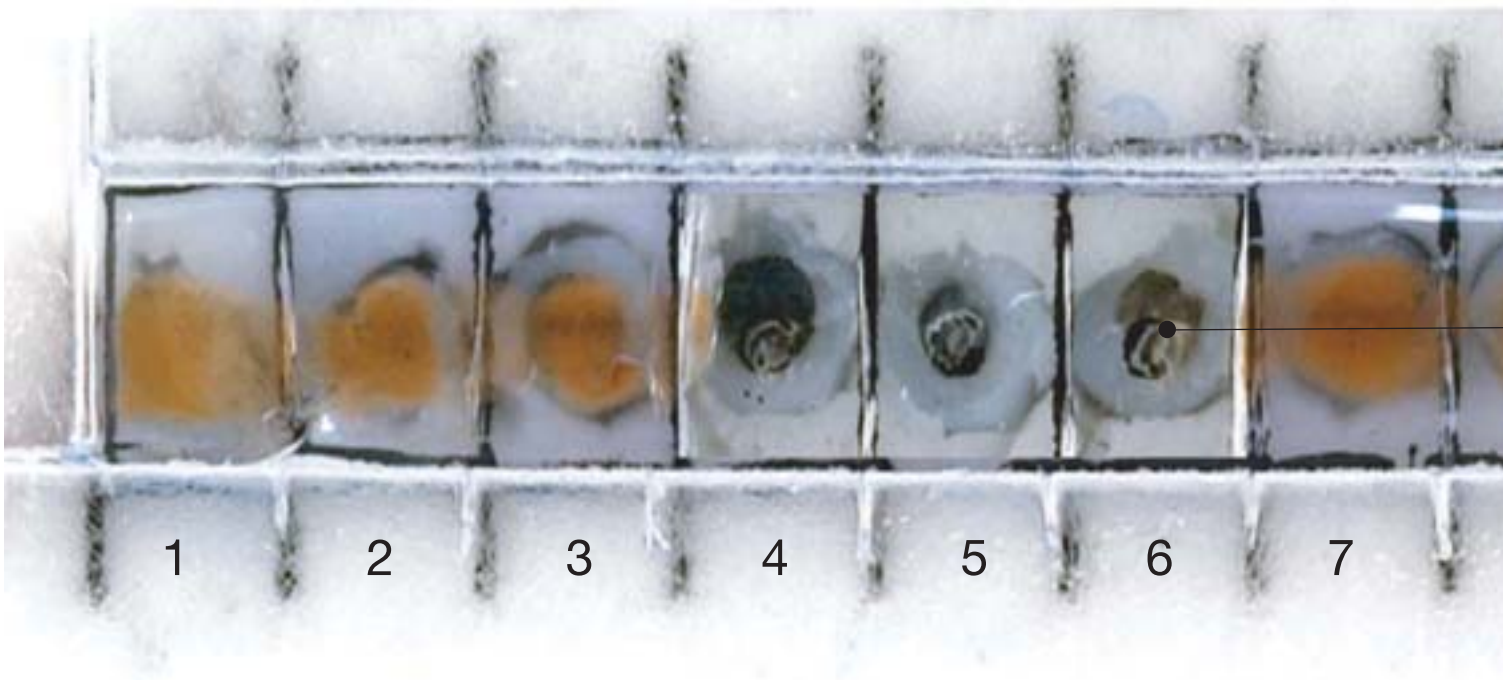
A similar geographic phenomenon can create structures within single cells. Eric Karsenti and Philippe Bastiaens showed that changes in concentrations of molecules guide the construction of the spindle that divides chromosomes in cell division (see page 24). And nearly twenty years ago, EMBL alumna Christiane Nüsslein-Volhard and her colleagues showed that molecules from a mother fruit fly flow through its huge egg cell, laying down patterns that will become body segments, a head-to-tail structure of the embryo that develops from this egg.

"These segments arise from a complex network of interactions between molecules," says Luis Serrano. "One phenomenon is that the concentration falls off farther from its source. Another is that other molecules respond differently when they encounter different amounts of the protein. They may, for example, actively repress it, or it might alter their own behavior."

When Luis looks at this phenomenon, he sees overlapping circuits, relays, inhibitors and networks. A fly egg is very complex – it's an environment consisting of several thousands of different types of molecules – but it's possible that only a few of these are truly calling the shots when it comes to building structures, and that they are pulling it off according to simple rules. If so, he says, you might be able to imitate it in the test tube. And now he and his colleagues in Heidelberg have done just that, using genes and proteins, microscopic magnets and a tiny plastic chamber.

* * *

Luis hasn't yet created artificial life, but every year he seems to be taking a step closer, and if you ask him he will probably admit that it would be an intriguing thing to do. A few years ago, trying to get a handle on self-regulating genes, he implanted a small artificial gene network in bacteria that switched on and off fluorescent proteins. The result was something like a lamp with a sensor which switches itself on when things get dark and turns itself off when it senses its own light. Cells flickered on and off, sometimes getting confused and getting stuck in



an intermediate state. This reflects what happens in cells sometimes when proteins regulate their own genes.

Last year his group developed a new technique that allowed them to precisely manipulate the positions of molecules on a surface. They coated beads with molecules and dropped them into extracts taken from cells. Since the beads were magnetic, they could be placed in precise positions and moved around.

Luis, Mark Isalan and Caroline Lemerle used the method to try to recreate what happens as molecules flow through a fly egg. "Instead of an egg we used a plastic chamber," Mark says. "And we didn't use fly molecules; what we were after was to demonstrate a principle. Developmental biologists have gotten a good idea of the types of gene interactions that create segments as the body forms. We hoped that by replicating similar types of interactions, we could produce stripe-like patterns in this little artificial system."

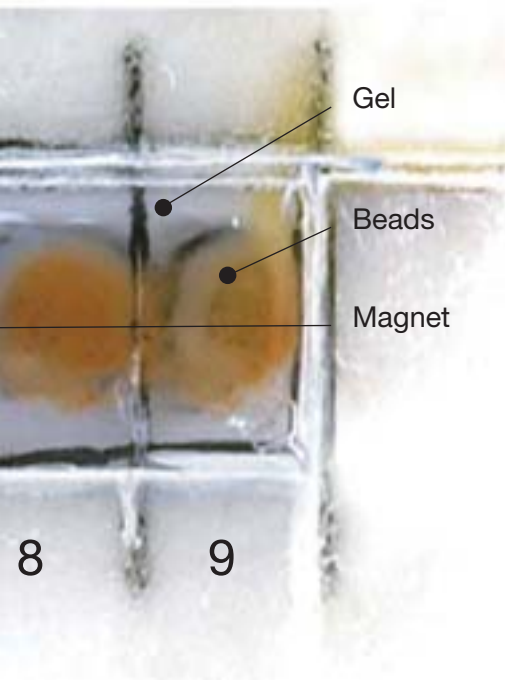
A very carefully controlled system would also allow them to test hypotheses that couldn't be directly tested in the fly, because there is not yet a technology to manage molecules so precisely in a living system.

The scientists wanted to reproduce two effects. First, it was necessary to create "slopes" of concentrations of different molecules. This would imitate the activity of a fly protein called bicoid, which is produced by the mother and enters one side of the egg. Concentrations of bicoid are highest at the point of entry, and they fall off as the molecule is transported across the cell.

The second phenomenon that had to be reproduced was a variety of ways of repressing the signal. "Recent evidence suggests that molecules along the route aren't simply reacting to different amounts of bicoid," Luis says. "The idea is that in some cases they are actively repressing it, and this creates new features in the landscape of the molecule in different regions of the cell. The question we hoped to answer was whether segments could arise simply from reactions to a gradient slope, or whether you also needed repression to get them."

The model system set up by the scientists, with the help of the Chemical Biology Core Facility, uses simple molecules that transcribe each other. This is much less complex than transcription in living cells, which involves too many molecules to be effectively controlled. Three signals were "broadcast" from the two ends of the artificial body, but they interact with each other. One signal (A) passes through the chamber without any interference – it has high concentrations at the broadcast point and lower in the middle. Signal B is also broadcast from both ends, but it is repressed by the first. So in places where A is high (at the ends), B is low. It is somewhat higher in the middle, where there is less A. Finally, signal C is produced everywhere. But since it can be repressed by either A or B, there is almost none of the protein at the ends of the tube, and somewhat higher concentrations in the middle.

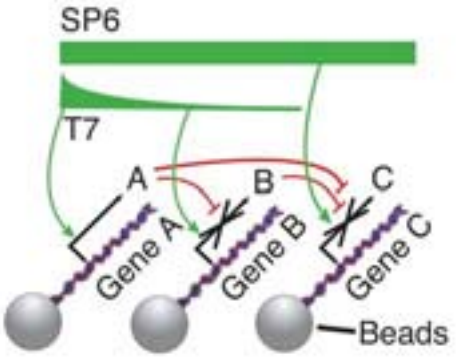
This was the simplest version of the experiment, and it already generated what Luis calls crude patterning behavior, because different amounts of the molecules in any particular location create different landscapes.

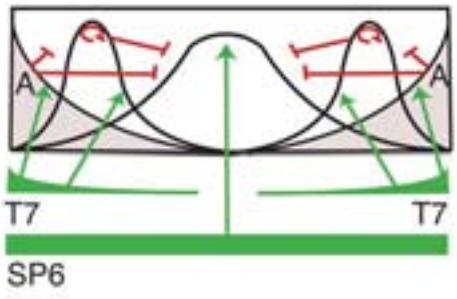


Left: attaching molecules to magnetic beads allowed the scientists to simulate the genetic activity that creates structures in developing organisms.

Right above: a schematic diagram of the molecules attached to beads.

Right below: Patterns of molecular expression that were observed as molecules regulated each others' expression.





Next the scientists began manipulating things. Using the beads, they added control elements at different places in the plastic chamber. Sometimes they removed repressors, in other situations they added various types of feedback loops. Each variation created a new genetic geography – in other words, a new "developmental pattern."

"What we discovered was that as more repressive elements are added to the system, there is a drop in the absolute amounts of proteins that get produced," Mark says. "On the other hand, patterns get sharper; boundaries become clearer."

Next the scientists created a computer model of the chamber and networks to manipulate the elements in an even more detailed way. One question concerned scale: the experimental chamber was 18 millimeters long, whereas a fly egg is about 35 times smaller. Since the amount of time it takes a molecule to spread through the egg, or the chamber, contributes to the formation of gradients, this difference in size might be important. But the simulations showed that this isn't the case. The same structures arise, Luis says – just at different times.

"What both the test tube and computer simulations show is that if you want to get clear patterns, genes have to exert control on each other," Luis says. "There need to be effects of cross-repression and feedback loops. This parallels what is being discovered about the interactivity of fly patterning genes. And although the system we have developed is much simpler than the living cell, it can already generate sharp and stable patterns. The earliest patterning networks in evolution might have functioned like this."

The next step for Luis and his colleagues will be to try to add layers of sophistication to the network, to imitate other types of regulation found in cells. What happens, for example, if a molecule is present that breaks down part of the signal in a controlled way? The experimental system they are using, Luis says, permits adding such features.

Doing those experiments will require crossing some very large geographical landscapes, though, as Luis packs up his lab for a move to Barcelona. After many years at EMBL, and five years as coordinator of the Structural and Computational Biology Unit, he will take up a new position as Director of the Systems Biology Unit and Vice-Director of the Centre for Genomic Regulation (CRG). ■

Evolution

THE THEORY OF EVOLUTION caused a revolution in our view of man's relationships to other species and the environment and thus opened entirely new modes of investigating human nature. Primatologists began a serious investigation of apes and chimpanzees, our closest living relatives, hoping for new insights into human biology, behavior and society. Yet equally important insights have come by comparing ourselves to strange, distant forms of life. The uniqueness of every species turns out to be mostly superficial, built on an immense architecture of traits shared with living beings that may be far away on the evolutionary tree. The similarities run so deep – from the forms and behavior of cells and tissues during development, to DNA sequences and the shapes of the molecules they encode – that every organism is a guidebook to every other. So many things are the same that it is sometimes surprising that species are different from each other at all.

The study of evolution is thus an investigation into similarities and differences. Why do small changes in the genetic code lead to dramatic changes in form? What does the information in a genome make possible, when put to different uses in various cell types, tissues and organisms? How did organs such as eyes and brains evolve from earlier sets of instructions and patterns? All of the cells in a person's body descend from a single fertilized egg, but they also descend from an ancient, one-celled ancestor. What traits in that ancestor would create a brain, and what genes would eventually lead to blood, muscle, and the immune system? Detlev Arendt and his colleagues think they may have a way to find out.

Answering these questions requires system approaches such as those described in the last chapter, and large-scale comparisons between the genetic programs used by cells in a wide spectrum of organisms. Nick Goldman's work suggests that sometimes creatures the farthest from our own lineage may reveal the most about ourselves. Christos Ouzounis believes that features of the last universal common ancestor of life on earth might help us find life on other planets.

New ways of comparing genomes have also led to fascinating new types of research, for example, metagenomic projects which Peer Bork ironically calls "GPS sequencing." Instead of sequencing specific organisms, these projects sample locations on the globe – from a cubic centimeter of farm soil to mossy clumps in the Sargasso Sea. Such studies are providing new information about the connections between the environment and biochemical processes; they may be the only way to truly measure the impact of our own behavior on the world. ■

Tobias Doerks, Christian von Mering,
Peer Bork, and Christopher Creevey,
up a tree, in descending order

A new tree of life

IN THE MARGINS of one of Charles Darwin's notebooks is a small, twig-like drawing – unimpressive until you realize that it represents an enormous intellectual leap, a milestone in human history. It is the first modern sketch of a tree of life, representing the fact that distinct species had common ancestors. For a century, naturalists had naming species and grouping them according to their similarities. Darwin suddenly understood that the classifications represented familial relationships.

Two decades later, another tree was meticulously composed by Ernst Haeckel, the great German naturalist and embryologist and a fanatical admirer of Darwin. Haeckel's chart attempts to synthesize the plant and animal kingdoms into a single genealogical record of life on earth. He got a lot of things right, but the tree goes back only so far. Once it got to one-celled organisms, he was stuck – scientists were only beginning to glimpse the amazing variety of such species alive on earth; they certainly didn't know enough to make a convincing phylogeny of events that happened before the divergence of plants and animals.

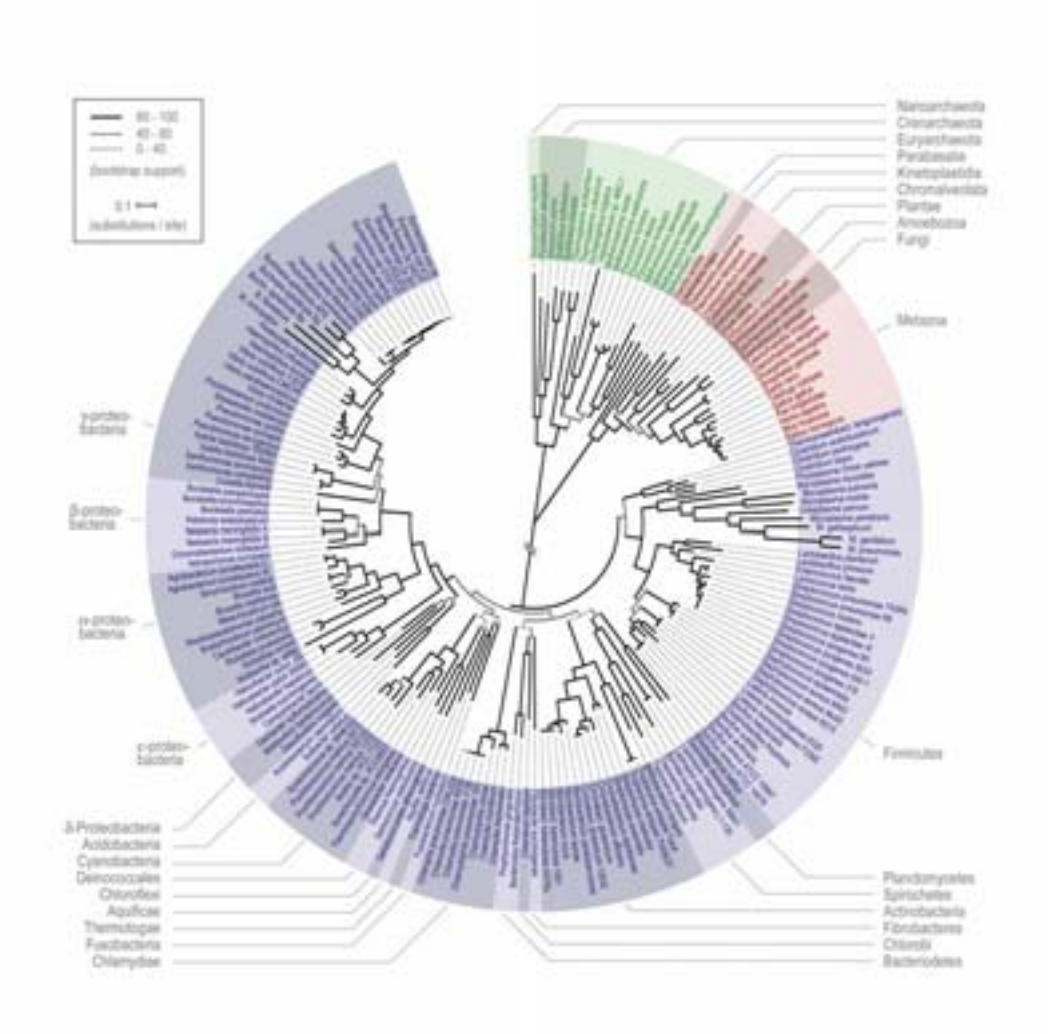
Since then, scientists have filled in branches and twigs, climbed down the trunk, and pushed deeply into the roots, drawing on the written record of evolution that is preserved in DNA. Still, questions remain, particularly in regards to the early history of life on earth. Peer Bork's group has now finished the highest-resolution tree of evolution that has yet been made. It will never be final – millions of species surely remain to be found, and the tree sprouts new branches and twigs faster than they can be described. But the current effort fills in many of the gaps, helping scientists sort out fragmentary clues of the existence of new organisms, and shedding light on the deepest roots of life.

* * *

Early in the earth's history, there existed an organism that would give rise to all the species known today. In 1994, Christos Ouzounis and Nikos Kyrpides gave this shadowy creature a nickname: *LUCA*, for the *last universal common ancestor*. Studies of DNA sequences taken from plants, fungi, animals, bacteria, and another form of one-celled organism called *Archaea* proved that it must have existed. But until recently, scientists could say very little else about it.

"Two things have changed," Peer says. "First is the immense amount of information we have from DNA sequencing – over 350 organisms have been completely sequenced, spread across the entire spectrum of life. This gives us a huge amount of data that can be compared to make a good tree and also to answer some questions about LUCA. Certain key genes can be found in all of them, and the chemical 'spelling' of these genes permits us to group them into families and historical relationships."

It also allows researchers to reconstruct hypothetical ancestors. A fundamental principle of evolution, called the *law of common descent*, states that if two organisms share features, it is almost always because they inher-



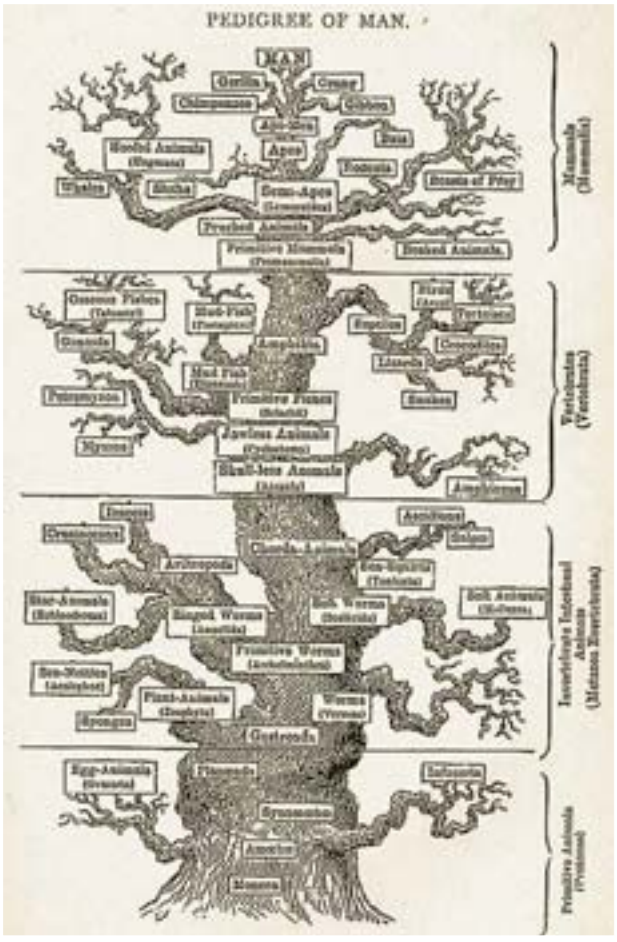
Left: the new tree of life, compared with Haeckel's (below)

Right: Francesca Ciccarelli

ited the characteristics from a common ancestor. So by comparing existing species, scientists can obtain a picture of more ancient forms of life.

"Over the past few decades, scientists have realized there is an important exception to this rule," Peer says. "Bacteria can swap genes with each other, and sometimes they can even steal a gene from a plant or an animal. Once that happens, they pass the gene on to their descendents. Such molecules have a completely different profile than genes inherited the normal way. It's like finding a branch from a tree that grows crosswise and fuses into another branch."

Peer says that attempts have been made to find such genes and eliminate them when building trees from DNA sequence data. But no one knew how often such events, called *horizontal gene transfer* (*HGT*), had happened, or had developed a convincing method for finding them. "For a while, it was almost as if the amount of data were increasing the problem rather than solving it," Peer says. "There were big debates, and the numbers of classifications were growing rather than reaching a consensus." Part of the problem lay in the fact that the work could only be done by computer in a highly automated way, due to the incredible amount of genomic data that had to be sifted through.



Francesca Ciccarelli, a postdoc in Peer's group, decided to tackle the problem of the tree anew and find a solution to the problem of the HGTs. She started by combing the complete genomes of 191 species for unique *orthologs* – genes that had clearly shared common ancestors. The task was difficult because it couldn't be completely automated. Francesca found 36 cases, five of which seemed to have been shuffled around through HGTs, so they were discarded.

Eliminating these from the analysis, the scientists could now build a complete tree by combining information from 31 genes. Peer was worried that some HGTs might have still have slipped in – a single mistake could spoil the quality of the tree. So the scientists put the computer to work doing some heavy lifting. The 31 genes were randomly divided into four groups. Trees were systematically drawn over and over again, for all of the genes in each group, with the exception of a single gene that was eliminated in each round. Then the results were compared. If the branches of the trees changed from pass to pass, an HGT was likely to be involved, and the gene was submitted to two more tests. In the end, the scientists found seven more candidates for HGTs, which they eliminated from their analysis.

The remaining information was combined into a supertree which was compared once again to trees based on individual genes in three different ways. "Any one of these methods on its own might have left a tree with some mistakes," Peer says, "but by combining them, we're able for the first time to compare the three domains of life – bacteria, *Archaea*, and eukaryotes – not only in terms of what has branched off from what, but also in terms of branch length. In other words, we can distinguish between fast- and slow-evolving species."

The results clear up some old controversies, for example, a debate about the very early evolution of animals. Some trees in the past proposed that the vertebrates (which include humans) split off from another branch which would remain united for a while before splitting into separate branches for worms and insects. The new version groups things differently: vertebrates and insects split off from the worms together, and diverge from each other later.

The higher resolution of the tree is also important, Peer says, because of *metagenomic* studies which are underway to sequence all the genes found in environments such as farm soil or ocean water. His group has participated in several such projects. "Most sequencing approaches start with a given organism and work through its whole genome systematically," he says. "Metagenomics is sequencing a *place* – like a GPS coordinate. In many cases we recover fragmentary traces of thousands of genes, and have no idea what organism they come from. Often these molecules represent creatures that have never been seen before." The breadth and detail of the new tree will allow scientists to make much better guesses about where such fragments fit in and what types of living beings they belong to.

Has the living world been fairly split up into major branches, limbs, and twigs, or have we overemphasized the prominence of our own lineage? A close look at the new tree shows the latter seems to be the case. The eukaryotes, which include yeast, plants and animals such as ourselves, are so visibly different from one another that scientists have pushed them apart from each other on the tree. Genetically speaking, however, the species are often much more closely related than many single-celled forms of life.

"Smaller genomes evolve faster," Peer says. "There isn't a single organism that has been sequenced that is both evolving fast and has a large genome. It suggests that some of the simplest species around have ended up that way because they have been pruned down. Evolution isn't always about acquiring complexity."

The study also yields insights into LUCA. "One very big question has been what the earliest bacteria were like when they split off from the Archaea. Bacteria are grouped into two classes, called *gram-positive* and *gram-negative*, based on features of their membranes. The new tree reveals that gram-positive bacteria evolved first. And if you look at their repertoire of genes, they seem to be suited to a very hot environment. The first Archaea were discovered in hot ocean vents, and most of the species alive today are thermophilic. It strongly suggests that LUCA was, too." ■

Christos Ouzounis
at EMBL-EBI

Extraterrestrials and the origins of life on earth

LIFE ON EARTH exists under the most extreme conditions – in ice, boiling water, acid, the water cores of nuclear reactors, salt crystals and toxic waste. Organic molecules have been detected in interstellar space, and some scientists believe that the bombardment of the early earth by meteorites brought along proto-organic material that may have played a role in the origins of life.

Whether or not that's true, Christos Ouzounis and his colleagues believe that trying to understand LUCA, the last universal common ancestor of life on earth, may be helpful in searching for life beyond our planet. If living organisms exist on the moons of Jupiter or elsewhere in the solar system, Christos says, they will have evolved under extreme conditions, and might share some of the characteristics of early *extremophiles* on earth.

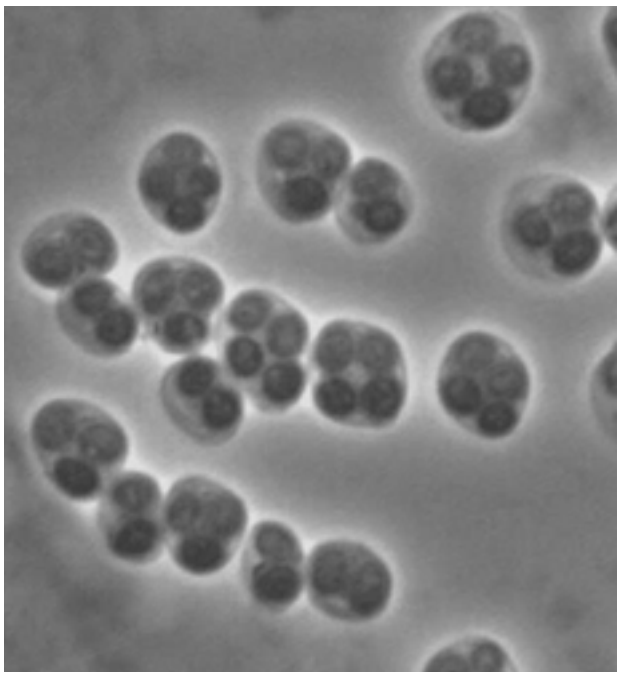
If living organisms exist on the moons of Jupiter or elsewhere in the solar system, it might share some of the characteristics of early extremophiles on earth.

Christos and his colleagues pulled together some of these ideas in a paper that was published just as he left EMBL to return to his native Greece. Christos was among the first research group leaders at EMBL-EBI, and the study pulls together several of the themes that his lab has worked on over the past nine years. Like the paper from Peer Bork's group (see previous story), the scientists drew on data from complete genomes and weeded out cases of horizontal gene transfer to list genes and biological functions that must have been present in LUCA.

"Past attempts to study these questions have been limited by horizontal gene transfers – HGTs," Christos says. "A lot of controversies have emerged, such as whether the organism was thermophilic, and whether it was simple or already quite complex. Most of the answers so far have been put together from partial information; our approach is to compare the content of entire genomes."

To do so, the scientists drew on a computer method called GeneTRACE, which they developed a few years ago. They applied three different techniques to build family trees of the organisms and to search for common features. "The goal was to say what types of genes LUCA had, and to try to determine what their functions might have been," Christos says.

The study revealed 669 common *ortholog families*, or groups of equivalent genes, which are known to participate in 561 biological functions. "This is a higher number and it reflects a higher complexity of biological functions than earlier studies have predicted," Christos says. "Some of those studies have been based on an analysis of the 'minimal genome' necessary to sustain simple organisms – a list of the minimal number of genes that such organisms need to survive – again, assuming that things have been getting more complex. LUCA might not have been so simple after all."

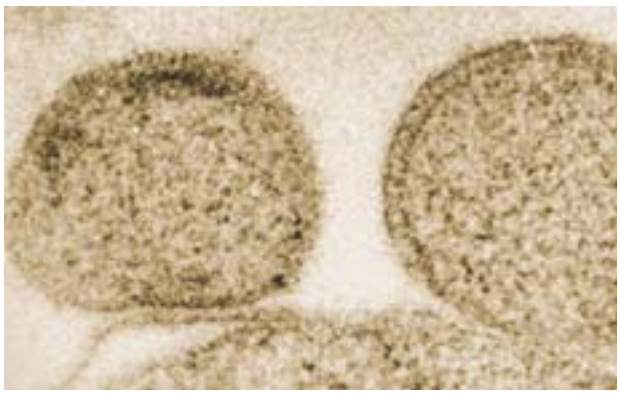


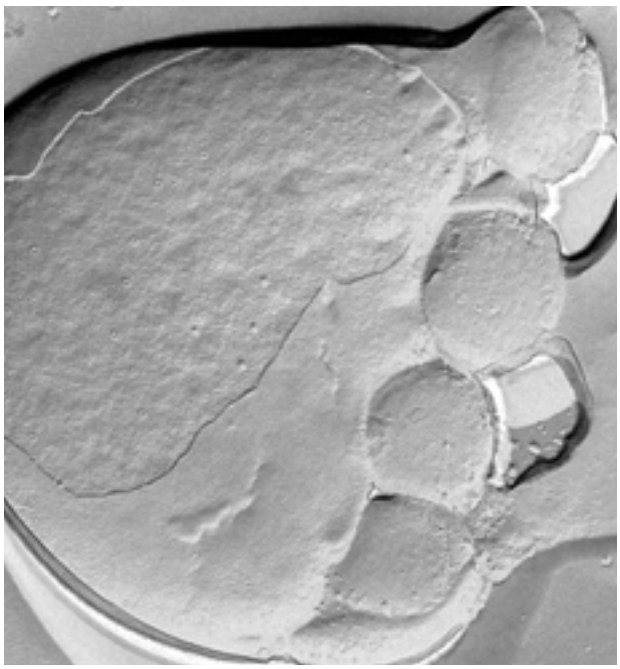
Ancestral processes included molecules that handle DNA replication, repair, and modification; machinery necessary to transcribe genetic information into RNA and to translate that into proteins; cellular processes such as signaling and protection from extreme changes in temperature; well-developed membranes and mechanisms for shuttling molecules between them; and a sophisticated set of metabolic enzymes which can produce nucleotides, amino acids, sugars and the components of membranes. LUCA probably didn't have many ways to process RNAs, which is remarkable, Christos says, considering the importance of this process in biological systems.

It reflects a higher complexity of biological functions than earlier studies predicted, LUCA might not have been so simple after all.

Another surprise was the list of molecules involved in electron transport – snatching charged particles from the environment in order to obtain energy, drive chemical reactions, and perform other subtle functions within cells. "Some of the components of this system seem to be suited to working in an environment with oxygen," Christos says. "The traditional view has suggested that air-breathing organisms evolved late, but recently there have been suggestions that there might have been some earlier."

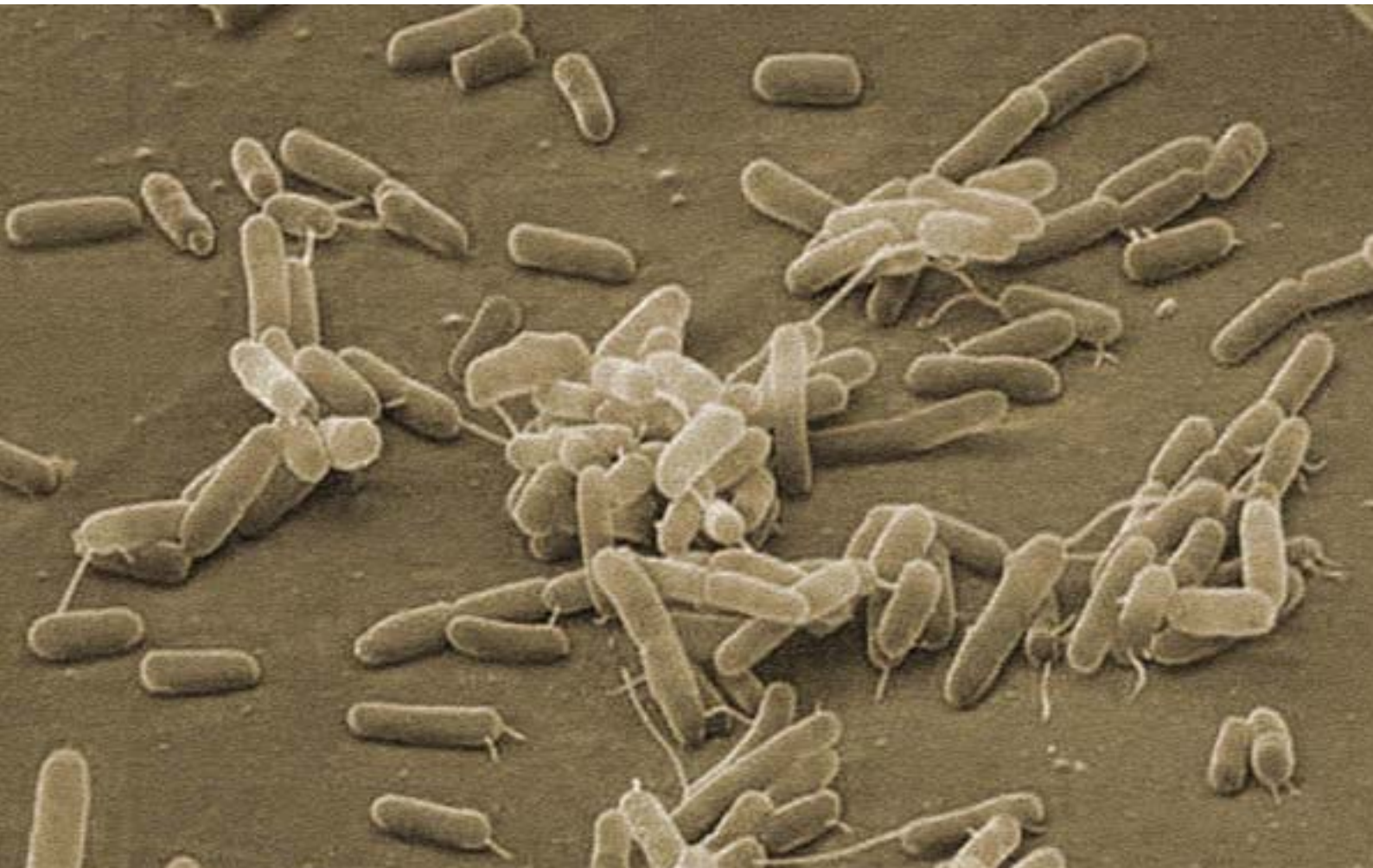
LUCA was not by any means the earliest form of life on earth. No one knows exactly when it lived, or how life





Eukaryotic cells (yeast, left), Archaea (above and right), and bacteria (below) make up the three kingdoms of life; comparisons of genomes from these three branches are giving scientists insights into the last universal common ancestor of all life on earth. The images of Archaea are reproduced by permission of the MacMillan publishing group: Huber et. al. (2002). Nature. 417(6884):63-7

developed before it arose. But because all living organisms descend from it, the approach of comparing genomes can't reach back farther than this creature. That history was clearly very complex, Christos says, because of the high complexity of LUCA itself. "By the time this creature lived, a great many of the biological processes that drive our own cells had already developed," he says. "Those basic processes were apparently very successful, because they have stood the test of time – billions of years of evolution, in an enormous variety of creatures. It may sound odd to compare ourselves to a one-celled organism that lived so long ago, but we may not be that much different after all." ■



Detlev Arendt, Peer Bork and Florian Raible, looking for the fastest and slowest evolvers

The tortoise, the hare and the worm

Platynereis dumerilii

IF THE FIRST genetically modern humans lived about 100,000 years ago, as many paleontologists believe, probably only six or seven thousand generations separate us from them. The comparable ancestor of a fruit fly – grandfather to the seven thousandth degree – lived only 230 years ago; 100,000 years from now, it would celebrate its three millionth generation. Because there is a connection between frequencies of generations and the mutations that creep into genomes, it should be little wonder that flies have evolved much faster than humans.

In fact, says Detlev Arendt, if evolution were a race, the entire branch of life that led to our species would count among the slowest. Detlev and his group in Heidelberg have been studying different animals in hopes of gaining insights into a significant moment in evolution. They want to describe one of the earliest animals, which no longer exists and from which no fossils remain. The organism is called *Urbilateria* because it is the last common ancestor of humans and flies and was symmetrical along the axis from head to tail.

Most work in laboratories has focused on insects such as the fruit fly, or mammals such as mice. Detlev's lab instead focuses on a tiny marine worm called *Platynereis dumerilii*. It's a useful model; not only does the animal only breed every three to four months – similar to common species of fish – but *Platynereis* also belongs to an animal branch that split apart from the other lineages shortly after Urbilateria lived, and studying it has allowed Detlev to gain a broader perspective on ancient events in animal evolution.

Most of his lab's work has focused on tracing the evolutionary origins of the brain and other organs. Recently Florian Raible, a postdoc who splits his time between the groups of Detlev and Peer Bork, began to investigate *Urbilateria* genes using the law of common descent: identical features shared between animals from different ends of the animal tree most likely have been passed on from the last common ancestor. While there is no complete genome for any *Lophotrochozoan* – the branch of life that includes Platynereis, mollusks and flatworms – Florian and his colleagues think that the sample of sequence information from Platynereis is sufficient to make a serious investigation of *Urbilateria* genes. In the process, the scientists have dispelled a case of bias in the way we think of evolution.

* * *

Humans are more complex than flies – and we have more genes. Is there a connection? Does our biological and social sophistication stem from some basic feature of our genome, for example the number or structure of genes it contains? Some species of rice have twice as many genes as humans. Perhaps the number matters less than the complex nature of genes themselves. In animals, protein-encoding regions of the genome (*exons*) are interrupted by non-coding regions called *introns*, which are removed as genetic information is transformed into proteins.

"At first glance, comparisons of genomes seems to reinforce the hypothesis that genes have become more complex through evolution, and that gene complexity has something to do with the complexity of organisms," Florian says. "In our comparison, human genes had on average 8.4 introns, whereas the average for flies and their relatives is between 2.4 and 5.4. If it was true that the earliest animal genes were very poor in introns, is the story of animal evolution one of acquiring more and more introns?"

Florian says the scientists were lucky to be able to draw on data from the honeybee genome, which was recently completed. "It gave us another point of comparison. Like other insects, bees have fewer introns that mammals. But we discovered that 25% of those they *do* have can be matched one-to-one with human introns, and that's a higher number than in other insects. Finding the same intron in bees and humans means that the ancestor had it, too. The sequences weren't invented twice."

But although bees have modified their genes less than the flies, they are still members of the same group. In con-

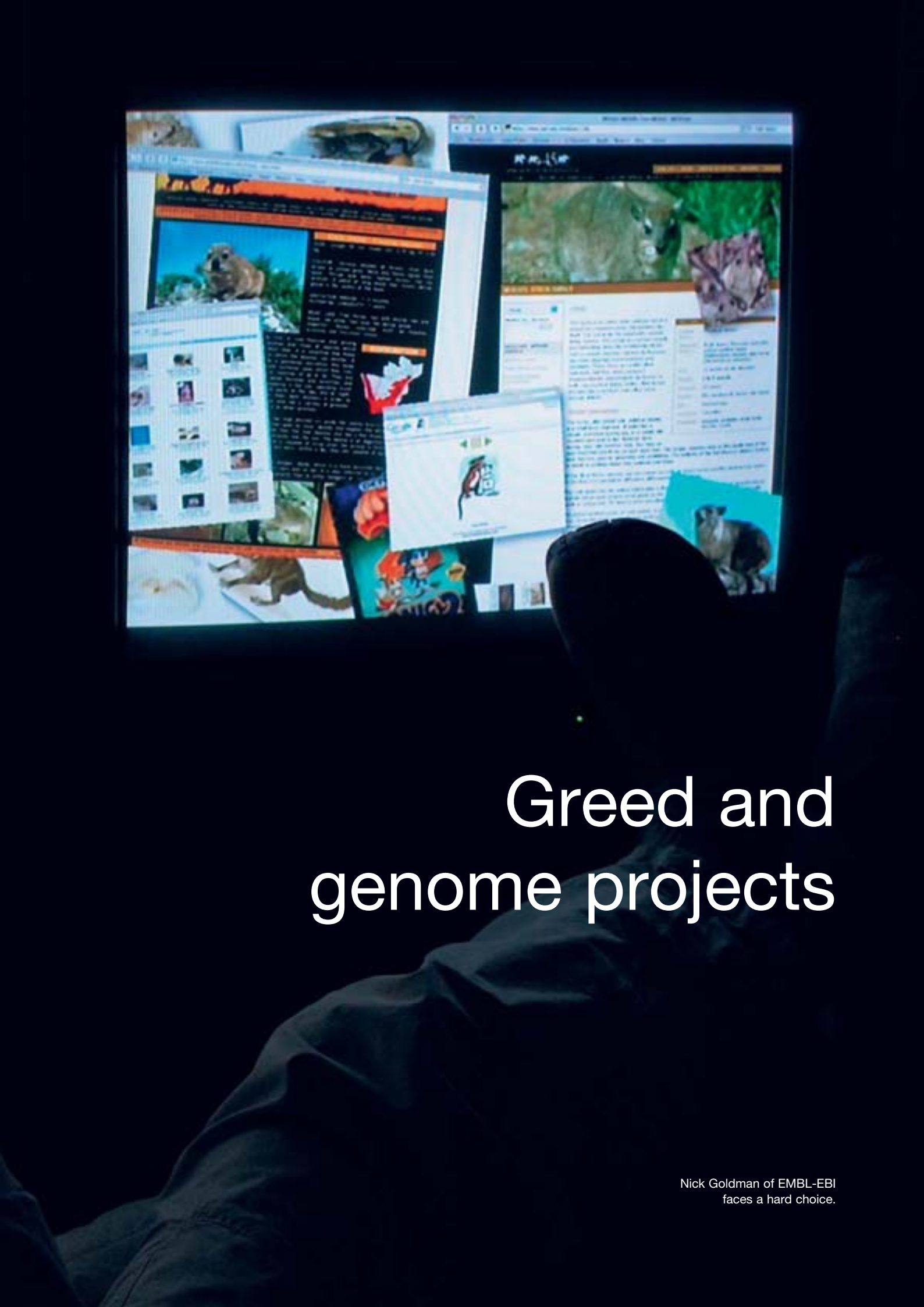
trast, *Platynereis* represents a large group of invertebrates not related to flies, and at the same time it has a relatively slow generation time. Therefore, a look at the *Platynereis* genome might provide much better insights into the true nature of Urbilateria genes.

Sequence data from Platynereis revealed 30 genes with clear matches in other animals. They contained an average of 7.8 introns per gene, slightly lower than the human average, but much higher than that of insects. Three-quarters of these can be matched to introns in the other branches, meaning that the *Urbilateria* had them as well.

"Most are shared with humans or fish," Florian says. "So then we did the opposite – we asked how many of the human introns are found in other descendants of the Urbilateria, and when we included *Platynereis*, we found that at least two-thirds of the introns we analyzed must have also been present in the ancestor, at precisely the same positions."

Did the pattern established for introns also hold for exons, the protein-encoding regions of genes? The scientists obtained a set of 442 proteins present in Platynereis that had clear relatives in the other branches of life. Again, Platynereis and humans were more closely related than insects and other organisms on their branch. While this small worm and our own species are separated by a vast evolutionary time – more than 600 million years in both lineages – they are less separated in evolutionary *space*: because they evolve slowly, they are both more similar to their common ancestor (and thus to each other) than most other species are.

This turns a bias on its head– that the complex mind and behavior of our own species might have something to do with the complexity of gene structure. Humans aren't special because their genes are becoming more complex; rather, this complexity is a heritage of our ancient past. Insects such as fruitflies and honeybees are evolving quickly, and in the process they have acquired "simpler genes" – losing introns. So to understand early animal evolution, Detlev says, we'll have to look in some unusual places. A simple marine worm may tell us more than the well-studied fruit fly. Rather than looking at the speediest evolvers, we'll have to pay more attention to the slowest – like ourselves. ■



Greed and genome projects

Nick Goldman of EMBL-EBI faces a hard choice.

Fabio Pardi

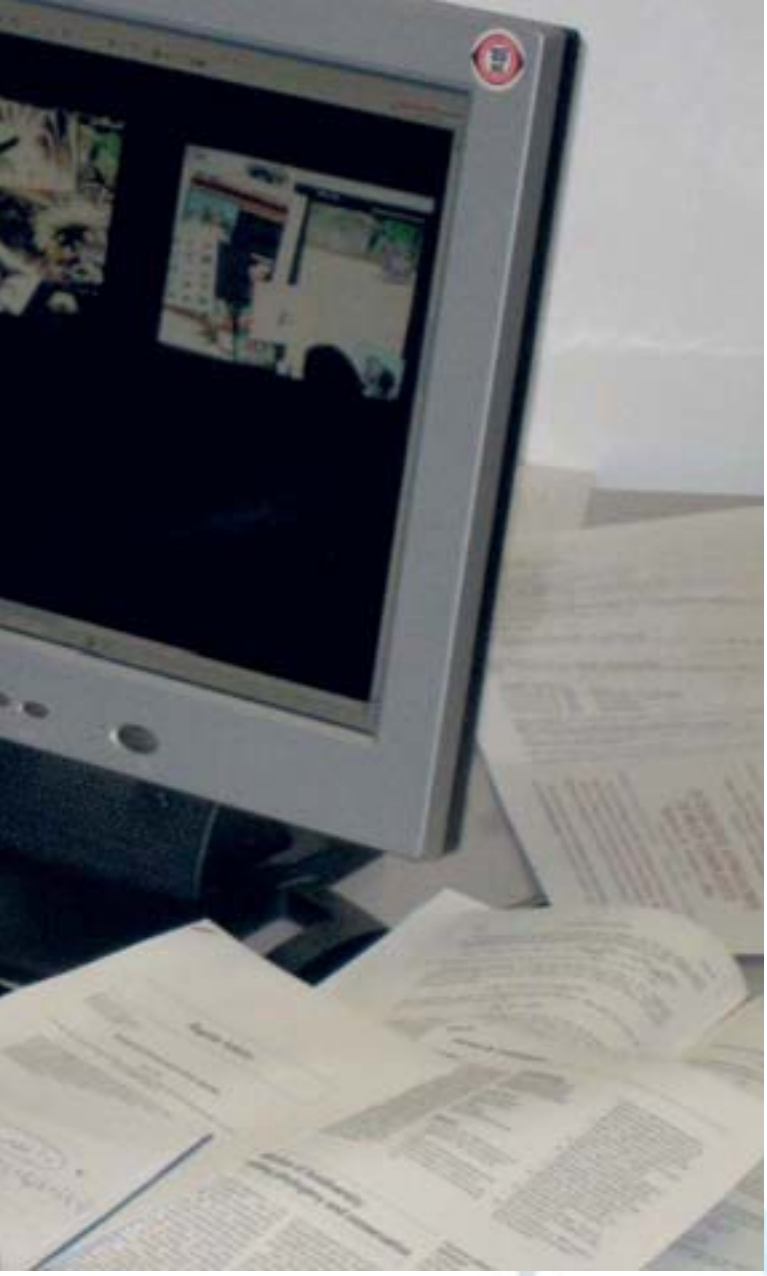
IT'S NOT WHAT YOU MIGHT THINK – Nick Goldman's group at EMBL-EBI isn't looking for ways to get rich off the DNA sequences of wallabies, or aardvarks, or strange types of fish only found in the deep sea. The "greed" in this case is a type of computer algorithm that Nick and PhD student Fabio Pardi have put to use in thinking about genomes.

Actually, there is a financial angle to the question that Nick and Fabio have been thinking about: What genome should be sequenced next? "Large-scale DNA sequencing projects consume lots of resources," Nick says, "and typically involve several sequencing centres. Each of them invests considerable time and money, so it's in everyone's best interests to pick the genome – or set of genomes – that will produce the best payoff."

Here he means information rather than monetary rewards. Because every organism on earth is related through evolution, every completed genome sheds light on every other. One way is by pointing out new genes. The human sequence has been finished for a couple of years, but we still haven't found all of the genes encoded in our DNA, partly because we don't know exactly what to look for.

"Knowing that humans are closely related to mice, for example, meant that their two genomes could be used as a sort of guidebook to each other," Nick says. "A particular gene might have unusual features and you wouldn't find it just by looking in the human genome, or that of the mouse. But if you look at both simultaneously, it suddenly pops up."

Of course scientists aren't only interested in human genes, he says. Although these are likely to have a more direct impact on medical research, discoveries in remote creatures – such as a one-celled bacteria living on the ocean floor – can reveal a new biological process that is ultimately more important. Thus picking the right genome is a question of context. If you want to find something specific, you pick a well-studied organism close to



another creature that you're interested in. EMBL scientists working on malaria-carrying mosquitoes profited immensely, for example, from the well-studied genome of another insect, the fruit fly.

But what if you want to cast the widest possible net, and collect as much *new* biological information as possible? Is there any way to predict what organism, or what groups, will be most helpful? That's the question that Fabio and Nick tackled. They were well prepared to take it on; one of the specialties of Nick's group is comparing different types of evolutionary analysis using statistics and mathematics.

"With the advent of massive amounts of DNA sequencing, most scientists completely changed their perspective on evolution," Nick says. "Instead of building trees of organisms based on their morphology – questions like body size and shape – they were now using information about genes. Those kinds of questions have to be tackled in new ways, and there are a lot of methods out there. We've developed tools to help referee between different models."

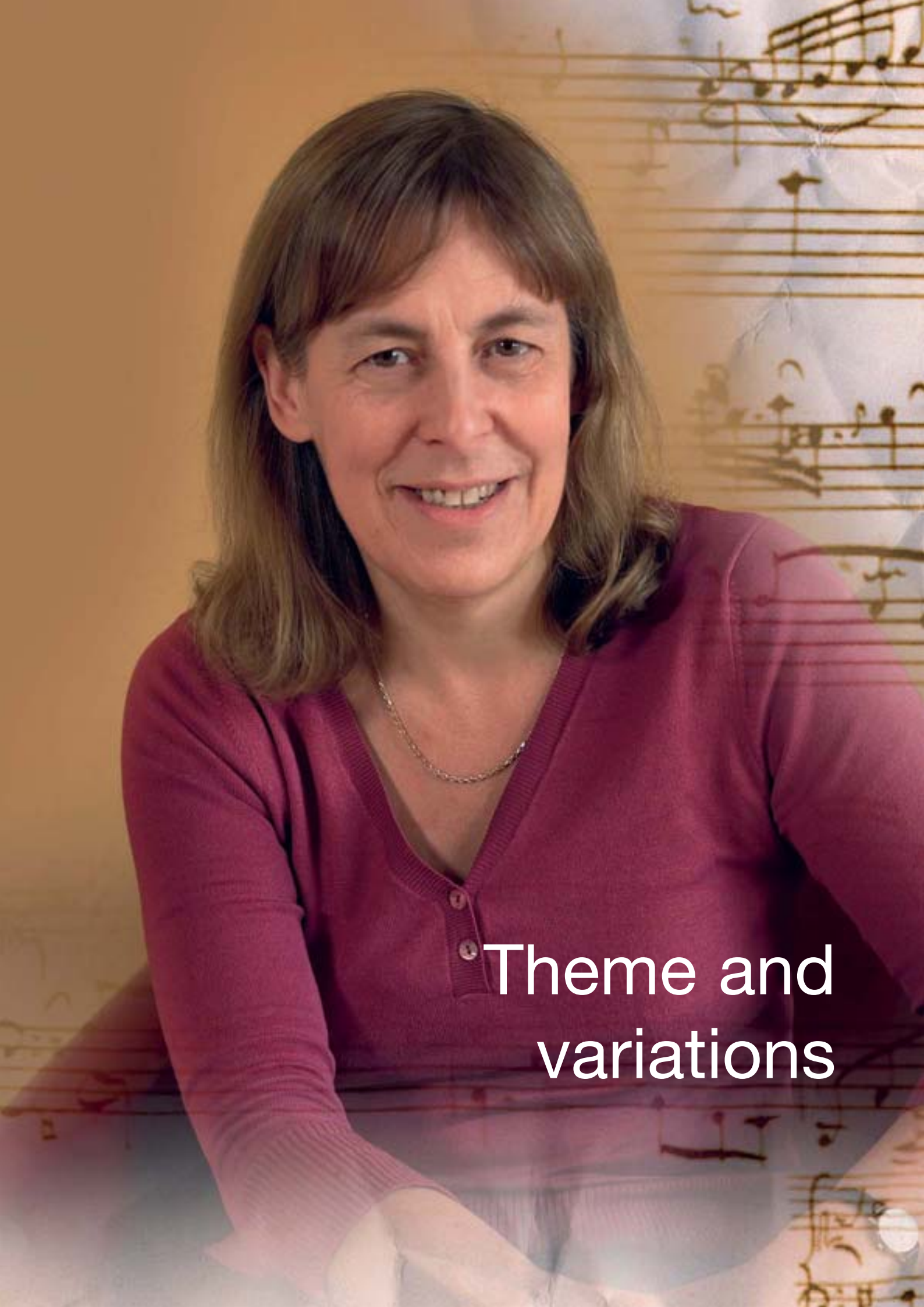
Before performing the calculations, Fabio says, he and Nick had a gut feeling about genome projects: the community ought to sequence several organisms at a time, hoping for the most new information, and the best results would come from big, collaborative efforts. "We translated this question into mathematics," Fabio says. "It sounds like this: Given that we have completed a certain number of species, and now we assume that we have the money to sequence a certain number more, what's the best strategy to obtain the richest amount of divergent information?"

The alternative to a multi-pronged, large-community attack on several genomes would be for one sequencing centre to pick one genome – the hot species of the moment. A "greedy" method would target a single organism that was most unlike the collection that had already been sequenced, and a single centre would do it alone. When that one was finished, someone would redo the calculation, once again pick the most "distant" genome, then focus all efforts on it.

"Contrary to our expectations, in this case the 'greedy' method works best," Fabio says. "No matter what other set of species you might pick, you get the best results by repeatedly tackling the one most divergent genome. And this strategy works very well for other types of decisions as well, for example, if you're considering a new model organism to bring into the laboratory, or if several species are about to go extinct and you can only save one."

One reason for the surprise is that in computer science, Nick says, greedy algorithms are rarely the best solution to a problem. "This strategy can be applied to a lot of different situations – for example, the 'traveling salesman' problem in which a person must choose the most economical route to visit many different cities. In that context, however, greed is not the best answer."

What about those budget-saving measures? Sequencing projects are generally decided upon and organized in large meetings, after lots of debate. If the goal of the genome project is to produce the most divergent set of data possible, Nick says, that will no longer be necessary. "A phone call will suffice," he says. "Just make sure you aren't going to do the same organism as somebody else." ■



Theme and variations

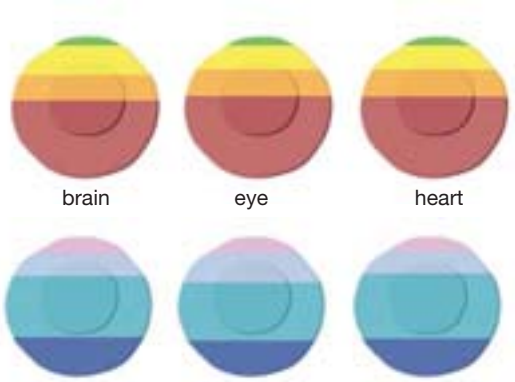
Janet Thornton, Head of EMBL-EBI, and
a theme and variations by J.S. Bach

CHARLES DARWIN spent nearly a decade dissecting some of the strangest creatures on earth: barnacles. Upon prying open the shell of the female of one species, he found it divided into compartments which served as "apartments" for her tiny mates: male barnacles that would live their entire lives without seeing the outside world. They were much simpler than the females – Darwin called them "bags of sperm" – almost like organs of the female rather than organisms in their own right. Barnacles come in such a bewildering array of shapes and forms that classifying them and mapping their evolutionary relationships to one another was incredibly difficult. Which species alive today is the most representative, the most closely related to the first ancestral barnacle? It took him over eight years to find an answer.

Just as every barnacle can be traced back to a common ancestor, each cell in our bodies is a direct descendant of the first cell that lived on earth. Our hundreds of cell types sometimes have dramatically different forms – from the doughnut-shaped red blood cells to the sprawling tree-like shapes of neurons or tadpole-like sperm. Even though they share the same genome, they achieve diversity by drawing on different parts of it. The genome itself is much larger and more complex than those of our early ancestors. Which part of it is the most closely related to them? Is there one type of cell more similar to the ancestor than others? What can we discover about our evolutionary history by looking at the patterns of genes that are switched on in different cell types?

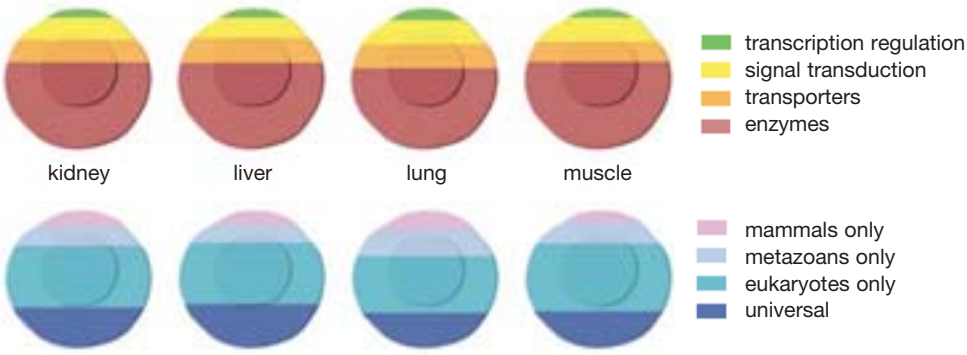
Janet Thornton's group at EMBL-EBI finished a project during the year that attempted to answer these questions by studying the genes active in different types of mouse cells and comparing them to genes common to all forms of life. "The way that we phrased this question was to ask if there is a relationship between an animal protein's origins – in other words, at what stage in evolution it appeared – and the role that it plays in cells and tissues," Janet says. "To try to answer this we can draw on the genomes from over 20 higher animals and a wide range of new experiments, and the way EMBL-EBI has set up its data resources makes it possible to carry out this type of study. Eight animal genomes have been added to Ensembl, our public collection of animal genomes, just in the last year. For many of these we have data about the tissues and circumstances under which particular genes are drawn upon. ArrayExpress, the database for this information, now holds data from over 1300 studies performed in 70 species."

Shiri Freilich, a PhD student in Janet's team, compared data from experiments on various mouse tissues with the help of Tim Massingham. The scientists started with a list of 6206 molecules which have clearly-related forms in a



wide range of organisms: 3516 of these occur in both single-celled life forms and animals; the other 2690 are found only in animals. "First we sorted proteins into functional categories: enzymes, signaling molecules, transporters, and transcription factors," Shiri says. "We studied what organisms had each protein, and in which tissues they appeared."

There is quite a variety in the way cells use their genome. Cells in the mouse eye use about 55% of their genes in the data set, whereas only 35% of the genes are active in muscle tissue. Shiri and her colleagues compared expression data to survey what types of proteins are encoded by these genes. "All tissues use all the functional types tested," she says. "It's not like one tissue is using a disproportionately high number of enzymes and another uses very few. Cells from all tissues produce about 60% enzymes, 20% transporters, 15% signaling molecules and 5% transcription regulators."



Top row: How many genes of specific functional types are found in certain types of cells?

Bottom row: What are the evolutionary origins of the genes found in those cell types?

This doesn't mean, though, that an ancestral one-celled organism had the same balance. "If you look at the evolutionary origin of the molecules, you find that our tissues contain about 25% universal proteins – found in every living cell," Shiri says. "40% of the proteins are common to all eukaryotes, 20% are found in multi-cellular organisms, and 15% are unique to mammals."

The next task was to look at individual molecules – were all tissues using the same universal proteins? It might be expected that every cell uses the same basic set of universal proteins, which is then supplemented with a tissue-specific set. To their surprise, Shiri and her colleagues found that less than one-third of the enzymes and transporters were expressed in all tissues. Even fewer – only one-tenth – of the signaling molecules and transcription factors were shared.

This upholds a prediction that the core set of genes needed in all types of cells also tend to be the oldest. Molecules that are used only in specific tissues tend to be recent evolutionary inventions, mostly involved in signaling and activating specialized genes. They have helped drive the development of highly-specialized tissues in animals by creating new ways for cells to use their genomes. As the work of Eric Karsenti and his colleagues has shown (page 24), a variety of structures can be created using the same molecular building materials. By changing its contents, the cell influences which molecules work together, what types of stimuli it can respond to, and eventually its shape and form.

Shiri says that two major messages emerged from the study. Since all of our cells descend from a one-celled ancestor, it's logical that the most ancient genes appear in nearly every species – genes which have survived for so long, in so many species, must be crucial. Genes with particular functions that evolved later, like transcription factors and other regulatory devices, tend to be tissue-specific. Late evolvers are usually not needed by all types of cells.

"That's no surprise," she says. "But while it's a clear trend, it isn't at all an exclusive relationship, and there are exceptions. If you classify proteins only by their age, and not their functions, the picture is much more diverse. Some proteins which appeared late in evolution, and are found only in animals, are produced in all tissues. And the flip side of this is that some ancient proteins have now become specialized in certain species. They're only found in particular tissues."

Thus the generic animal cell is different from its ancient ancestor. It has acquired new functions and capabilities, thanks to the recent invention of new molecules. It has also found new uses for old molecules: some of them are no longer needed in all cells and have specialized instead. Each tissue has found its own uses for what it has inherited. Evolution is built on themes and variations: new organisms are the product of changes in the genome, but equally important is the way they mix and match the ingredients already at hand. ■

Science and Society at EMBL

6th EMBL/EMBO joint conference

On 28-29 October 2005, the EMBL/EMBO annual Science and Society conference was on the theme of "Science and Security". Organised over two days, this multidisciplinary event brought together scientists, philosophers, social scientists, policy makers, consumer associations and members of the public for inspiring debate. The annual conferences have become an important forum for breaking down communication barriers between scientists and non-scientists and for promoting mutual interest, understanding and dialogue on subjects that concern everyone.

The first session of the 2005 conference focused on how to prevent the misuse of biological knowledge. How does our perception of risk match reality? Can measures be applied to restrict access to only those with "good intentions", and to control the application of biological knowledge, products or processes? These were some of the questions that the speakers in the first session addressed.

In the second session, attention turned towards security and freedom of research. Can publication and exchange of information or materials in the academic world endanger national security? Should restrictions on the freedom of research be applied? These and related issues, which

Left and below left: Attendees and panel members shared the discussion at the 6th EMBL/EMBO joint conference.

have grown in importance following a number of recent incidencts – such as the increasing restrictions on academic institutions that have resulted from heightened security consciousness after the 11 September 2001 terrorist attacks, for example – were presented and analysed in this session.

The third main topic of the conference concerned the production of new knowledge and technologies for identification. Here, the audience learned how the shape of our ears is almost as reliable as our fingerprints for physiological identification, and the pattern of our irises is even better. Iris scan technology is already used in some airports in Europe and the USA to speed up passenger transit. Speakers in this session presented the state of the art in biometrics, biological forensics and the science of identification, after which data were presented for assessing the advantages as well as disadvantages of these new technologies and how they are viewed by the public. There was a lively discussion in which panelists and members of the audience reflected on how these advances are viewed by the public and what the future may hold in terms of perceived needs and corresponding innovations.

The last part of the conference focused on information technology in the knowledge society. More personal information is collected from us now than ever before in history – for either our convenience or security – and the trend is continuing, with concern for security never greater. What is the state of the art of information interpretation/annotation technology? How can we ensure that this information is used properly? The speakers and panelists in the concluding session addressed the vital question of how to strike a balance between protecting society and protecting the individual. Participants discussed how information should best be applied at personal, societal, national and international levels, and how it should best be regulated to promote security without restricting our freedom.

A free copy of the DVDs with a selection of talks from the conference is available upon request from stefanss@embl.de.

The 7th annual EMBL/EMBO joint conference on Science and Society, "Genes, Brain/Mind and Behaviour", will take place at EMBL Heidelberg on 3-4 November 2006. For more information, visit www.embl.org/aboutus/sciencesociety/conferences/2006/scope06.html. ■

EMBL Science and Society Forum in 2005

Heidelberg Forum in 2005

January

13 January
Der ungeborene, der alte und der kranke Mensch: Fragen zur Menschenwürde
Prof. Dr. h. c. Rüdiger Wolfram

25 January
Impact of agricultural biotechnology on biodiversity, myths and facts
Klaus Ammann, Botanical Garden, University of Bern, Switzerland

February

11 March
Religion as a Natural Phenomenon
Daniel C. Dennett, Professor of Philosophy, Tufts University, USA

March

April

8 April
The evolution of female promiscuity
Olivia Judson, Imperial College, UK

May

25 May
Über die neurobiologischen Grundlagen des Bewusstseins
Professor Christof Koch

June

EMBL Science and Society Forum seminars were launched in 1998 as an initiative to promote awareness of the impact that work within the life sciences is having on society. Events are open to the public, and every year sees a fascinating line-up of guest speakers.

July

August

The Heidelberg Forum on the Biosciences and Society is an initiative sponsored by the Manfred Lautenschläger foundation and run by EMBL, DKFZ, the Center for Molecular Biology and the Medical Faculty of the University of Heidelberg to promote public understanding of science.

September

9 September
Science in a democratic society
Philip Kitcher, Professor of Philosophy, Columbia University, USA

October

31 October
Missbrauch der Wissenschaft: Doping im Sport
Professor Werner Franke

December

Science and Art summer festival

On 15 July 2005, a Science and Society mini-symposium on "Art in Science in Art" looked at ways in which science and art can complement and inspire one another.

The afternoon began with talks from speakers with links to both disciplines: Suzanne Anker from the School of Visual Arts in New York; Arthur I. Miller, philosopher and author of *Einstein, Picasso: Space, Time and the Beauty that Causes Havoc*; and Christa Sommerer of the University of Art and Design in Linz. They demonstrated how artists and scientists have long acted as catalysts for creative innovation, often allowing previously held ideas or theories to be realised or surpassed. If society expects scientists to provide new knowledge and solve practical problems, artists are expected to bring a personal vision to bear on the experiences of life. Professor Peter Weibel of Karlsruhe's Zentrum für Kunst und Medientechnologie then chaired a panel discussion in which representatives of EMBL and the speakers talked about creativity at the interface between science and art. During the coffee breaks, attendees could enjoy a colourful exhibition of artworks created by members of the EMBL community, which decorated the Operon Foyer. Exhibits included sculpture, jewellery, nature photographs, microscope images and videos, demonstrating that scientists can be artists too, and that research can sometimes produce unexpected and spontaneous artworks. ■

Genes and Behaviour

A one-day symposium in Monterotondo

Modern biology has been shedding light on how genes relate to behaviour. Traditional research strategies in human behavioural genetics include studies of twins and adoptees using approaches designed to sort biological from environmental influences. In their laboratory work, biologists use selective and systematic knockout technologies to assess links between specific genotypes and phenotypes.

At the same time, the discovery of DNA polymorphism has resurrected research into human genetic variation that takes samples from distinct populations that are associated with particular behaviours. Analytical methods for assessing disease risk factors in the interactions between genes and the environment have vastly improved. Understanding these factors and their interactions could lead to major improvements in diagnostics, preventive medicine and therapeutics.

The symposium in Monterotondo on 9 December 2005 started with two experimental biologists – Cornelius Gross, Group Leader from EMBL Monterotondo, and Peter McGuffin of King's College, London – who gave talks about their work in progress relating to the genetic basis of depression and anxiety. During the second half of the programme two social scientists – Nikolas Rose of the London School of Economics and Karin Knorr Cetina from the University of Konstanz – talked about the social and ethical implications of these research areas within the life sciences. Judging from the response of the audience, the meeting achieved its goal of promoting enthusiastic dialogue between experimental biologists and social scientists.

The symposium concluded with an open discussion on how to assess and interpret the genetic component in the relative risks people face throughout their lives of developing specific patterns of behaviour detrimental to their wellbeing. ■

EICAT – EMBL International Centre for Advanced Training

Opening New Perspectives

Anne Ephrussi, Matthias Haury and Matthias Hentze discuss ATC floor plans

Since EMBL opened its doors in 1974, advanced scientific training has been one of the cornerstones of its overall mission to promote scientific excellence in the life sciences throughout Europe. In 2004, we decided to draw together the numerous and diverse activities that constitute our advanced training programme into one organisational unit, the EMBL International Centre for Advanced Training (EICAT). EICAT has now moved forward with the recruitment of a Coordinating Manager, Matthias Haury.

Together with EICAT Coordinator Anne Ephrussi and EMBL Associate Director Matthias Hentze, who also jointly oversee the EMBL International PhD Programme, Matthias Haury has already begun to bring the various activities together, further enhancing an already strong programme. The team is not only concentrating on graduate education, but also working to promote the new Postdoctoral Programme and to introduce a Vocational Training Programme for scientists and administrative staff at all EMBL sites, in collaboration with the EMBL Personnel Section. The excellent EMBL Visitors and Scholars Programme, as well as the Collaborative Training Programme, are also part of EICAT activities.

Although it has been active only since 2003, the European Learning Laboratory for the Life Sciences (ELLS) has already established itself as a major player in European efforts towards the training of science teachers. Within the framework of EICAT, ELLS fulfils an important function, fostering EMBL outreach towards the non-scientific community and bringing the forefront of scientific research closer to schools.

One of EICAT's most important commitments regards EMBL Conferences, Courses and Workshops. With ever-increasing numbers of applicants, these advanced training initiatives have challenged the EMBL conference capacities to their limits; it has frequently been necessary to decline over 80% of applicants who apply to participate in EMBL-organised workshops. New and improved facilities are urgently required to maintain EMBL's mission to provide the scientific community with a world-class advanced training programme in the biological sciences.

Facing this challenge, the EMBL Council very recently approved the construction of a new state-of-the-art Advanced Training Centre (ATC) on the EMBL Heidelberg campus. The ATC will enable EICAT to enhance and expand EMBL Courses and Conferences activities and to further develop its collaboration with EMBO in this area. ■

Plan of the EMBL campus showing the round ATC

The EMBL International PhD Programme (EIPP)

Founded in 1983, the EMBL International PhD Programme has established itself as a reference point in international graduate education in the life sciences and is a centrepiece of EICAT activities. The outstanding reputation of this world-renowned programme annually attracts over 700 of the best European and international students. The student body currently includes 180 students from 30 countries, all of whom were admitted after a competitive selection and interviewing process. After the first two months, in which thety attend the "Core Course in Molecular Biology", new students begin their PhD projects in laboratories at the main EMBL campus in Heidelberg or one of the Outstations. Students at EMBL are closely supervised throughout their PhD by a thesis advisory committee, composed of three EMBL group leaders and one university professor, to guarantee their scientific progress and guide them towards producing excellent research.

In 1997 EMBL was awarded the right to grant its own PhD degree, however, the EIPP encourages students to register and obtain a joint degree with one of 25 EIPP Partner Universities in 18 countries. Students may also register at a national university and defend their theses there. The list of Partner Universities is increasing; EIPP aims to have at least one or two excellent Partner Universities in each of the EMBL Member States.

More than half of EMBL PhD students receive fellowships from the EIPP. Others are supported by external grants managed by their group leaders. Some students participate in and are funded by the E-STAR project, generously financed by a Marie Curie Early Stage Training grant from the EU. Also, two students from an Eastern European member of the Council of Europe are selected each year for fellowships sponsored by the Fondation Louis-Jeantet de Médecine in Geneva.

The EIPP is enhanced by activities beyond the core course and lab work. For several years, EMBL's PhD students have organised an annual EMBL International PhD symposium. In 2004 students secured an independent, three-

year grant from the EU to sponsor these symposia. The event in 2005 was entitled "Biology at Work – A Journey Through Applied Life Sciences" and covered five scientific areas: neurobiology, plant sciences, biomedicine, environmental biology and evolution. Participants from almost all European countries discussed a wide spectrum of topics ranging from "Golden Rice" (designed to save lives in the developing world) to "20 Years of Cochlear Implants" (helping deaf children to acquire hearing and language). A writing competition for the communication of scientific results to the general public, organised in collaboration with the ELLS and EMBO, saw predoc Markus Elsner take the prize for his essay "What Burns Night can teach you about biology".

Additional opportunities for PhD students include participating in the ELLS activities. And, like EMBL postdocs, the students also organise their own retreat (see box). All of these activities are an important part of the extracurricular events for the PhD students, bridging gaps between students at different stages of their studies and those working at different EMBL sites. The EMBL PhD experience is thus an integral programme that not only prepares the students for a scientific career, but also encourages the development of interdisciplinary and personal skills, which are important for their successful future. ■

Exploring alternative careers at a medieval castle

What happens if you bring together 80 EMBL predocs in a medieval knight's castle in Wernfels, Northern Bavaria? Apart from spending some sunny leisure time canoeing and sightseeing in Nürnberg, the first-ever EMBL predoc retreat, organised September 3 – 5 2005, offered an exceptional opportunity to discuss the PhD programme, as well as science and life beyond EMBL, in a relaxed environment far from the lab and group leaders.

The event was fully funded from the solidarity fund, which was established through voluntary donations from predocs who are supported by fellowships from the E-STAR (Early-Stage Training in Advanced Life Science Research Across Europe) programme.

Part of the retreat was spent listening to academic talks by predocs, but probably the best remembered and valued talks were those on alternative careers presented by invited speakers:

Uli Weihe, formerly a predoc in Steve Cohen's lab at EMBL Heidelberg and now a consultant at McKinsey and Co. in Frankfurt, described the mindset of top management consulting as very similar to that of science: the daily business is to tackle problems through creative solutions.

Anna Eichhorn, founder of Humatrix, a market-leading company in DNA diagnostics, reported on the process of becoming self-employed, a step she took while still studying for her PhD.

Tine Walma, formerly a structural biologist and now assistant editor of *FEBS* Letters, talked about the world of editing and publishing, discussing science communication as a major job opportunity and presenting critical steps in publishing a paper from the perspective of a publishing house.

Klaus Müller, from Hoffmann-La Roche, gave an outline of the challenges and opportunities for a young scientist applying to enter industry, pointing out that an additional two years of postdoc experience in a complementary field provides an ideal basis for an industrial career.

It is well known that not all PhD programme graduates will wish to pursue an academic scientific career, and as alternative options usually do not receive very much exposure at EMBL, these talks were greatly appreciated, as they opened students' minds to a new horizon of possibilities that are accessible once you have an EMBL PhD in your pocket.

Malgorzata Duszczyk

The EMBL Postdoctoral Programme

The EMBL Postdoctoral Programme is a formal name for a set of activities that have been initiated at EMBL to make the time spent as a postdoc more attractive, interesting and constructive for the future. There are about as many postdocs at EMBL as there are predocs. Postdocs are a heterogeneous group, consisting of individuals who spend very different amounts of time at EMBL, with funding from many sources, and with a great variety of plans for the future. It is a group EMBL needs to pay attention to for many reasons, including the need to attract talented scientists to research careers in Europe and keep them here.

The postdoctoral programme has two special characteristics: the very active participation of interested postdocs themselves, reflecting the maturity and independence of the group, and the voluntary nature of all the activities. Only one senior EMBL faculty member is directly involved.

The first EMBL internal workshop dedicated specifically to the needs of postdocs was held in the autumn 2005. For this intense workshop, "Preparing for the academic job market" each participating postdoc was "on" for one three-hour session in addition to preparing written material. Feedback was given by EMBL faculty members Pernille Rørth and Elena Conti, as well as each of the participants. The workshop was much appreciated and two similar workshops are planned for 2006.

Additional aims for the postdoctoral programme – together with the Postdoc Association – include the establishment of special postdoc mailing lists, highlighting postdoc work through a dedicated series of short talks at EMBL's annual Lab Day, initiating an Alternative Careers Day, updating the websites (internal and external) and interacting with the Personnel Section to clarify issues concerning postdocs.

The Postdoc Retreat

Many EMBL postdocs, myself included, found it a little difficult to meet and interact with their peers outside of their immediate environment when they first arrived in their new labs. In comparison with the predocs – who are thrown together from the moment they arrive at EMBL, and thus have an instant support network and contacts in other labs – postdocs may find themselves a little isolated, both scientifically and socially. In addition, until very recently there has been no formalised training programme for postdocs, who consider themselves independent researchers but would still like a helping hand in furthering their careers. These two considerations were major driving forces behind the creation of the Postdoctoral Association.

The starting point was the first ever EMBL Postdoc Retreat, held at Mont Saint Odile Monastery in Alsace last April. Over one hundred postdocs from all EMBL units (approximately half the total postdoc population) gathered in these beautiful surroundings for two days of seminars, informal scientific discussions and social events. The retreat kicked off with a highly informative and entertaining talk from Martin Raff, Emeritus Professor of Biology at University College London, who gave a personal retrospective of his life in science, peppered with advice on how to further our own careers. On the second day, a career development session featured personal perspectives from Oliver Gruss, an ex-EMBL postdoc who recently set up his own group, Anna Migliazza, a postdoc-turned-industrial executive and journalist Paul Smaglik from Nature Jobs. These speakers provided much useful food-for-thought for those of us often left perplexed when considering our own future direction.

Overall, the retreat was a great success and has acted as a spring-board for further developments within the EMBL postdoc community. One session at the retreat provided an initial forum for discussion of

The postdoctoral programme would like to initiate closer interactions between EMBL postdocs and EMBL Alumni in various (European) countries. The rationale is that many postdocs will leave EMBL to become group leaders elsewhere in Europe and may need help learning about opportunities and grant systems in these countries. Other EMBL postdocs will decide to take another path, and input from those who have already done so would be very useful. EMBL Alumni could offer valuable advice in these areas.

Many postdocs need to secure individual fellowships or grant funding before joining EMBL, as most labs do not have dedicated postdoc positions. As a unique initiative, the Spanish Ministry of Science Fellowship Program (for international organisations) has recently dedicated a few fellowships to postdocs wishing to work at EMBL. The first selection round has just been completed and we hope that this very positive initiative will have a brilliant future. ■

postdoctoral issues and concerns, and led to the establishment of a voluntary Postdoctoral Committee. The committee now represents postdoctoral interests and is involved in setting up a number of initiatives designed to provide training and career development, which have until recently been somewhat lacking. To this end, the Postdoc Association, in conjunction with EMBLEM, will host an "Alternative Careers Day" in June 2006, and Pernille Rørth has set up an intensive course for senior postdocs who are interested in applying for group leader positions. On the social side, we now hold bi-monthly welcome receptions for new postdocs. Further events and courses will also follow, not least the second Postdoc Retreat to be held in october this year. Although the Postdoctoral Programme is still in its infancy, it is turning the postdoc population into a true community, and we are already reaping the benefits.

Katherine Brown

The EMBL Visitors Programme

Each year more than 1000 visitors from very different backgrounds come to the various EMBL sites to work for periods ranging from a day to several months or even more than a year. Whoever they are – diploma students or facility users, sabbatical visitors or long-term collaborators – the EMBL Visitors Programme facilitates their stay, striving to make their visit to EMBL a profitable experience. Visitors play a central part in the daily life of EMBL, as they contribute both scientifically and culturally to the very international and interactive atmosphere of the Laboratory. Many visit the EMBL Outstations in Hamburg and Grenoble to collaborate with leading scientists in structural biology, while others spend time at the EMBL-EBI in Hinxton where they encounter unique expertise in bioinformatics and computational biology. Those curious to learn about advanced mouse biology benefit from the scientific excellence in mouse genetics and functional genomics at the Outstation in Monterotondo, ideally situated next to the European Mouse Mutant Archive (EMMA). EMBL visitors become EMBL's ambassadors, taking with them a bit of the unique spirit of the Laboratory when they return to their home institutions. ■

Please refer to the Director General's Report on page xiii for further information.

The European Learning Laboratory for the Life Sciences (ELLS)

"Real" science in European classrooms

The European Learning Laboratory for the Life Sciences (ELLS) is EMBL's facility dedicated to improving science education in schools throughout Europe. Created in 2003 within the framework of an EU-funded project co-ordinated by EMBL's sister organisation EMBO, ELLS brings high-school teachers and scientists together in LearningLABs: three-day courses built around hands-on practicals, bioinformatics activities, educational games, as well as science and society forums dealing with controversial issues. The activities are developed by ELLS staff, Alexandra Manaia and Julia Willingale-Theune, together with EMBL scientists. ELLS follows the philosophy that excellent science teaching in high-school classrooms promotes the understanding of science in the general population and helps attract talented students to careers in science.

Over the last three years, 26 LearningLABs have been held across Europe: in France, Germany, Italy, Portugal and the Ukraine, reaching more than 350 teachers from 15 different nationalities and involving around 150 EMBL scientists.

2005/06 was a busy year, highlighted by the development of innovative educational projects in collaboration with the EMBL PhD International Programme.

A close collaborative project between the EMBL International PhD Programme and ELLS was initiated in February 2005. Since then, 18 students have worked with ELLS to design and produce innovative materials which have been tested, refined and disseminated in the context of ELLS courses and other outreach activities.

The Ukrainian connection

The export of LearningLABs to new European locations was a major activity, presenting a new model for sustainability of nationally organised and funded initiatives. Tanya Klymenko, a postdoctoral fellow from Jürg Müller's group, played a crucial role in establishing this model in the Ukraine. Through Tanya's connections with the Ukrainian education and scientific communities, the first LearningLAB, "Exploring the molecules of life" was organised in collaboration with the Bogdan Khmelnitsky Cherkasy National University in May 2005. The course had such a positive impact that the Ukrainian education authorities supported a second wave of LearningLABs that were held throughout the Ukraine in four locations: Cherkasy 28-29 October 2005, Kharkiv 27-28 January 2006, Lutsk 2-3 February 2006 and Kherson 7-8 April 2006. ■

Three projects among many illustrate the innovation and input of advanced skills from our PhD students:

A First Look at the Code of Life – an introduction to bioinformatics by Cleopatra Kozlowski

"I came up with the idea of designing a bioinformatics course aimed at students with little background in biology during the EMBO Teacher's Workshop in May 2005. As accessibility was my main concern, the four lessons were designed to use only paper and pencil and the concepts were explained using metaphors and games. In the "mutation game", nucleotide cards are used to create "mutations" in RNA sequences. Lesson three explains how the accumulation of DNA mutations allows bioinformaticians to deduce phylogenetic relationships between organisms. The last lesson explores the concept of mobile DNA, comparing it to scrambling up words in a cooking recipe, to see if new recipes appear. The course has been very well received by teachers, who have provided invaluable feedback for improvement."

A hands-on practical on protein expression and purification by Philipp Gebhardt

"Designing a science classroom activity is quite a challenge for a bench scientist. One has to focus on the basic concepts underlying biological processes and then try to convey them in easily understandable terms to a non-expert audience. Simplifying the principle of protein affinity purification and putting it into a hands-on activity format was a very interesting experience for me. I also developed an accompanying teaching system, the Modular Extensible Magnetic Array (MEMA), comprising different magnetic elements to be used on a traditional classroom board. The teachers' feedback helped me to further refine the system. It is inspiring to participate in these activities: there are mutual benefits for scientists and teachers."

"Explorer les molécules: de la structure à la fonction" – designing and instructing a LearningLAB at the EMBL Grenoble outstation by Jeanne Morinière and Elena Seiradake

"Getting official accreditation from the Grenoble Rectorat was the first step towards enabling French teachers to attend the course. We then concentrated on putting the programme together, involving EMBL Grenoble scientists in teaching and preparing the handbook. The final programme covered a classic protein crystallisation work-flow at the EMBL: purifying and crystallising the protein, freezing the crystals, working on the beamlines at the European Synchrotron Radiation Facility (ESRF), processing diffraction images, visualising 3D-model, and demonstrating classroom kits designed by Manfred Weiss from EMBL Hamburg. The 12 participants were highly qualified biology and physics teacher-trainers, possessing a strong scientific background. They were so eager to learn about our work; this made us feel special and really increased our regard for our own research!"

Left: PhD student Philipp Gebhardt explains basic biological processes to teachers

Right: French teachers working on protein purification and crystallisation at EMBL Grenoble

Courses, Conferences and Workshops

The EMBL Heidelberg Course and Conferences Office (from left to right): Bettina Schäfer, Antje Seeck, Sylke Helbing, Doros Panayi (Head of Services),and Emma Fassmann.

For the past 20 years, the EMBL Courses and Conferences Office (CCO) has successfully managed and organised scientific conferences, courses and workshops, many sponsored or co-sponsored by EMBO. Over the years, these scientific events at EMBL have become more and more frequent and have attracted an increasing number of attendees, reaching a frequency that now stretches EMBL's conference capacities to the limits. In the last year, five large conferences, six courses, and seven workshops have taken place at EMBL Heidelberg, and in the last five years, more than 10 000 scientists from over 60 countries have participated in over 100 of those events. All Outstations participate very actively in this programme, and the EMBL-EBI in 2005 alone organised more than 60 courses in 17 countries, teaching scientists how to best exploit the information available in the various EMBL-EBI databases.

The CCO provides a wide variety of services. In the planning phase of an event, the CCO staff manages everything related to with event promotion, websites, communication with participants, registration, housing and budgeting. At a later stage they ensure the logistics and technical support surrounding the event, manage public relations and acts as the interface to catering and transport services. They also work alongside with their colleagues from the EMBL Photolab, who provide audio-visual expertise. Further in-house services that contribute to the smooth running of conferences include Catering, Building Maintenance, Office of Information and Public Affairs, Scientific Instrument Maintenance, IT, and Finance.

Given the importance of scientific meetings for EMBL in fostering the exchange of scientific ideas with scientists at the Laboratory and within Europe, it became clear that the current EMBL conference facilities are no longer appropriate to provide the most attractive conference experience. To allow EMBL to build on its leading role in the international conference scene, new facilities are needed. The recently approved Advanced Training Centre will soon provide an infrastructure ideally suited for international workshops, courses, and conferences, and will also change the way the CCO operates. With new informatics support, the complete conference management system will be streamlined, allowing conference officers to actively develop new avenues in areas such as sponsoring, advertising and customer support. With the ATC, additional activities involving a greater number of attendees will be possible, and will provide a new challenge for the CCO in the future. ■

The Advanced Training Centre

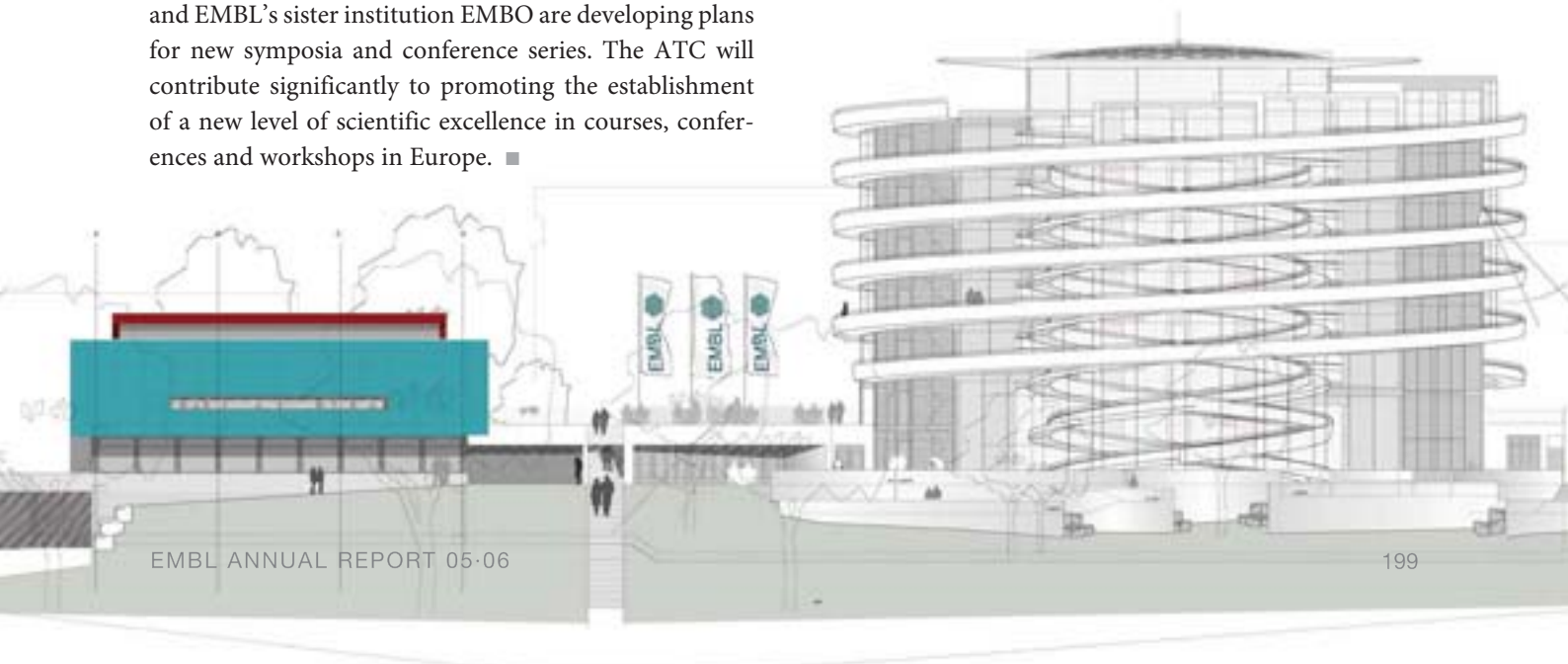
In an extraordinary session on 22 March 2006 EMBL Council approved the construction of a new Advanced Training Centre (ATC) on the EMBL campus in Heidelberg. This project has been made possible by a generous donation from the Klaus Tschira Stiftung, a local private foundation, and a large contribution from EMBL Heidelberg's host country, Germany.

The new building, constructed in the form of a DNA double helix, will provide the infrastructure to integrate all of EMBL's training activities and will house all the activities of EICAT. It will also be home to the Office of Information and Public Affairs, the EMBL Science and Society Programme, the Szilárd Library, and the Photolab.

The centrepiece of the building will be a 450-seat auditorium, as well as spectacular exhibition space for 300 posters, teaching labs for 60 participants and a 30-seat computer teaching laboratory. The ATC will also house three medium-sized seminar rooms and several meeting rooms for use by courses and conferences visitors and EMBL staff. The EMBL canteen and kitchen will move into an adjacent new building, allowing high-quality catering to both meeting participants and EMBL staff. The large ATC foyer, together with the rooftop terrace and the lounge, will offer ample space for social events, and informal scientific exchange in a relaxed atmosphere.

Not only unique in its architecture, but also innovative in its design as a conference facility, the ATC will enable EICAT to significantly enhance its already internationally reputed conference and workshop programme, thus ensuring that EMBL remains one of the most attractive scientific conference venues in Europe. Together, EICAT and EMBL's sister institution EMBO are developing plans for new symposia and conference series. The ATC will contribute significantly to promoting the establishment of a new level of scientific excellence in courses, conferences and workshops in Europe. ■

With an overall floor space of approx. 16,000 sq.m, the ATC will offer ample room for conferences and poster exhibitions



The customer is always right

The EBI is an unusual part of EMBL because it focuses predominantly on the development and maintenance of bioinformatics services: threequarters of its staff (around 225 people) are dedicated to this task and the EBI website attracts over 2.2 million users each month. Consequently, even though all our services are freely available to researchers without restriction, some of the metrics valued by commercial organisations are important to us: like any business, we depend on our user community for our survival, and for many of us, production deadlines, usage statistics and market research are therefore part of our daily lives.

How does any service-based organisation ensure that it is giving its users what they want? A good place to start is to ask them what they think of our products – and that is exactly what we did for a six-week period in spring 2006. During this time, more than 900 of our users provided valuable feedback on the EBI website through a web-based survey hosted by surveymonkey.com. Our objective was not to get detailed feedback on individual data resources (this can be addressed using surveys specific to a particular database) but to get an overall feel for what our users think of our website, with the aim of using their feedback to improve it.

A typical respondent

The majority of people who completed the survey are academic researchers, working in Europe, who consider bioinformatics to be important to their work. This doesn't quite match with the demographics of our user community, which includes a higher proportion of users from the United States and from the commercial sector. Over 70% of our respondents use the EBI website regularly, with over one-tenth of them using it more than 50 times a week. Around half of the users who completed the survey were experimental biologists and the other half were computational biologists.

How are we doing?

Gratifyingly, the most popular answer to "What limits your use of the EBI?" for both groups was "nothing". When asked to rate various aspects of the website on a sliding scale from "terrible" to "excellent", we were relieved to find that the vast majority of survey respondentss consider the EBI website to be "good" or "OK" in terms of its graphics, accuracy of the data, timeliness of information, reliability, speed and ease of navigation. Nevertheless, a quarter of respondents agreed that they find some things hard to find, and 18% thought the user interfaces unfriendly.

What's important to our users?

Over 80% of our survey respondents classed our databases as vital or important. Our most well established resources proved to be the most popular: nucleotide sequences and eukaryotic genomes were the most frequently cited essential databases (with eukaryotic genomes being a particularly high priority to the experimental biologists); protein sequence, protein families and protein–protein interactions were close behind (see figure). Analysis tools and the training section of the website were also viewed as important by a significant proportion of respondents. Only a small percentage of respondents

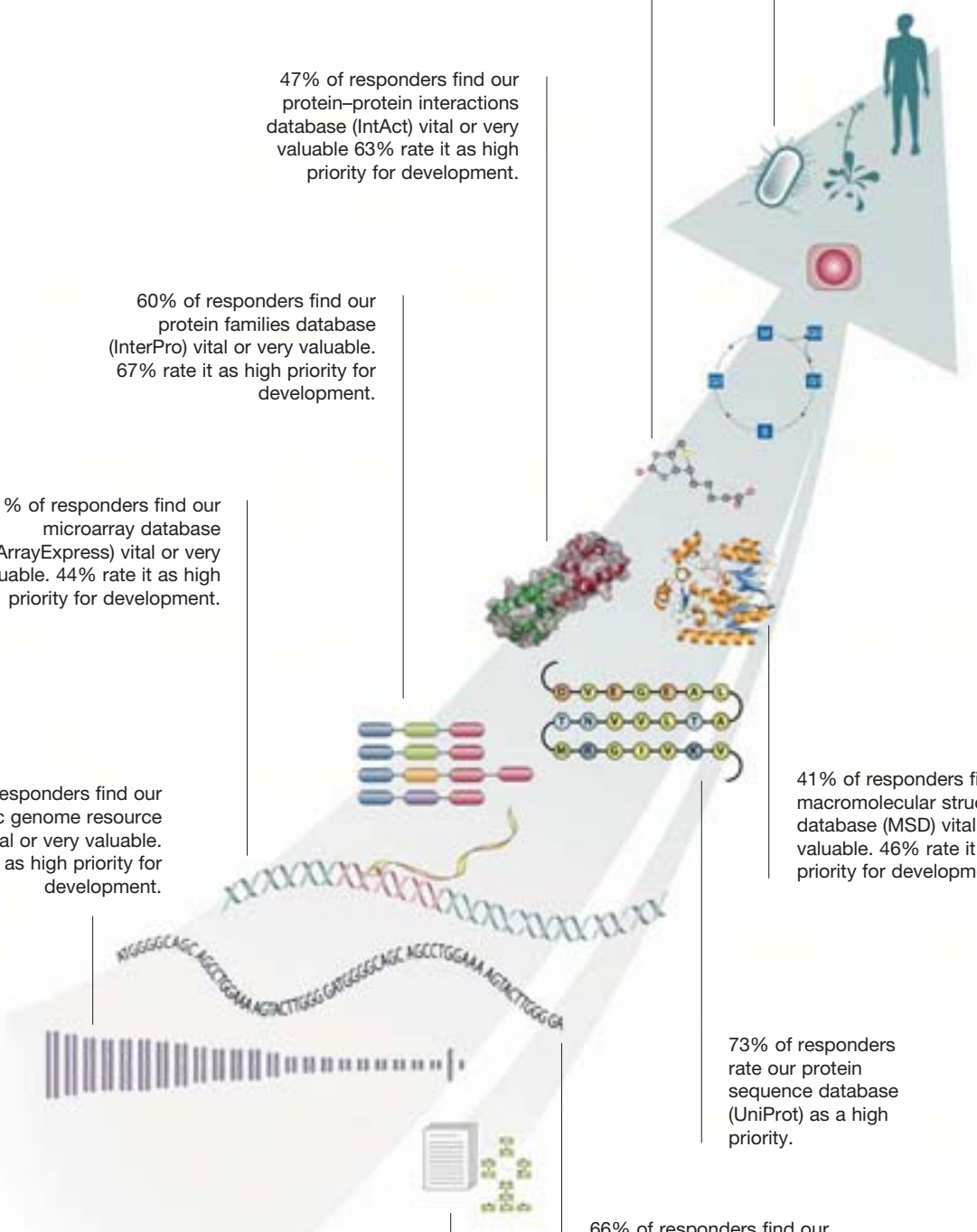
18% of responders find our chemical entities database (ChEBI) vital or very valuable. 27% rate it as high priority for development.

38% of responders rate the development of our biological models database (BioModels) as a high priority.

47% of responders find our protein–protein interactions database (IntAct) vital or very valuable 63% rate it as high priority for development.

60% of responders find our protein families database (InterPro) vital or very valuable. 67% rate it as high priority for development.

31% of responders find our microarray database (ArrayExpress) vital or very valuable. 44% rate it as high priority for development.



51% of responders find our eukaryotic genome resource (Ensembl) vital or very valuable. 63% rate it as high priority for development.

41% of responders find our macromolecular structure database (MSD) vital or very valuable. 46% rate it as a high priority for development.

73% of responders rate our protein sequence database (UniProt) as a high priority.

66% of responders find our nucleotide sequence database (EMBL-Bank) vital or very valuable. 68% rate it as high priority for development.

38% of responders find our literature and text-mining tools (CitExplore and EBIMed) vital or very valuable.

56% of responders rate the development of the Gene Ontology as a high priority.

had tried our data integration and browsing tools, although those who were familiar with them rated them highly; BioMart proved to be the bioinformaticians' preferred analysis tool whereas experimental biologists were more likely to use the Sequence Retrieval System, SRS.

When we asked what priority we should give to developing our services, nucleotide sequences, eukaryotic genomes, protein sequences, families and interactions, and Gene Ontology received the most votes. This provides a conundrum for us, as it is far more challenging to obtain resources for the development and maintenance of well-established data resources than it is to fund new ones. Indeed, EMBL-Bank, our nucleotide sequence database and the foundation upon which almost all our other data resources are built, receives no external funding whatsoever despite the fact that the environmental genome sequencing projects are producing new data at an alarming rate and the database must constantly evolve to meet the changing needs of its users.

When it comes to developing our tools for searching and analysing data, similarity searches, sequence analysis, protein structure analysis and protein structure searches came highest in the respondents' list of priorities; literature, text mining and tutorials came next. There was uncertainty amongst the respondents as to whether we should develop SOAP-based web services (a means of gaining programmatic access to data resources so that users can query them as though they were installed on their local computers).

What else do our users want?

We asked our users to tell us what new services they would like us to deliver. The responses could be divided into three categories. Firstly, training materials were a popular request: respondents asked for tutorials, handbooks and worked examples. Secondly, better ways of accessing the data were requested: our respondents wanted more web services, integrated search and analysis software and better navigation tools. Finally, there are new types of data: requests included data on post-translational modifications, pathways, alternative splice forms and regulatory regions. This illustrates the challenge of providing transparent access to the huge number of data resources that the EBI hosts: we do, in fact, host databases of post-translational modifications (Resid, which is used to annotate UniProt entries), pathways (Reactome) and alternative splice forms (ASD), but not all of our users have been able to find what they are looking for.

Our final question was "Is there anything else you'd like to tell us?" The respondents had a huge amount of praise for the EBI's user support. The most frequent negative comments related to the observation that there is too much information on our website; users complained that they have to click too many times to get the information they need. "The EBI offers too much for a new user to take in" and "too big a choice of search algorithms, databases and post-processing options" summarised these sentiments succinctly. "Every tool should be accessible by people normally at the bench" encapsulated where we would like to be. Surprisingly, a higher proportion of informaticians (31%) than biologists (18%) found things hard to find on the website, perhaps because they look more deeply into what is available.

Conclusions and next steps

An important purpose of the EBI user survey was to generate results that we can act upon to improve our services to users. The key conclusions that we have made from this survey, and other ongoing input from our users, are:

- the database search tools are difficult to use for some users
- the website is difficult to navigate for some users
- some pages are geeky, making it difficult for users new to bioinformatics to understand what is available.

The current EBI website has served our users well for the past five years but it was designed when we offered a smaller range of services - it now needs to be revamped in a way that simplifies data access for the general user without losing the capabilities for sophisticated data analysis for which our expert users rely upon us. Our External Service team has already begun the design of a more intuitive search interface to the main databases, and is making this much more high-profile on the main page of the EBI website. The general design of menus and pages is being re-visited to address general navigation issues, removing confusing icons and terms on high-level pages. Many of the project pages will be updated, using simpler explanations to document what each resource is about and moving some of the geeky underlying technology issues away from the frontline.

A wide variety of specific issues was raised in the survey as free-text responses. These ranged from comments about website design to remarks on specific features of databases or analysis tools. More than 200 suggestions for new services or improvements were made, and over 100 respondents left their email addresses for follow-up discussions. We will be dissecting these and passing them on to the appropriate development groups, and responding to the individual users as far as is possible. ■

EMBL Alumni – Seeding Europe with top scientists

In the three decades of EMBL's existence, the body of alumni has grown to include almost 4000 former scientists, students and support staff spread across Europe and the world. They are important both as individuals and as a network of people who have worked in a unique, international and interdisciplinary environment. Upon leaving EMBL the majority move on to the Member States and continue a career in their national systems, thus strengthening European science.

Alumni matters are organised at EMBL in two ways: First through staff members who maintain statistics for all EMBL Alumni and support and interface with the Alumni Association; and second through the Alumni Association, a body of approximately 1050 registered alumni with an elected board.

2005-2006 saw many new faces at both levels of organisation.

EMBL's interaction with the Association

Associate Director Matthias Hentze is now EMBL's interface to the Alumni Association Board, taking over from Iain Mattaj. In-house support has now been taken over by the Director General's office. Mehrnoosh Rayner currently supports the Association administratively.

One important relationship between alumni and the Laboratory concerns the EMBL International PhD Programme (EIPP). Every year the EIPP has to turn down highly qualified applicants to the Programme simply for reasons of capacity. Thus the EIPP launched a "Shared Applicant Pool" in December 2005. Data on students who apply but are not accepted, and who register and give their consent, can be accessed by interested alumni who have become independent investigators and are seeking students. This service proved enormously popular and has so far been used by 115 EMBL Alumni.

Alumni Association

Board Elections

Alumni Board elections, which take place every three years, were held online in September 2005. Almost half the seats on the Board were renewed, with five members stepping down and eight new representatives elected by members of the Association. There are now 17 Board members in total.

EMBL Alumni Board

New Members	Continuing Members
Oddmund Bakke	Colin Dingwall
Freddy Frischknecht	Angus Lamond (Chair)
Bernard Hoflack	Daniel Louvard
Tony Hyman	Konrad Müller
Claudia Koch-Brandt	Annalisa Pastore
Richard Morris	Albert Stegmüller
Giovanni Paolella	Renata Stripecke
Niovi Santama	Juan Valcárcel
	Fotis Kafatos

In May 2006, the newly elected board of the EMBL Alumni Association gathered in Heidelberg for their annual meeting. Back row from the left: EMBL Associate Director Matthias Hentze, Bernard Hoflack, Oddmund Bakke, Angus Lamond, Freddy Frischknecht, Richard Morris. Front row from the left: Mehrnoosh Rayner from the EMBL Director General's office, Giovanni Paolella, Claudia Koch-Brandt, Albert Stegmüller, Daniel Louvard, Renata Stripecke

Local Chapters

The Alumni Association local chapters were launched in 2004. Many alumni find that linking up with former colleagues in their area can be useful, and a good chapter can act as a helpful local support structure. They have become increasingly active, with four local chapter meetings held in Dresden, Vienna, Barcelona and London in 2005-2006.

Programme details for all local chapter meetings can be found on the Alumni Association web pages http://www.embl.org/aboutus/alumni/chapters.html, as can information on all current Alumni Association local chapters: Austria, France, Germany, Greece, Italy, Scandinavia, Spain and Portugal, Switzerland, the UK and the USA.

Statistics

In terms of nationality, the largest number of alumni with a scientific background still come from the four biggest EMBL Member States, Germany, the UK, France and Italy. These four countries also have the highest number of EMBL Alumni residents. Perhaps not surprisingly, the smallest number originates from Croatia and Iceland (the most recent member states). An impressive average of 88% of EMBL Alumni for whom we have an address return to an EMBL Member States. ■

Science in School

Science is moving more rapidly than ever; one groundbreaking discovery chases the next at an incredible speed. School teachers have trouble keeping up with the pace, and yet many pupils call science classes "boring". In response to this crisis, EMBL and its partners have launched *Science in School*, a free journal to promote inspiring science teaching across Europe and across scientific disciplines.

Interpreting "science" in the broadest sense, *Science in School* includes not only biology, physics and chemistry, but also maths, earth sciences, engineering and medicine, highlighting the best in teaching and cutting-edge research, and focusing on interdisciplinary work.

Science in School features news about the latest scientific discoveries, teaching materials, interviews with inspiring teachers and scientists, reviews of books, films and websites, events for teachers and many other useful resources for science teachers. Contributors to the first issue included renowned neurologist and author Oliver Sacks, and scientists and teachers from nine countries.

Science teachers, scientists, education ministries, parents and children across Europe have received the first issue of *Science in School* with enthusiasm. The journal looks well set to achieve its aim of bridging the gap between the worlds of research and school, by encouraging communication among all stakeholders in science education. One powerful tool to achieve this will be the journal's online discussion forum, enabling direct dialogue across national and subject boundaries.

Supported by the European Commission's Science and Society Programme, *Science in School* is part of a larger science education project, NUCLEUS. Appearing quarterly online and in print, it is published by EIROforum, a partnership between Europe's seven intergovernmental research organisations. Whereas the print copy is in English, online articles are provided in many European languages.

EIROforum's other major education activity is the "Science on Stage" teaching festival (www.scienceonstage.net) at which teachers selected in national competitions meet to exchange their most innovative teaching ideas. Some of the best projects from last year's "Science on Stage" festival will be featured in *Science in School.*

Science in School is freely available online at www.scienceinschool.org. To receive an alert when a new issue is published, please send an email alert with the subject "Subscribe to *Science in School*" to scienceinschool@embl.de. Include your postal address to receive a free print copy.

Submissions are sought from scientists who enjoy communicating science to a broad audience. Accessibly written articles on cutting-edge developments or reviews of current topics in science are particularly welcome. ■

A Year in the Life of EMBL

April

Postdoc retreat in Alsace

For the first-ever postdoc retreat 107 young scientists gathered in a monastery in France. The purpose of the retreat was to encourage scientific exchange, address career-related issues and establish a committee to give this community a voice.

May

EBI expansion goes ahead

The EBI has received a big boost from The Wellcome Trust, the Medical Research Council and the Biotechnology and Biological Sciences Research Council, which have given funds to extend the EBI building in Hinxton. The new development will provide 1,500 square metres of space which, together with the existing 3,000-square-metre building, will house over 400 staff.

6th EMBL International PhD Student Symposium

For the first time the PhD Student Symposium was organised outside Heidelberg as a result of the close collaboration between EMBL Monterotondo students and colleagues from the University of Rome “Tor Vergata”. Divided into four sessions, the symposium highlighted the role of animal models in research, and challenged misconceptions about the impact of basic sciences on cutting-edge technology.

2005

June

Lab Day 2005

Scientists from all the EMBL sites gathered in Heidelberg on June 7 to celebrate Lab Day. The schedule included talks presenting interesting science from many of EMBL's units. Students received their PhD diplomas as part of the International PhD Programme's graduation ceremonies.

Grants Office information day

With a practically oriented FP6 information day, the Grants Office at EMBL Heidelberg aimed to aid individuals looking for funding and helped to clarify the grants process. EMBL scientists, staff and administrators working on various EU-funded projects, as well as external visitors learned how to deal with legal documents cost breakdowns and justifications from legal advisors from the EU Office of the Federal Ministry for Education and Research.

Juli

Art in Science – Science in Art

EMBL celebrated the various aspects of creativity generated by scientists and artists alike and hosted a one-day festival called "Art in Science in Art". Speakers included artists like Suzanne Anker from the School of Visual Arts, New York, and Christa Sommerer from the University of Art and Design, Linz, as well as philosopher Arthur I. Miller from University College London. A colourful exhibition of artworks, created by members of the EMBL community, demonstrated that scientists can be artists too and that research can sometimes produce unexpected and spontaneous artwork.

Molecular Medicine Partnership Unit

EMBL and the Medical Faculty of the University of Heidelberg have initiated a second phase of the Molecular Medicine Partnership Unit (MMPU), a research unit jointly established by the two institutions. Since its creation in 2002 the MMPU has created valuable insights into the mechanisms underlying some of the most common human genetic diseases. The new agreement will put the unit on a firmer basis for the long term, strengthening ties between the institutions and intensifying activities in post-graduate training.

New beamline for Hamburg

EMBL Hamburg completed X12, Its newest energy tuneable beamline for macromolecular crystallography. X12 was designed and built by the EMBL Hamburg instrumentation group, with an emphasis on simplicity of operation, fast energy tuneability and, above all, user-friendliness. It will help reduce the experimental load on BW7A, the most overbooked of the EMBL Hamburg crystallography beamlines.

Collaboration started with Japan

Renowned Japanese scientist and current president of the National Institutes of Natural Sciences (NINS) Professor Yoshimo Shimura paid a visit to the Heidelberg Laboratory. He came to sign an agreement between EMBL and one of the five member institutions of NINS, the NIBB (National Institute of Basic Biology), to promote collaborations via joint conferences and to facilitate the exchange of scientists by introducing Shimura Awards, travel grants for EMBL scientists to visit scientific institutions in Japan.

August

EMBL Hamburg at IUCr 2005

EMBL Hamburg headed to Florence at the end of August to be present at this year's congress of the International Union of Crystallography (IUCr). At this important meeting for the structural biology community EMBL Hamburg highlighted its new High-Throughput Crystallisation Facility and the plans for a Life Science Centre at the future PETRA III synchrotron.

September

PhD students' retreat

PhD students from all five EMBL locations embarked on their first-ever retreat. The event gave the predocs a chance to meet, discuss science and figure out ways to work together. Generously funded by the E-Star students – recipients of Marie Curie Early Stage Training in Advanced Life Science Research fellowships – the event hosted a series of talks on alternative careers to academia. The retreat was also an opportunity for many PhD students to present and get feedback on their work.

EMBL's biggest-ever conference

Over 320 scientists signed up for the EMBO Conference on "Protein Synthesis and Translational Control" held in mid-September at EMBL's main Laboratory – making it the largest conference ever held on the Heidelberg campus. This conference raised very high interest because it addressed a traditional topic in biology that plays a crucial role in development, the central nervous system, memory and disease.

2005

October

Science and Security

With the theme “Science & Security”, the 2005 EMBO/EMBL joint Science & Society conference addressed a hot topic. Speakers from various institutions, including the US Chemical and Biological Arms Control Institute and Interpol, talked about the use and the abuse of biological knowledge, information technology in the knowledge society, and freedom of research.

EMBL Grenoble opens its doors

EMBL Grenoble took part in the city’s “Fête de la Science” with a joint EMBL/ESRF/ILL stand, which was visited by about 16,500 people during the three-day weekend, devoted to promoting communication between scientists and the general public, including a day specifically aimed at school pupils.

November

Inauguration of the High-Throughput Facility at EMBL Hamburg

EMBL Hamburg took a major leap forward with the services it provides to researchers in the EMBL Member States with the inauguration of Europe’s largest high-throughput crystallisation facility. The resource was officially opened on the DESY/EMBL Hamburg site in the presence of the Director General, campus partner DESY, the ESRF and the facility’s main sponsors, the BMBF and the EU.

EMBL Hamburg hosts Italian journalists

Sixteen Italian science journalists made their way from Rome, Milan and Bologna to meet the community at EMBL Hamburg. The visit was led by Istvan Palugyai, the vice-president of the European Union of Science Journalists (EUSJA) and stemmed from his interest in the BIOXHIT project. As well as BIOXHIT, lectures covered major EMBL Hamburg projects including X-MTB and the PETRA III collaboration with DESY.

December

7th PhD Symposium “Biology at work”

In December 2005,150 participants from 29 countries made their way to Heidelberg in December 2005 to gather for “Biology at Work – A Journey Through Applied Life Sciences” the seventh international EMBL PhD Student Symposium. Scientific sessions were complemented with talks on a wide range of different topics, ranging from “Golden Rice” designed to save lives in the developing world, to “20 years of Cochlear Implants” helping deaf children to acquire hearing and language.

EMBL/EMBO/IBC-CNR Mini-Symposium on “Genes and Behaviour” in Monterotondo

In 2005, EMBL Monterotondo hosted a major Science and Society event for the first time, with a one-day symposiumon “Genes and Behaviour”. The programme covered the scientific and societal implications of the complex relationships between genes and behaviour, as well as new analytical methods for assessing disease risk factors. This was a privileged occasion to introduce Science and Society issues, not only to the staff at the EMBL Outstation, but also to a much wider audience at the IBC-CNR campus in Monterotondo.

January 2006

Carl-Ivar-Brändén Building opening

The new Carl-Ivar-Brändén Building (CIBB), adjacent to the EMBL Grenoble Outstation, opened its doors as a new centre of excellence for European structural biology. The building will house research groups from the Partnership for Structural Biology, including the ESRF, ILL and EMBL, as well as the Institute de Virologie Moléculaire et Structurale, associated with the Université Joseph Fourier. The facilities will include a complete pipeline for carrying out high-throughput structural investigations of proteins and other molecules.

EU Commissioner Janez Potočnik visits EMBL

“A model institution for science in Europe” was Commissioner Janez Potočnik’s conclusion when he visited EMBL Heidelberg. Two projects in which the Commissioner showed particular interest were the EMBL International Centre for Advanced Training (EICAT) and the bioinformatics services and databases of the EBI.

March

Advanced Training Centre (ATC)

At an extraordinary meeting EMBL Council approved the construction of an Advanced Training Centre (ATC) to start in October 2006. The ATC will be a unique European centre that combines cutting-edge facilities for practical laboratory and computer training courses with the infrastructure required to host medium-sized international conferences. The project has been made possible by a large contribution from the host country, Germany, and a donation from the Klaus Tschira Foundation.

EIROforum launches *Science in School*

A new European journal to promote inspiring science teaching, *Science in School*, was launched at EMBL by the EIROforum. *Science in School* is Europe's first international, multidisciplinary journal for science teaching. It is aimed at secondary school teachers, scientists and other stakeholders in European science education, and highlights the best in teaching and cutting-edge research.

EBI welcomes master's students

The EBI master's students' open day welcomed students from up and down the UK, across Europe and further afield. The event is designed to motivate participants studying for a bioinformatics master's degree. Half-hour lectures from research group members offered insights into the EBI's scientific work while the service teams demonstrated their resources and answered questions about how to enter a career in bioinformatics.

Cambridge Science Festival

For the second time the EBI and Wellcome Trust Sanger Institute combined forces at the Cambridge Science Festival. There was something for each of the 1700 visitors, including "build a bug" using balloons and wool, DNA modelling activities, using computers to decipher the genetic code and look up the functions of proteins, and a discussion board that tackled the issue of "Who benefits from biological research?"

2006

Selected literature

Aloy, P., Böttcher, B., Ceulemans, H., Leutwein, C., Mellwig, C., Fischer, S., Gavin, A.C., Bork, P., Superti-Furga, G., Serrano, L. & Russell, R.B. (2004). Structure-based assembly of protein complexes in yeast. Science, 303, 2026-2029

Arendt, D. (2005). Genes and homology in nervous system evolution: Comparing gene functions, expression patterns, and cell type molecular fingerprints. Theory in Biosciences, 124, 185-197

Bairoch, A. et al. (2005). The universal protein resource (UniProt). Nucleic Acids Res., 33, D154-D159

Beckmann, K., Grskovic, M., Gebauer, F. & Hentze, M.W. (2005). A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in Drosophila. Cell, 122, 529-540

Brennecke, J., Stark, A., Russell, R.B. & Cohen, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol., 3, e85

Cappelletti, G., Surrey, T. & Maci, R. (2005). The parkinsonism producing neurotoxin MPP+ affects microtubule dynamics by acting as a destabilising factor. FEBS Lett., 579, 4781-4786

Carninci, P. et al. (2005). The transcriptional landscape of the mammalian genome. Science, 309, 1559-1563

Carola, V., Frazzetto, G. & Gross, C. (2006). Identifying interactions between genes and early environment in the mouse. Genes Brain Behav., 5, 189-199

Carola, V., Frazzetto, G., Lesch, K.-P.*, and Gross, C. (2006). Loss-of-function mutation in 5-HT transporter modulates long-term effects of rearing environment in the mouse. In press.

Castagnetti, S., Hentze, M.W., Ephrussi, A. & Gebauer, F. (2000). Control of oskar mRNA translation by Bruno in a novel cell-free system from Drosophila ovaries. Development, 127, 1063-1068

Castoldi, M., Schmidt, S., Benes, V., Noerholm, M., Kulozik, A.E., Hentze, M.W. & Muckenthaler, M.U.(2006). A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA). RNA, 12 (5), 913-920

Caudron, M., Bunt, G., Bastiaens, P. & Karsenti, E. (2005). Spatial coordination of spindle assembly by chromosome-mediated signalling gradients. Science, 309, 1373-1376

Ciccarelli, F.D.*, Doerks, T.*, von Mering, C., Creevey, C.J., Snel, B. and Bork, P. (2006). Towards automatic reconstruction of a highly resolved tree of life. Science, 311,1283-1287

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.M. & Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. Proc. Natl. Acad. Sci., 103 (14), 5320-5325

Deidda G, Rossi N, Tocchini-Valentini GP. (2003). An archaeal endoribonuclease catalyzes cis- and trans- nonspliceosomal splicing in mouse cells. Nat Biotechnol. 21, 1499-504

Di Segni G, Borghese L, Sebastiani S, Tocchini-Valentini GP. (2004). A pre-tRNA carrying intron features typical of Archaea is spliced in yeast. RNA, 11, 70-76

Duncan, K., Grskovic, M., Strein, C., Beckmann, K., Niggeweg, R., Abaza, I., Gebauer, F., Wilm, M. & Hentze, M.W. (2006). Sex-lethal imparts a sex-specific function to UNR by recruiting it to the msl-2 mRNA 3'UTR: translational repression for dosage compensation. Genes Dev., 20, 368-79

Durinck, S. et al. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics, 21, 3439-3440

Freilich, S. et al. (2005). Relationship between the tissue-specificity of mouse gene expression and the evolutionary origin and function of the proteins. Genome Biol., 6, R56

Freilich, S. et al. (2005). The complement of enzymatic sets in different species. J. Mol. Biol., 349, 745-763

Gavin, A.C. et al. (2006). Proteome survey reveals modularity of the yeast cell machinery. Nature, 440, 631-636

Gross, C., Zhuang, X., Stark, K., Ramboz, S., Oosting, R., Kirby, L., Santarelli, L., Beck, S. & Hen, R. (2002). Serotonin1A receptor acts during development to establish normal anxiety-like behaviour in the adult. Nature, 28, 396-400

Gurniak, C.B., Perlas, E. & Witke, W. (2005). The actin depolymerizing factor n-cofilin is essential for neural tube morphogenesis and neural crest cell migration. Dev. Biol., 278, 231-241

Hart, D.J. & Tarendeau, F. (2006). Combinatorial library approaches for improving soluble protein expression in Escherichia coli. Acta Crystallogr D Biol Crystallogr, 62, 19-26

Herrmann, O., Baumann, B., de Lorenzi, R., Muhammad, S., Zhang, W., Kleesiek, J., Malfertheiner, M., Kohrmann, M., Potrovita, I., Maegele, I., Beyer, C., Burke, J.R., Hasan, M.T., Bujard, H., Wirth, T., Pasparakis, M. & Schwaninger, M. (2005). IKK mediates ischemia-induced neuronal death. Nat. Med., 11, 1322-1329

Hucka, M. et al. (2003). The Systems Biology Markup Language (SBML): A medium for representation and exchange of biochemical network nodels. Bioinformatics, 19, 524-531

Isalan, M., Lemerle, C. & Serrano, L. (2005). Engineering gene networks to emulate Drosophila embryonic pattern formation. PLoS Biol., 3, e64

Isalan, M., Santori, M.I., Gonzalez, C. & Serrano, L. (2005). Localized transfection on arrays of magnetic beads coated with PCR products. Nat Methods, 2, 113-118

Karsenti, E. & Vernos, I. (2001). The mitotic spindle: a self-made machine. Science, 294, 543-547

Klemenhagen, K.C., Gordon, J.A., David, D.J., Hen, R. & Gross, C.T. (2006). Increased fear response to contextual cues in mice lacking the 5-HT1A receptor. Neuropsychopharmacology, 31, 101-111

Klymenko, T. & Muller, J. (2004). The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. EMBO Rep., 5, 373-377

Klymenko, T., Papp, B., Fischle, W., Kocher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M. & Müller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes Dev., 20, 1110-1122

Kunin, V. et al. (2005b). The net of life: Reconstructing the microbial phylogenetic network. Genome Res., 15, 954-959

Le Novère, N. (2005) BioModels.net, tools and resources to support Computational Systems Biology. In "4th Workshop on Computation of Biochemical Pathways and Genetic Networks", Kummer, U. et al. (eds.), Logos, 69-74

Le Novère, N. et al. (2005). Minimum information requested in the annotation of biochemical models (MIRIAM). Nat. Biotechnol., 23, 1509-1515

Le Novère, N. et al. (2006). BioModels Database: A free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucleic Acids Res., 34, D689-D691

Lingel, A. & Sattler, M. (2005). Novel modes of protein-RNA recognition in the RNAi pathway. Curr. Opin. Struct. Biol., 15, 107-115

Lingel, A., Simon, B., Izaurralde, E. & Sattler, M. (2005). The structure of the flock house virus B2 protein, a viral suppressor of RNA interference, shows a novel mode of double-stranded RNA recognition. EMBO Rep., 6, 1149-1155

Lorentzen, E. & Conti, E. (2005). Structural basis of 3' end RNA recognition and exoribonucleolytic cleavage by an exosome RNase PH core. Mol. Cell, 20, 473-481

Lorentzen, E., Walter, P., Fribourg, S., Evguenieva-Hackenberg, E., Klug, G. & Conti, E. (2005). The archaeal exosome core is a hexameric ring structure with three catalytic subunits. Nat. Struct. Mol. Biol., 12, 7, 575-581

Madden, D.R., Armstrong, N., Svergun, D., Perez, J. & Vachette, P. (2005). Solution X-ray scattering evidence for agonist- and antagonist-induced modulation of cleft closure in a glutamate receptor ligand-binding domain. J. Biol. Chem., 280, 23637-23642

Márquez , J.A., Hasenbein, S., Koch, B., Fieulaine, S., Nessler, S., Russell, R.B., Hengstenberg, W. & Scheffzek, K. (2002). Structure of the full-length HPr kinase/phosphatase from Staphylococcus xylosus at 1.95 A resolution: Mimicking the product/substrate of the phospho transfer reactions. Proc. Natl. Acad. Sci. USA, 99, 3458-3463

Márquez, J.A., Smith, C.I., Petoukhov, M.V., Lo Surdo, P., Mattsson, P.T., Knekt, M., Westlund, A., Scheffzek, K., Saraste, M. & Svergun, D.I. (2003). Conformation of full-length Bruton tyrosine kinase (Btk) from synchrotron X-ray solution scattering. EMBO J., 22, 4616-4624

Martens, L. et al. (2005). PRIDE: The proteomics identifications database. Proteomics, 5, 3537-3545

Mendjan, S. et al. (2006). Nuclear pore components are involved in the transcription regulation of dosage compensation in Drosophila. Mol. Cell, 21, 811-23

Morand P, Budayova-Spano M, Perrissin M, Muller CW, Petosa C. (2006). Expression, purification, crystallization and preliminary X-ray analysis of a C-terminal fragment of the Epstein-Barr virus ZEBRA protein. Acta Crystallograph Sect F Struct Biol Cryst Commun., 62, 210-4.

Neduva, V. & Russell, R.B. (2005). Linear motifs: evolutionary interaction switches. FEBS Lett., 579, 3342-3345

Neduva, V., Linding, R., Su-Angrand, I., Stark, A., de Masi, F., Gibson, T.J., Lewis, J., Serrano, L. & Russell, R.B. (2005). Systematic discovery of new recognition peptides mediating protein interaction networks. PLoS Biol., 3, e405

Nerlov, C. (2004). C/EBPa mutations in acute myeloid leukaemias. Nat. Rev. Cancer, 4, 394-400

Omenn, G.S. et al. (2005). Overview of the HUPO Plasma Proteome Project: Results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database. Proteomics, 5, 3226-3245

Orchard, S. et al. (2005). Annotating the human proteome. Mol. Cell. Proteomics, 4, 435-440

Orchard, S. et al. (2005). The use of common ontologies and controlled vocabularies to enable data exchange and deposition for complex proteomic experiments. Pac. Symp. Biocomput., 186-196

Ouzounis, C.A. (2005). Ancestral state reconstructions for genomes. Curr. Opin. Genet. Dev., 15, 595-600

Ouzounis, C.A., Kunin V., Darzentas N. & Goldovsky L. (2005). A minimal estimate for the gene content of the last universal common ancestor – exobiology from a terrestrial perspective. Res. Microbiology, in press

Pal, C., Papp, B. & Lercher, M.J. (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat. Genet., 37, 1372-1375

Panjikar, S., Parthasarathy, V., Lamzin, V.S., Weiss, M.S. & Tucker, P.A. (2005). Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the

validation of an X-ray diffraction experiment. Acta Crystallogr. D. Biol. Crystallogr., 61, 449-457

Papp, B. & Müller, J. (2006) Histone tri-methylation and the maintenance of transcriptional ON and OFF states by PcG and trxG proteins. In press.

Pardi, F., Goldman, N. (2005). Species Choice for Comparative Genomics: Being Greedy Works. PLoS Genet., 1, e71

Petosa C, Morand P, Baudin F, Moulin M, Artero JB, Muller CW. (2006). Structural basis of lytic cycle activation by the Epstein-Barr virus ZEBRA protein. Mol. Cell, 21, 565-72

Petoukhov, M.V. & Svergun, D.I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys. J., 89, 1237-1250

Porse, B.T., Bryder, D., Theilgaard-Monch, K., Hasemann, M.S., Anderson, K., Damgaard, I., Jacobsen, S.E. & Nerlov, C. (2005). Loss of C/EBPa cell cycle control increases myeloid progenitor proliferation and transforms the neutrophil granulocyte lineage. J. Exp. Med., 202, 85-96

Porse, B.T., Pedersen, T.A., Hasemann, M.S., Schuster, M.B., Kirstetter, P., Luedde, T., Damgaard, I., Kurz, E., Schjerling, C.K. & Nerlov, C. (2006). The proline-histidine-rich CDK2/CDK4 interaction region of C/EBPa is dispensable for C/EBPa-mediated growth regulation in vivo. Mol. Cell Biol., 26, 1028-1037

Raible, F., Tessmar-Raible, K., Osoegawa, K., Wincker, P., Jubin, C., Balavoine, G., Ferrier, D., Benes, V., de Jong, P., Weissenbach, J., Bork, P. & Arendt, D. (2005). Vertebrate-type intron-rich genes in the marine annelid Platynereis dumerilii. Science, 310, 1325-1326

Rehwinkel, J. et al. (2005). Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. RNA, 11, 1530-1544

Rehwinkel, J., Behm-Ansmant, I., Gatfield, D. & Izaurralde, E. (2005). A crucial role for GW182 and the DCP1:DCP2 decapping complex in miRNA-mediated gene silencing. RNA, 11, 1640-1647

Schleifenbaum, A., Stier, G., Gasch, A., Sattler, M. & Schultz, C. (2004). A genetically encoded FRET probe for PKC activity based on pleckstrin. J. Am. Chem. Soc., 126, 11786-11787

Seitz, A. & Surrey, T. (2006). Processive movement of single kinesins on crowded microtubules visualized using quantum dots. EMBO J., 25, 267-77

Stark, A., Brennecke, J., Bushati, N., Russell, R.B. & Cohen, S.M. (2005). Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3'UTR evolution. Cell, 123, 1133-1146

Stark, A., Brennecke, J., Russell, R.B. & Cohen, S.M. (2003). Identification of Drosophila microRNA targets. PLoS Biol., 1, E60

Surrey, T., Nédélec, F., Leibler, S. & Karsenti, E. (2001). Physical properties determining self-organization of motors and microtubules. Science, 292, 1167-1171

Svergun, D.I., Petoukhov, M.V. & Koch, M.H. (2001). Determination of domain structure of proteins from x-ray solution scattering. Biophys. J., 80, 2946-2953

Taxis, C., Keller, P., Kavagiou, Z., Jensen, L.J., Colombelli, J., Bork, P., Stelzer, E.H. & Knop, M. (2005) Spore number control and breeding in Saccharomyces cerevisiae: a key role for a self-organizing system. J. Cell Biol., 171, 627-640

Tocchini-Valentini GD, Fruscoloni P, Tocchini-Valentini GP. (2005). Coevolution of tRNA intron motifs and tRNA endonuclease architecture in Archaea. PNAS, 102, 15418-22

Velankar, S. et al. (2005). E-MSD: an integrated data resource for bioinformatics. Nucleic Acids Res., 33, D262-D265

Weissenhorn, W. (2005). Crystal structure of the endophilin-A1 BAR domain. J. Mol. Biol., 351, 653-661

Wichmann, O., Wittbrodt, J. & Schultz, C. (2006). A small-molecule FRET probe to monitor phospholipase A(2) activity in cells and organisms. Angew. Chem. Int. Ed. Engl., 45, 508-512

Zou, P., Pinotsis, N., Lange, S., Song, Y.H., Popov, A., Mavridis, I., Mayans, O.M., Gautel, M. & Wilmanns, M. (2006). Palindromic assembly of the giant muscle protein titin in the sarcomeric Z-disk. Nature, 439, 229-233

Notes

Notes

Notes

Index

Annual Report 2005-2006

Written by
Russ Hodge (Scientific Report)
Iain Mattaj (Director General's Report)

Photography
EMBL Photolab: Maj Britt Hansen, Udo Ringeisen, Marietta Schupp, Doros Panayi

Graphics
Petra Riedinger

Layout and Design
Nicola Graf, Vienna Leigh, Russ Hodge

Contributors
Silke Schumacher, Lena Raditsch, Vienna Leigh
Christian Boulin, Cath Brooksbank, Katherine Brown, Malgorzata Duszczyk, Anne Ephrussi, Emma Fassmann, Fabian Filipp, Caroline Hadley, Matthias Haury, Eleanor Hayes, Matthias Hentze, Alexandra Manaia, Doros Panayi, Thomas Portmann, Mehrnoosh Rayner, Pernille Rørth, Bettina Schäfer, Antje Seeck, Halldór Stefánsson, Henriette Uhlenhaut, Stephanie Weil, Julia Willingale-Theune

Additional Photography: Manfred Bernhardt (191, 199, 211 left), Andrey Bogomolov (208, top), Fabian Filipp (192, 208 middle), Heinz-Dieter Gentz (207, middle right), Andy Giddings (211, bottom right), Alexandra Manaia (197), Lisa Mullan (211, top right), Arne Seitz (206, left), Hartwig Valdmanis (209, bottom)

Cover
Photograph by Maj Britt Hansen
Inside cover: Anne-Claude Gavin and Petra Riedinger (design)

DVD
Layout and editing: Vienna Leigh
Design: Francesco Sottile
Production: Sonopress GmbH

Translations
Nicola Fischer (German), Cathérine Moerman (French)

Exposure and printing
ColorDruck Kurt Weber GmbH

Special thanks to the EMBL Group Leaders and Heads of Units, Glauco Tocchini-Valentini, Rainer Bender, Jürg Müller, Stephen Cusack, Dietrich Suck, Eric Karsenti, Walter Witke, the EMBL Mechanical Workshop, Achilleas Frangakis, and Cambridge University Library.

EMBL Heidelberg
Meyerhofstraße 1
69117 Heidelberg
Germany
Tel. +49 (0)6221 387 0
Fax +49 (0)6221 387 8306

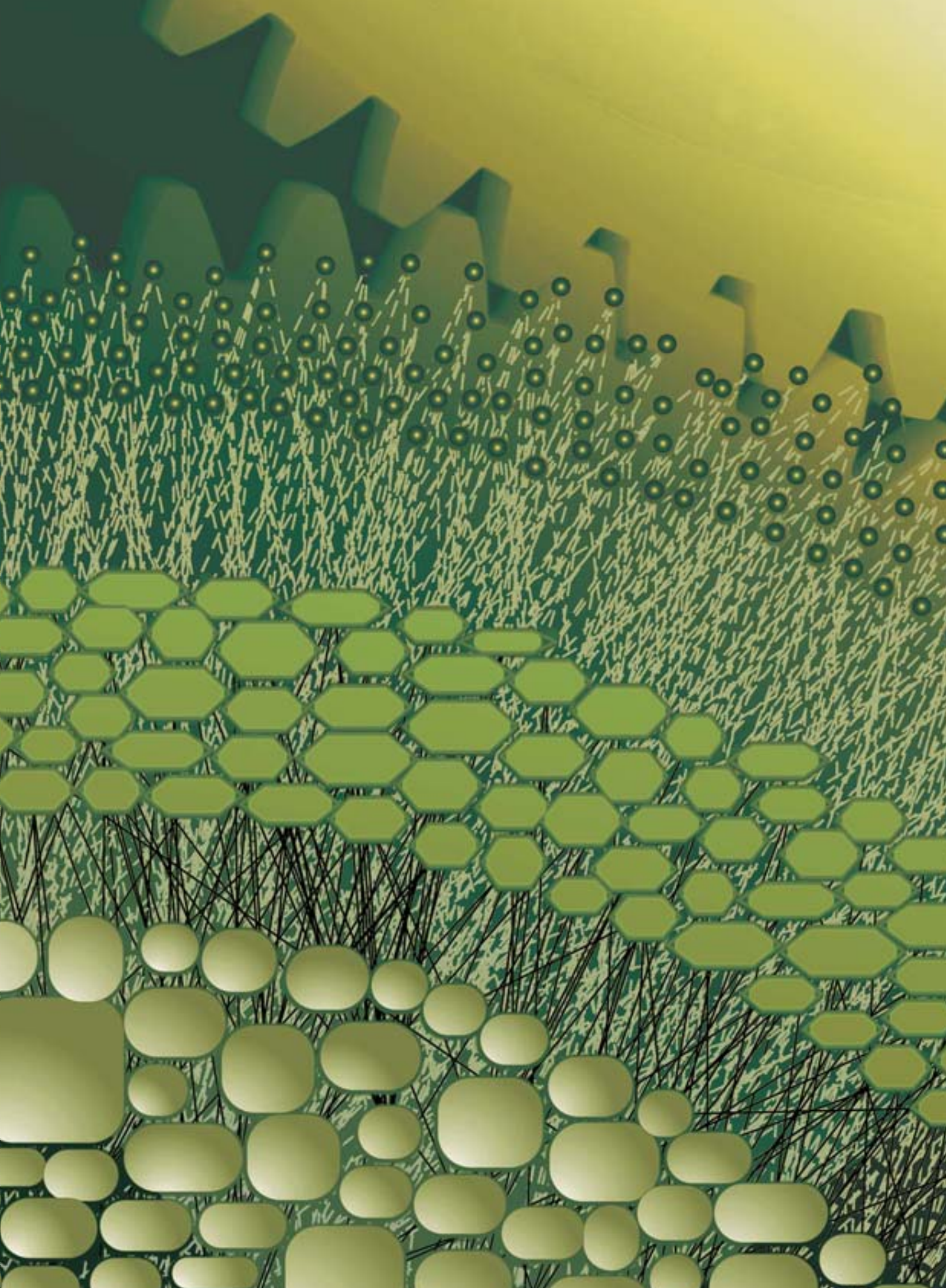
EMBL Grenoble
6, rue Jules Horowitz, BP 181
38042 Grenoble, Cedex 9
France
Tel. +33 (0)4 76 20 72 69
Fax +33 (0)4 76 20 71 99

EMBL Hamburg
c/o DESY
Notkestraße 85
22603 Hamburg
Germany
Tel. +49 (0)40 89 902 0
Fax +49 (0)40 89 902 104

EMBL-EBI
Wellcome Trust Genome Campus, Hinxton
Cambridge CB10 1SD
United Kingdom
Tel. +44 (0)1223 494 444
Fax +44 (0)1223 494 468

EMBL Monterotondo
Adriano Buzzati-Traverso Campus
Via Ramarini, 32
00016 Monterotondo (Rome)
Italy
Tel. +39 06 900 912 85
Fax +39 06 900 912 72

For more information about EMBL and its activities, visit www.embl.org



EMBL

Annual Report

2005-2006

EMBL Member States:
Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom