EMBL Research at a Glance 2007-2008

EMBL

EMBL Heidelberg
Meyerhofstraße 1
69117 Heidelberg
Germany
Tel. +49 (0)6221 387 0, Fax +49 (0)6221 387 8306
www.embl.org
info@embl.org

EMBL Grenoble
6, rue Jules Horowitz, BP181
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 90 20, Fax +49 (0)40 89 90 21 04

EMBL-EBI (European Bioinformatics Institute)
Wellcome Trust Genome Campus, Hinxton
Cambridge CB10 1SD
United Kingdom
Tel. +44 (0)1223 494444, Fax +44 (0)1223 494468

EMBL Monterotondo
Adriano Buzzati-Traverso Campus
Via Ramarini, 32
00015 Monterotondo (Rome)
Italy
Tel. +39 06 90091285, Fax +39 06 90091272

EMBL's Member States:

Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and United Kingdom

Texts:
EMBL Group and Team Leaders and Heads of Units

Layout and editing:
Vienna Leigh, EMBL Office of Information and Public Affairs

Cover Design:
Vienna Leigh

Exposure and printing:
ColorDruck Kurt Weber GmbH, Leimen

European Molecular Biology Laboratory

Heidelberg • Grenoble • Hamburg • Hinxton • Monterotondo

Contents

EMBL – Europe's flagship laboratory for basic research in molecular biology

The vision of the European nations who founded the European Molecular Biology Laboratory was to create a centre of excellence where Europe's best brains would come together to conduct basic research in molecular biology. During the past three decades, EMBL has grown and developed substantially. Its member states now number nineteen. Over the years EMBL has become the flagship of European molecular biology and is ranked as one of the top research institutes worldwide.

EMBL's missions are to perform cutting-edge research in molecular biology, to offer services to European scientists, to provide advanced training to researchers at all levels, to develop new technologies and instrumentation and to actively engage in technology transfer to the benefit of scientists and society.

In research, the five EMBL sites (a central laboratory in Heidelberg, with outstations in Grenoble, Hamburg, Hinxton and Monterotondo) put strong emphasis on interdisciplinarity and collaboration, and when the researchers leave to assume key positions in the member states, they export their unique experience of working in a very energetic and international environment. Freedom, flexibility and a regular turnover of staff allows EMBL to pursue the most exciting themes in molecular biology as they arise. Our long-standing tradition of organising excellent courses, conferences and workshops and an extensive outreach programme ensure that know-how spreads further and engages the public in a dialogue about the impact of modern biology on our lives.

In *Research at a Glance* you will find a concise overview of the work of our research groups and core facilities. Science at EMBL covers themes ranging from studies of single molecules to an understanding of how they work together in complex systems to organise cells and organisms. Our research is loosely structured under thematic units, giving scientists the intellectual freedom to pursue the topics that most interest them. The research units are complemented by "Centres", inter-unit collaborations, bringing people with common interests together. Networking and training within the Centres further support scientists working on interdisciplinary projects. Increasingly, our young scientists come with physics, chemistry, mathematics and computer science backgrounds, bringing in expertise that helps us to move into the growing field of systems biology.

EMBL combines a critical mass of expertise and resources with organisational flexibility, enabling us to keep pace with today's biology. The impact of the laboratory's scientific work, the quality of its services and its continued attractiveness to world-leading young scientists are testimony to EMBL's success and show that we are well-equipped for the future.

Iain Mattaj
EMBL Director General

EMBL Heidelberg, Germany

A city of about 130,000 inhabitants, Heidelberg is home to Germany's oldest university, as well as leading biological and medical science institutes such as the Centre for Molecular Biology, the German Cancer Research Center (DKFZ) and the Max Planck Institute for Medical Research, making it an ideal site for EMBL's Main Laboratory. Nestling in the wooded hills above the city, the complex is home to five of EMBL's scientific Units, namely Gene Expression, Cell Biology and Biophysics, Developmental Biology, Structural and Computational Biology and Directors' Research, as well as Core Facilities and the central administration, from which service functions are provided for the use of staff working at all five EMBL sites.

Today more than 800 staff members are located at EMBL Heidelberg, and the close proximity of the other excellent institutes has led to numerous long-term collaborations. EMBL shares a campus with its sister organisation, the European Molecular Biology Organization. The two share strong historical ties and work together in many ways; for example, they combine to stage many highly-recognised international courses and conferences.

Heidelberg is also home to EMBLEM, the Laboratory's technology transfer company. Other important central functions can be found on the Heidelberg campus, such as the offices of the International PhD Programme and the Science and Society Programme.

Cell Biology and Biophysics Unit

The cell is the basic unit of life. Interestingly, living cells occupy the precise midpoint between the molecular and macroscopic scales. Thus, in order to understand how organisms are built and how they function, we need to understand the molecular mechanisms and physical principles that give rise to cellular organisation.

All cells (including prokaryotes) are divided into functional domains, each with different molecular compositions. In addition, eukaryotes have compartments such as the nucleus, the cytoplasm, the cytoskeleton and the membrane trafficking system.

Research in the Cell Biology and Biophysics Unit focuses on the mechanisms and principles that underlie the dynamic organisation of these compartments and the distribution of specific molecules to each cellular sub-system. As a cell enters mitosis, all the microtubules suddenly disassemble to reorganise into the mitotic spindle. At the same time, the nuclear envelope disappears, the Golgi complex fragments and membrane traffic ceases. During development, when progenitor cells differentiate into new cell types, the daughter cells receive a complement of chromosomes and organelles from the parent cell, while the genetic program is changed. A reorganisation of cellular architecture takes place, guided by mechanisms that begin to be unraveled. The elucidation of such mechanisms and the principles that govern them is a major challenge to contemporary biology.

This is the challenge that the Cell Biology and Biophysics Unit pursues, both conceptually and methodologically. The areas that we are presently concentrating on are membrane trafficking, signalling and cytoskeletal networks and their role in cell organisation and organogenesis as well as mitosis. From this work, a new vision of evolution is also beginning to emerge. Physicists working together with cell biologists are trying to elucidate the fundamental rules that root the organisation properties of living matter. This involves the use of ever better technology and analytical tools. Novel developments in fluorescence microscopy, the identification of gene products and their localisation, as well as in the use of computer simulations to understand complex self-organisation processes are a particular strength of the unit. New concepts applied from other fields of science such as chemistry, statistical physics and soft matter physics have also begun to diffuse through the unit and help understand some surprising self-organisation properties of cell components, groups of cells and gene networks.

Eric Karsenti
Head of Cell Biology and Biophysics Unit

Eric Karsenti

PhD 1979, University of Paris, France.
Postdoctoral research at University of California, San Francisco, USA.
Group Leader at EMBL since 1985.
Head of Unit since 1998.

Self-organisation principles in cell morphogenesis

Previous and current research

We have been working on the mechanism of mitotic spindle assembly in *Xenopus* egg extracts and cells using live cell imaging, FRET technology and computer simulations, in collaboration with physicists and colleagues in the Cell Biology and Biophysics Unit. We found that in large cells and frog eggs, spindle assembly involves two self-organisation principles: 1) the establishment of a gradient of regulators that affect microtubule nucleation and dynamics in the cytoplasm surrounding the chromosomes, and 2) the collective behaviour of dynamic microtubules and motors that results in the assembly of a steady state bipolar structure.

These observations reveal the existence of general principles in the formation of dynamic large cellular structures like the mitotic spindle. We can define such principles as: 1) the stochastic random motion of molecules interacting specifically with each other in the cytoplasm; 2) reaction-diffusion processes that define spatial distributions of regulatory molecules; and 3) collective behaviour of cytoskeletal systems that self-organise in large structures. Principle 2) may coordinate 1) and 3) to generate a precise dissipative structure that maintains its shape over time.

Initial experiments carried out on muscle cell morphogenesis suggest that similar principles are at work for the determination of cell shapes during cell differentiation.

Another important question concerns the nature of interphase and mitotic cytoplasmic 'states'. Frog eggs are interesting because following fertilisation they undergo a series of twelve divisions during which the cell cycle alternates only between two states: S-phase, during which a nucleus forms and DNA is replicated, and mitosis, during which a spindle forms and chromosomes are segregated. The switch between the two states involves only the continuous synthesis of cyclin coupled to its periodic degradation. When cyclin accumulates to a threshold, it triggers the activation of cdc2 kinase that sends the egg cytoplasm in the mitotic state. The problem is to define what we mean by 'cytoplasmic state'. We need to take one morphological effect of the interphase and mitotic states and check what is different in the cytoplasm and how such a difference can be brought about by cdc2 kinase. We have chosen to look at microtubule dynamic states. In interphase, microtubules are stable and long; in mitosis, they are short and dynamic. To characterise what could determine this difference, we looked at the interaction state of several molecules involved in the regulation of microtubule dynamics and correlated the state of this network with the effect of removing each component one-by-one on microtubule dynamics in interphase and mitosis. We found that two microtubule stabilisers need to interact in the mitotic situation to permit microtubule growth, whereas in interphase only one stabiliser is enough to generate very long and stable microtubules. On the other hand, a strong microtubule destabiliser is active in mitosis and almost entirely inactive in interphase, because its activity is damped through an interaction with several of the stabilisers. Therefore, the cdc2 kinase changes the state of interaction between a series of regulators when cells enter mitosis and the interphase and mitotic 'states' can therefore be defined by the state of this network as far as microtubule dynamics are concerned.

Future projects and goals

Concerning the mechanism of spindle assembly, there are issues that still need to be addressed generally concerning the principles underlying spindle morphogenesis during evolution. We will use computer simulations in collaboration with the Nédélec lab to establish the parameter space within which various spindles may form. We hope, in the near future, to have a full predictive description of the conditions that would support spindle assembly in vertebrate cells. Similar approaches will be used to examine the conditions of spindle assembly in various species.

Concerning the establishment of interphase and mitotic microtubule dynamic states, we are in the process of identifying the phosphorylation sites involved in the change in the pattern of MAP interactions between interphase and mitosis. We therefore hope to characterise not only the structure of the MAP interaction network but also the structure of the regulatory system, including the nature of the kinase and phosphatase networks involved.

Selected references

Bastiaens, P., Caudron, M., Niethammer, P. & Karsenti, E. (2006). Gradients in the self-organization of the mitotic spindle. *Trends Cell Biol*., 16, 125-134

Brunet, S., Zimmermann, T., Reynaud, E.G., Vernos, I., Karsenti, E. & Pepperkok, R. (2006). Detection and quantification of protein-microtubules interactions using green fluorescent protein photoconversion. *Traffic*, 7, 1283-9

Karsenti, E., Nedelec, F. & Surrey, T. (2006). Modelling microtubule patterns. *Nat Cell Biol*., 8, 1204-1211

Pizon, V., Gerbal, F., Diaz, C.C. & Karsenti, E. (2005). Microtubule-dependent transport and organization of sarcomeric myosin during skeletal muscle differentiation. *EMBO J*., 24, 3781-3792

Claude Antony

PhD 1984, University Paris VI, France.
Postdoctoral research at EMBL.
Group Leader at CNRS 1994-2003.
Facility Head and Team Leader at EMBL since 2003.

Electron tomography of fission yeast cytoskeleton

Previous and current research

The microtubule interphase array in fission yeast: an electron tomography approach. Our focus of interest is the organisation of the fission yeast microtubular cytoskeleton using electron tomography (Johanna Höög and Helio Roque, PhD students). Electron tomography is our method of choice as it not only allows the reconstruction, modelling and quantifying of subcellular elements, but also enables the visualisation of structural features in very fine detail, for instance microtubule ends morphology. These data allow a better understanding of the functional organisation of the cytoskeleton and its relationship to membranous organelles.

This year, full fission yeast volume was reconstructed in an eukaryotic cell for the first time (Johanna Höög). As a result, we can show and analyse microtubule bundles in their entirety, with information on microtubule polarity, nucleus position and relationship between mitochondria and microtubules. We also modelled all membrane-enclosed organelles at the cell ends, and showed that among three identified categories of transport vesicle, none seemed to be associated with the MT bundles in the cytoplasm (Höög *et al.*, *Dev. Cell*, 2007).

Interphase microtubule bundling factors. Ase1p function in *S. pombe* is to organise MTs in overlapping antiparallel bundles in interphase. To understand the inter-microtubule bonds, we reconstructed MT arrays in an ase1 deletion mutant by electron tomography (Helio Roque). We showed that the bundles are highly disorganised with individualised MTs, while the overall MT array had lost its parallel orientation to the cell axis. The SPB bundles contain fewer MTs which are also shorter than in WT. Finally, the spacing of MTs in ase1 deletion was reduced compared to WT cells. Nevertheless, some anchoring points between MTs remained, as well as some MT overlapping zones. Therefore, other factors may be involved to explain the residual MT anchoring at their tips to the lattice of other MTs.

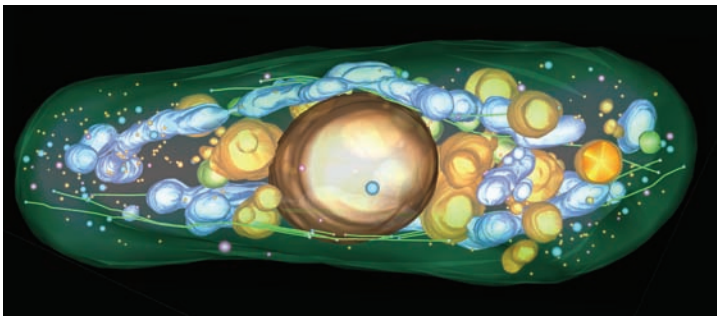
Budding yeast spindle organisation. Trainees Xavier Heiligenstein and Wandrille George have been investigating the structural organisation of the mitotic spindle in ipl1-321 mutant cells in budding yeast that affect kinetochore function and show defects in chromosome biorientation.

Future projects and goals

The mutant analysis of Mal3 (EB1) and Tip1 (CLIP 170) deletions (Johanna Höög, in collaboration with Damian Brunner) needs to be completed to characterise precisely their aberrant MT phenotype. Tip1 deletion also showed reduced MT polymers, and the MT to SPB connection needs to be carefully investigated. Finally, cells treated with a depolymerizing drug and then allowed to repolymerise their MTs will be analysed by 3DEM to understand MT function and dynamic end structures.

Interaction of microtubules and the nuclear envelope will be carried out by Johanna Höög and Helio Roque in collaboration with Damian Brunner. Nuclei volumes and the associated MTs will be reconstructed using tomography to visualise physical connections between the two.

Concerning the bundling factors, Helio Roque will further identify the bridges between the microtubules in the ase1 mutant as well as in the double deletion mutant of klp2 ase1, as the motor protein may also be responsible for connecting MT ends.



Electron tomography reconstruction of the full cell volume of S. pombe *WT with all microtubule bundles (green), mitochondria (blue), vacuoles (yellow), and trafficking vesicles.*

Selected references

Höög, J.L.., Schwartz, C., Noon, A.T., O'Toole, E.T., Mastronarde, D.N., McIntosh, J.R. & Antony, C. (2007). Organization of interphase microtubules in fission yeast analyzed by electron tomography. *Dev. Cell,* 12, 349-361

Höög, J.L. & Antony, C. (2007). Whole-cell investigation of microtubule cytoskeleton architecture by electron tomography. *Methods Cell Biol.*, 79, 145-67

Carazo-Salas, R.E., Antony, C. & Nurse, P. (2005). The kinesin Klp2 mediates polarization of interphase microtubules in fission yeast. *Science*, 309, 297-300

Tanaka, K., Mukae, N., Dewar, H., van Breugel, M., James, E.K., Prescott, A.R., Antony, C. & Tanaka, T.U. (2005). Molecular mechanisms of kinetochore capture by spindle microtubules. *Nature*, 434, 987-994

Damian Brunner

PhD 1995, University of Zurich, Switzerland.
Postdoctoral research at the Imperial Cancer Research Fund, London, UK.
Group Leader at EMBL since 2000.
Joint appointment with the Developmental Biology Unit.

Cell morphogenesis and spatial microtubule organisation

Previous and current research

To create a defined morphology, cells need to polarise and correctly orient their polarity axis. Both processes require a defined intracellular order, which is based on the specific sub-cellular arrangement of actin and microtubule filaments. Our investigations focus on the contribution of microtubules. Microtubule organisation varies tremendously between different cell types and also in individual cells during different developmental or cell cycle stages. Little is known about how this variability is achieved and how cells switch from one organisational state to another. We address these questions in two model organisms, the unicellular fission yeast *Schizosaccharomyces pombe* and the fruit fly *Drosophila melanogaster*.

In the cylindrical fission yeast cells we describe the cell autonomous machinery that organises and maintains a defined interphase microtubule distribution. In these cells, approximately 30 anti-parallel microtubules form 3-6 bundles that are arranged parallel to the long cell axis. Microtubule minus-ends overlap in the cell centre. From there the plus-ends grow to the cell poles where they switch to shrinkage, an event termed catastrophe (Fig. 1). This localised catastrophe is fundamental to proper cell morphology. Our findings suggest a model where conserved proteins at the growing microtubule plus-ends (+TIPs) mediate cell-pole targeting in two steps. First, the yeast EB1 homolog Mal3p promotes growth until cortical regions are encountered. In central regions of the cell cortex, the CLIP-170 homolog Tip1p then prevents premature catastrophes by suppressing Mal3p removal, which keeps microtubules growing below the cortex until the cell poles are reached (reference 1).

A central question concerning +TIP function is how these proteins can accumulate at growing microtubule plus-ends. We discovered that Tip1p is transported there by the Tea2p motor protein (reference 2). Mal3p in contrast, seems to 'treadmill', preferentially binding to plus-ends followed by rapid unbinding. In an attempt to further describe Mal3p plus-end binding (collaboration with Andreas Hoenger, EMBL/Boulder), we discovered that Mal3p also binds and stabilises the microtubule lattice seam, explaining its weak localisation all along the microtubules (reference 3). This provides a new twist to the model of how microtubule dynamics are controlled and shows that microtubules have two different surfaces for molecular interactions.

In fruit flies we explore to what extent the basic machinery found in fission yeast is used to maintain microtubule organisation in a multi-cellular organism and how non-autonomous cells achieve and coordinate changes in microtubule distribution. We have shown how microtubules become re-organised during embryonic dorsal closure (DC), a wound healing-related process. Thereby, anti-parallel microtubules transiently form bundles in the epidermal cells that move dorsally to close a cavity (Fig. 2). Surprisingly, these microtubules are essential exclusively for the final step, the fusion of the epithelium (reference 4).

Future projects and goals

In fission yeast we now focus on two topics. First, we want to understand how catastrophes are induced at cell poles and second we want to identify the critical molecules/processes for switching between the 7 different microtubule arrangements found in *S. pombe* cells.

In flies, we are trying to identify the signals triggering microtubule reorganisation during DC and we want to uncover the molecular mechanisms driving the process.

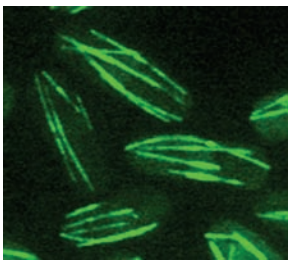


Figure 1: Interphase microtubule bundles in fission yeast cells.

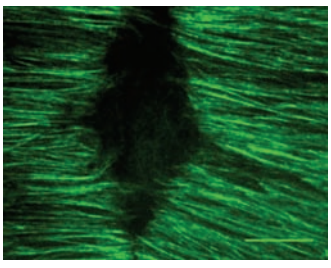


Figure 2: Microtubule bundles in the epidermal cells during dorsal closure in Drosophila melanogaster.

Selected references

Sandblad, L., Busch, K.E., Tittmann, P., Gross, H., Brunner, D. & Hoenger, A. (2006). The *Schizosaccharomyces pombe* EB1 homolog Mal3p binds and stabilizes the microtubule lattice seam. *Cell*, 127, 1415-1424

Jankovics, F. & Brunner, D. (2006). Transiently reorganized microtubules are essential for zippering during dorsal closure in *Drosophila melanogaster*. *Dev. Cell*, 11, 375-385

Busch, K.E., Hayles, J., Nurse, P. & Brunner, D. (2004). Tea2p kinesin is involved in spatial microtubule organization by transporting tip1p on microtubules. *Dev. Cell*, 6, 831-843

Busch, K.E. & Brunner, D. (2004). The microtubule plus end-tracking proteins mal3p and tip1p cooperate for cell-end targeting of interphase microtubules. *Curr. Biol.*, 14, 548-559

Darren Gilmour

PhD 1996, Cambridge University, UK.
Postdoctoral research at the Max Planck Institute for Developmental Biology, Tübingen, Germany.
Group Leader at EMBL since 2004.

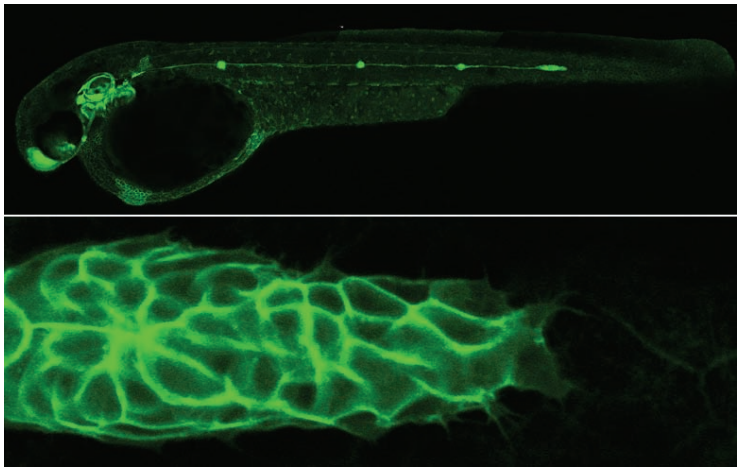
The role of collective cell migration during organ morphogenesis

Previous and current research

The coordinated migration of cohesive groups of cells is a hallmark of both morphogenesis and tumour metastasis. Such 'collective' cell migration sculpts the shape of many complex organs and yet very little is known about the logic underlying these coordinated movements. The zebrafish lateral line primordium is a cluster of 200 migrating cells that has several innate features, such as excellent imaging potential and genetic tractability, that allow this process to be studied in the four-dimensional context of the intact developing embryo. We have generated a set of transgenic reporter lines that allow the migration of cells of the lateral line to be captured by multicolour timelapse. In addition, genetic screens have lead to the isolation of a number of molecules required for this process, most notably CXCR4, a chemokine-receptor required for the guidance and shaping of this migrating tissue. In embryos lacking this receptor or its ligand SDF-1, the cells of the cluster attempt to move in random directions but remain together, resulting in a primordium that is misshapen and immotile. We are currently carrying out a genetic mosaic analysis to determine the role of this chemokine receptor and other key signalling pathways in controlling individual cell behaviours within this moving tissue.

Future projects and goals

Our aim is to understand the mechanisms that coordinate cell behaviour and morphology across a migrating tissue. We will develop probes that allow us to quantify the activity of CXCR4 and other key regulators to determine the extent to which local changes in signalling levels correlate with differences in migratory behaviour *in vivo*. Dynamic interactions between migrating cells are likely to play an equally important role and we are currently screening for molecules mediating these cell–cell interactions using genetics and expression profiling. Previous data suggests that such mediators encode cell adhesion molecules whose activities are dynamically regulated. It is our hope that 4D-imaging of cytoskeletal dynamics in living embryos will reveal how changes in cell organisation spread across moving tissues during organogenesis.



The zebrafish migrating lateral line organ allows collective migration to be easily studied in vivo.

Selected references

Gilmour, D., Knaut, H., Maischein, H.M. & Nüsslein-Volhard, C. (2004). Towing of sensory axons by their migrating target cells *in vivo*. *Nat Neurosci.,* 7, 491-2

Gilmour, D.T., Maischein, H.M. & Nüsslein-Volhard, C. (2002). Migration and function of a glial subtype in the vertebrate peripheral nervous system. *Neuron,* 34, 577-88

Haas, P. & Gilmour, D. (2006). Chemokine-signaling mediates self-organizing tissue migration in the zebrafish lateral line. *Developmental Cell,* 10, 673-680

Lecaudey, V. & Gilmour, D. (2006). Organizing moving groups during morphogenesis. *Curr. Opin. Cell Biol.*, 18, 102-7

Gareth Griffiths

PhD 1975, University of Wisconsin, USA.
Postdoctoral research at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany.
At EMBL since 1977.

Cell biology of pathogens

Previous and current research

Our group analyses how mycobacteria interact with macrophages. They enter by phagocytosis, a process which delivers the bacteria inside phagosomes. This membrane organelle normally fuses with several organelles of the endocytic pathway of macrophages in order to provide an environment in the phagosomal lumen that can kill the pathogen. Phagosomes enclosing pathogenic mycobacteria such as *M. tuberculosis* fail to fuse with lysosomes and can thus survive and even grow. We showed recently that pro-inflammatory lipids help to kill *M. tuberculosis* while anti-inflammatory lipids facilitate pathogen growth in macrophages (Anes *et al.*, 2003, *Nature Cell Biol.*; Griffiths, 2004, *Trends Cel. Biol.*).

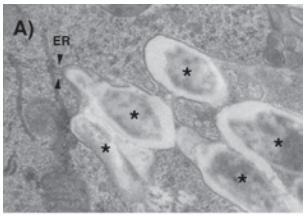
We also use latex bead phagosomes (LBP) as a model to describe the mechanisms of phagosome maturation in more detail. The focus over the past 5-6 years has been on the mechanisms by which the phagosomal membrane assembles actin filaments *de novo*. Our data have led to a hypothesis that the role of this process is to provide tracks for phagosomes to fuse with lysosomes (Kjeken *et al.*, 2004. *Mol. Biol. Cell*). Our work with mycobacteria (Anes *et al.*, 2003; 2006) supports this model.

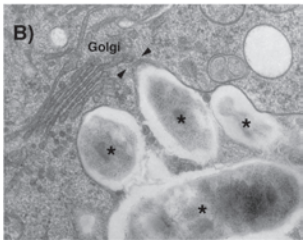
In the LBP system we study the detailed signalling networks regulating actin assembly by phagosomes. This included a systems biology analysis in collaboration with bioinformatic specialists (Thomas Dandekar, Jens Reich). A number of phosphorylated lipids, such as the phosphoinosites PI,4P and PI4,5P2 and sphingosine-1-phosphate (S1P), can stimulate the ability of LBP (and mycobacterial phagosomes) to nucleate actin assembly. Our experiments show that these phosphorylated lipids induce the transport of ADP across the phagosomal membrane into the lumen. There, it is converted to ATP by an adenylate kinase activity. We are speculating that this ATP binds to the ATP receptor P2X7, found in the phagosome, which then interacts on its cytoplasmic side with the actin assembly machinery. Other studies in the lab argue that this machinery involves ezrin and/or moesin in a complex with N-Wasp (Sabrina Marion).

Support for this model was also obtained on the plasma membrane of macrophages. There the addition of serum or S1P induces an immediate release of ADP and a soluble adenylate kinase activity into the culture medium; ADP that comes out is converted to ATP. This binds an ATP receptor that stimulates plasma membrane actin assembly after 5 sec. Patch clamp analysis by Fritz Marqwardt has identified the ATP receptor that is linked to actin assembly as the P2X7 receptor (Kuehnel *et al.*, submitted).

In the mycobacterial project, a collaboration with Elsa Anes's group in Lisbon, is now focused on the mechanisms by which macrophages kill mycobacteria. We initially addressed how the non-pathogenic *M. smegmatis* is killed (Anes *et al.*, 2006, *Cell Microbiology*) and more recently we have started focus also on the pathogens *M. tuberculosis* and *M. bovis*. In the *M. smegmatis* system much information has been provided by RNA expression microarray analysis of the genes up-regulated in response to *M. smegmatis*. We are especially interested in pro-inflammatory gene products under the control of the inflammatory transcription factor NfkB that are up-regulated in response to infection.

The ultrastructure of macrophages infected with mycobacteria using high pressure freezing and freeze substituion is an additional ongoing project. The figures shows examples of RAW mouse macrophages infected for 2h with *M. smegmatis* prepared by high pressure freezing and freeze substitution. These images reveal likely connections between the phagosome and the ER (a controversial issue) and with the Golgi complex.





Future projects and goals

In macrophages: to identify the S1P receptor that responds to S1P to induce ADP release and to identify the adenylate kinase that is released, and the mechanism of release. The role of different proteins implicated in the actin assembly such as P2X7, ezrin/moesin will be addressed by RNAi approaches, in combination with functional assays including patch clamp analysis. In the mycobacterial project the major goal is to understand how macrophages kill mycobacteria.

Selected references

Anes, E., Peyron, P., Staali, L., Jordao, L., Gutierrez, M.G., Kress, H., Hagedorn, M., Maridonneau-Parini, I., Skinner, M.A., Wildeman, A.G., Kalamidas, S.A., Kuehnel, M. & Griffiths, G. (2006). Dynamic life and death interactions between *Mycobacterium smegmatis* and J774 macrophages. *Cell Microbiol.*, 8, 939-960

Kalamidas, S.A., Kuehnel, M.P., Peyron, P., Rybin, V., Rauch, S., Kotoulas, O.B., Houslay, M., Hemmings, B.A., Gutierrez, M.G., Anes, E. & Griffiths, G. (2006). cAMP synthesis and degradation by phagosomes regulate actin assembly and fusion events: consequences for mycobacteria. *J. Cell Sci.*, 119, 3686-3694

Anes, E., Kuhnel, M.P., Bos, E., Moniz-Pereira, J., Habermann, A. & Griffiths, G. (2003). Selected lipids activate phagosome actin assembly and maturation resulting in killing of pathogenic mycobacteria. *Nat. Cell Biol.*, 5, 793-802

de Bernard, M., Moschioni, M., Habermann, A., Griffiths, G. & Montecucco, C. (2002). Cell vacuolization induced by *Helicobacter pylori* VacA cytotoxin does not depend on late endosomal SNAREs. *Cell Microbiol.*, 4, 11-18

Marko Kaksonen

PhD 2002, University of Helsinki, Finland.
Postdoctoral research at the
University of California, Berkeley, USA.
Group leader at EMBL since 2006.

Dynamics of membrane trafficking

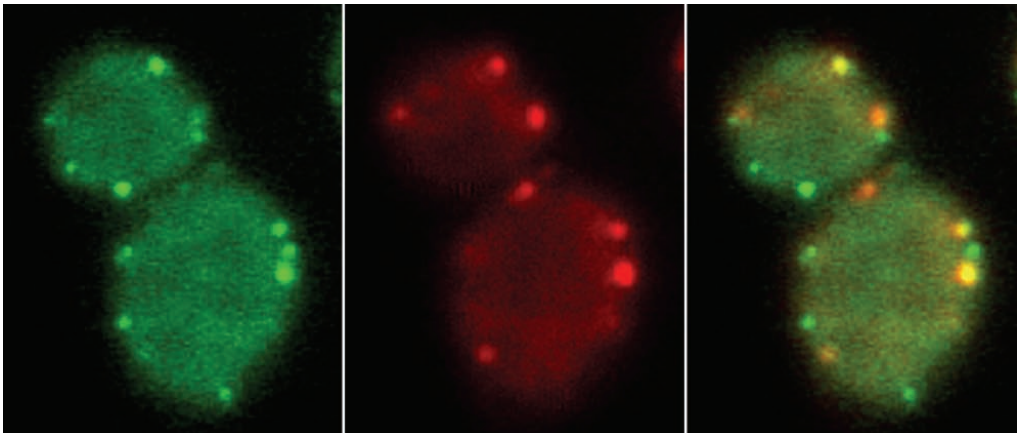
Previous and current research

Many biological processes at the cellular level are based on complex networks of macromolecular interactions. These networks have modular organisation, where the modules form dynamic molecular machines that drive processes such as signalling, cell motility, cytokinesis and vesicle trafficking. Our laboratory's long-term goal is to contribute to the understanding of the general principles governing the assembly and function of these supramolecular machines.

More specifically, we are interested in the formation of cargo-loaded transport vesicles, such as endocytic vesicles. The formation of the endocytic vesicle is driven by a highly dynamic molecular machinery composed of more than 50 different protein species and several thousand individual protein molecules. Our main experimental organism is budding yeast, *Saccharomyces cerevisiae*. We combine powerful yeast genetics with quantitative live-cell imaging methods. Using live-cell imaging we have shown that the endocytic proteins assemble at the endocytic sites in a highly regulated sequence and form modular machinery that drives vesicle formation. Using mutant yeast strains we have revealed specific roles for numerous proteins in this process.

Future projects and goals

In the future, we will continue to study the membrane trafficking events in budding yeast using live-cell imaging combined with yeast genetics. We will focus on the mechanisms of the assembly of the clathrin-based endocytic machinery and the mechanisms of selective recruitment of cargo molecules into the endocytic vesicle. We will also extend our work to trafficking events at the Golgi complex. These membrane trafficking events are highly conserved elemental processes that are involved in multiple biological phenomena ranging from cell polarisation to neural plasticity. As most of the yeast trafficking proteins are widely conserved in eukaryotes, we believe that the mechanisms we unravel in yeast cells will be applicable to eukaryotes in general.



A yeast cell expressing fluorescently-labelled endocytic proteins. The first two images show Sla1 (green) and Abp1 (red) proteins. The last image shows both channels merged. The spots at the cell surface reveal accumulation of the proteins at endocytic sites. The protein composition of endocytic machinery changes dynamically during vesicle formation.

Selected references

Kaksonen, M., Toret, C.P. & Drubin, D.G. (2006). Harnessing actin dynamics for clathrin-mediated endocytosis. *Nat. Rev. Mol. Cell Biol.*, 7, 404-14

Liu, J., Kaksonen, M., Drubin, D.G. & Oster, G. (2006). Endocytic vesicle scission by lipid phase boundary forces. *Proc. Natl. Acad. Sci. USA*, 103, 10277-82

Kaksonen, M., Toret, C.P. & Drubin, D.G. (2005). A modular design for the clathrin- and actin-mediated endocytosis machinery. *Cell*, 12, 305-20

Kaksonen, M., Sun, Y. & Drubin, D.G. (2003). A pathway for association of receptors, adaptors, and actin during endocytic internalization. *Cell*, 2, 115, 475-87

Michael Knop

PhD 1995, University of Stuttgart, Germany.
Postdoctoral research at the Max Planck Institute for Biochemistry, Munich, Germany, and the Beatson Institute for Cancer Research, Glasgow, Scotland.
Group Leader at the Max Planck Institute for Biochemistry, Munich, Germany.
Group Leader at EMBL since 2001.

Meiosis and sex in budding yeast

Previous and current Research

Our laboratory is interested in sexual cell division (meiosis) and reproduction with a focus on the connection of the evolutionary/population genetic implications with the underlying molecular mechanisms. We use bakers yeast, *Saccharomyces cerevisiae*, and other yeast species as model systems.

In budding yeast, we have shown that meiosis is governed by a regulatory mechanism that generates a particular outcome (spores) in such a way that their number is adjusted to available resources (nutrients) and that simultaneously they match a particular and advantageous mating behaviour.

We are also studying the ways in which basic molecular machinery, such as exocytotic vesicle fusion or organisation of the cytoskeleton, becomes regulated during meiosis, so that it generates the different output seen for the meiotic type of cell division (spores) as compared to mitosis (bud).

Future projects and goals

Among other projects, we are currently working on the following processes:

Regulation of spindle pole body function during meiosis. This project is based on a series of previous work (e.g. Gordon *et al.*, 2006) and aims to understand the cell cycle regulation of spindle pole body functions in meiosis II.

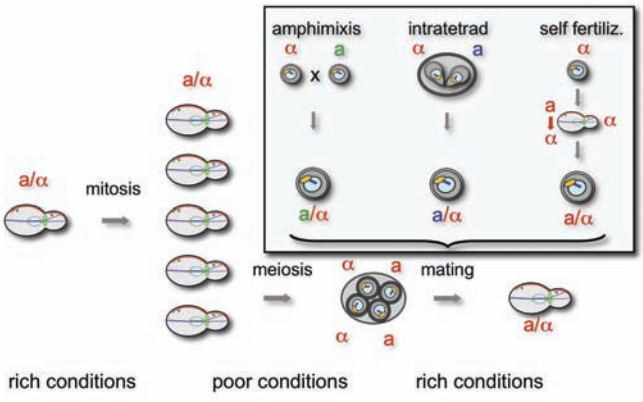
Cell shape generation during meiosis. This project aims to explore the molecular mechanisms that control the round shape of the spores. Currently we investigate the function of a series of process-specific proteins using the various methods available with yeast. This project is also connected to the analysis of the function of the actin cytoskeleton in meiosis (Taxis *et al.*, 2006) and of a new, septin like molecule (Maier *et al.*, 2007).

Impact of different breeding strategies on genome organisation. This project was initiated by our finding that essential genes tend to cluster next to centromeres (Taxis *et al.*, 2005). Here we use Monte Carlo Simulations to explore the forces that act during genome evolution, and how they impact gene order. This work is done in collaboration with Philippe Keller (see Stelzer Group, page 17).

Other yeast-related projects:

MAP kinase signalling. Besides meiosis and sporulation, we are also interested in the process of mating, where two haploid cells fuse in order to form a diploid cell. Here, we investigate the MAP kinase signalling cascade that is needed for the cells of opposite mating type to sense each other and to trigger the cell fusion processes.

We primarily focus on complexes between MAP-kinases and scaffolding molecules, and we use quantitative microscopic approaches to study the abundance of these proteins, their complex formation and subcellular localisation as a function of the signalling processes. This work is done in collaboration with the Bastiaens' lab (visiting).



The life cycle of budding yeast is adapted to fluctuating environments in which poor and rich phases follow each other. The spores constitute the life cycle stage adapted to adverse conditions. Upon return to a rich environment, spores germinate and seek mating partners in order to return to a diploid life cycle stage and to prepare for a subsequent period of poor nutrient supply. Upon encounter of it, they induce meiosis and form spores.

Selected references

Maier, P., Rathfelder, N., Finkbeiner, M. G., Taxis, C., Mazza, M., Le Panse, S., Haguenauer-Tsapis, R. & Knop, M. (2007). Cytokinesis in yeast meiosis depends on the regulated removal of Ssp1p from the prospore membrane. *EMBO J.*, 26, 1843-52

Gordon, O., Taxis, C., Keller, P.J., Benjak, A., Stelzer, E.H., Simchen, G. & Knop, M. (2006). Nud1p, the yeast homolog of Centriolin, regulates spindle pole body inheritance in meiosis. *EMBO J.*, 25, 3856-3868

Knop, M. (2006). Evolution of the hemiascomycete yeasts: on life styles and the importance of inbreeding. *Bioessays*, 28, 696-708

Taxis, C., Maeder, C., Reber, S., Rathfelder, N., Miura, K., Greger, K., Stelzer, E.H. & Knop, M. (2006). Dynamic organization of the actin cytoskeleton during meiosis and spore formation in budding yeast. *Traffic*, 7, 1628-1642

Taxis, C., Keller, P., Kavagiou, Z., Jensen, L.J., Colombelli, J., Bork, P., Stelzer, E.H.K. & 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

François Nédélec

PhD 1998, Université Paris 11, France.
Postdoctoral research at EMBL.
BioMS Group Leader since 2002.
Joint appointment with the Structural and Computational Biology Unit.

Cellular architecture

Previous and current research

Modern microscopy has shown us the dynamic nature of biological organisation. Let's look, for example, at eukaryotic cell division and one of its essential parts: chromosome segregation. This is accomplished by a structure called a mitotic spindle, made of chromosomes, microtubules (polar filaments) and numerous associated proteins. All these elements are connected into a solid structure, which remains highly dynamic: the main components – microtubules – are in rapid turnover. The microtubules grow, shrink and disappear in a matter of minutes, while the mitotic spindles can subsist for hours. In fact, none of the microtubule-associated proteins, such as molecular motors, remain for long, yet their permanent stochastic interactions at the molecular level result in a stable overall structure: a spindle conserves its shape and size, and applies the balanced forces necessary to position and segregate the chromosomes precisely.

The spindle is thus a fascinating structure, which illustrates a central question in biology: how can the uncoordinated and inevitably imperfect actions of proteins and molecules result in a structure able to fulfill its biological function with the utmost accuracy?

Obviously, some kind of averaging is going on, but deciphering how multiple proteins cooperate to produce cellular order is not at all straightforward. It is a challenging problem for at least two reasons: 1) there are many different types of protein implicated; and 2) most of their interactions are dynamic and largely uncharacterised.

Within the field of cytoskeletal morphogenesis, we try to address these two fundamental difficulties in practical terms, by developing *in vitro* experiments and modelling tools. The *in vitro* approach allows us to reduce the number of components in the system; we can either remove a specific protein, or start from scratch by mixing purified components. Modelling allows us to recapitulate the process of protein organisation in a framework in which all the interactions are known exactly and can be specified at will. In practice, we develop innovative numerical methods to simulate the collective behaviour of multiple polar fibres and of their associated proteins. Simulations are often used to validate or refute pre-existing ideas, but they can also be used in a more creative way. We generate systematically various properties for the molecules, and use simulations to test their ability to form stable structures. Successful scenarios are identified automatically, leading to the formulation of new hypotheses, which can later be tested experimentally.

Future projects and goals

We want to study systems in which experiments and theory can be synergistically combined. For example, our current work on the mitotic spindle is to further characterise chromosome-microtubule interactions by experimentation, in order to include chromosomes into the simulations. We are generally interested in any cytoskeletal process observed at the cellular level, such as morphogenesis in *S. pombe*, or the generation of asymmetry in the first division of the *C. elegans* embryo.

Selected references

Kozlowski, C., Srayko, M. & Nédélec, F. (2007). Cortical microtubule contacts position the spindle in *C. elegans* embryos. *Cell*, in press

Janson, M.E., Loughlin, R., Loiodice, I., Fu, C., Brunner, D., Nédélec, F.J. & Tran, P.T. (2007). Crosslinkers and motors organize dynamic microtubules to form stable bipolar arrays in fission yeast. *Cell*, 128, 357-368

Karsenti, E., Nédélec, F. & Surrey, T. (2006). Modelling microtubule patterns. *Nat. Cell Biol.*, 8, 1204-1211

Goshima, G., Nédélec, F. & Vale, R.D. (2005). Mechanisms for focusing mitotic spindle poles by minus end-directed motor proteins. *J. Cell Biol.*, 171, 229-240

Nédélec, F. (2002). Computer simulations reveal motor properties generating stable antiparallel microtubule interactions. *Journal of Cell Biology*, 6, 1005-1015

Rainer Pepperkok

PhD 1992, University Kaiserslautern, Germany.
Postdoctoral research at University of Geneva, Switzerland.
Lab Head at the Imperial Cancer Research Fund, London, UK.
Team Leader at EMBL since 1998.

Membrane traffic in the early secretory pathway

Previous and current research

Transport between two adjacent membranes in the secretory pathway is mediated by coated vesicular carriers and involves at least four basic steps:

1) budding of vesicles from donor membranes;

2) transport towards the target membrane;

3) docking and fusion of vesicles with the target membrane; and

4) recycling of the transport machinery back to the donor membrane.

To warrant both the specificity of delivery of cargo and the maintenance of the unique protein and lipid compositions of the organelles involved, these four steps must be tightly regulated.

Our research focuses on how membrane traffic between the endoplasmic reticulum and the Golgi complex is regulated in space and time. To investigate this, we have developed light microscopy approaches to directly visualise in living cells the kinetics of secretory and organelle markers simultaneously with vesicular coat molecules (COPI and COPII) and their regulators. We also use fluorescence recovery after photobleaching (FRAP) and fluorescence resonance energy transfer measurements (FRET) together with mathematical modelling of the data in order to understand the mechanistics of the temporal and spatial regulation of the molecular interactions involved.

We are also interested in the trafficking of a specific class of protein toxins travelling the secretory pathway in reverse from the plasma membrane to the endoplasmic reticulum. Using this system we could obtain evidence for a novel COPI-independent recycling pathway from the Golgi complex to the endoplasmic reticulum.

In order to identify further specific regulators of the alternative transport pathways and potential regulators of COPI and COPII in microtubule dependent transport, we systematically localise novel human proteins in cells in collaboration with the department of Annemarie Poustka at the DKFZ in Heidelberg. Using the high content screening microscopy platform, that we have developed in the past, we use automated microscopy-based assays that reveal a potential functional implication of the proteins we systematically localise in secretory transport, Golgi morphology or microtubule stability. We also exploit our developments on high content screening microscopy and transfections on siRNA arrays to conduct cell-based 'genome-wide' siRNA screens with the aim of identifying the interaction networks that regulate ER-exit of cargo and its link to subsequent transport along microtubules.

Future projects and goals

We will study the novel proteins, which we reveal in our screens to be involved in the early secretory pathway, in further detail. An important question in this context will be if and how they participate in the temporal and spatial organisation of ER-exit sites and their function, and the biogenesis of the Golgi complex.

Selected references

Forster, R., Weiss, M., Zimmermann, T., Reynaud, E.G., Verissimo, F., Stephens, D.J. & Pepperkok, R. (2006). Secretory cargo regulates the turnover of COPII subunits at single ER exit sites. *Curr. Biol.*, 16, 173-179

Runz, H., Miura, K., Weiss, M. & Pepperkok, R. (2006). Sterols regulate ER-export dynamics of secretory cargo protein ts-O45-G. *EMBO J.*, 25, 2953-2965

Pepperkok, R. & Ellenberg, J. (2006). Innovation – High-throughput fluorescence microscopy for systems biology. *Nat. Rev. Mol. Cell Biol.*, 9, 690-696

Watson, P., Forster, R., Palmer, K.J., Pepperkok, R. & Stephens, D.J. (2005). Coupling of ER exit to microtubules through direct interaction of COPII with dynactin. *Nat Cell Biol.*, 7, 48-55

Ernst H.K. Stelzer

Diploma (Physics) 1983, Max-Planck-Institute for Biophysics, Frankfurt/Main.
PhD (Physics) 1987, University of Heidelberg.
Project leader EMBL Physical Instrumentation Programme 1987-1989.
Group Leader, Physical Instrumentation and Cell Biology Programmes since 1989.
Group Leader, Cell Biology and Biophysics Unit since 1996.

In vitro based optical nanotechnologies for physiological and clinical approaches to a modern biology

Previous and current research

Modern biophotonics provides technologies that operate in a nanodomain. The resolution of optical microscopes is several 100nm, the precision of optical tweezers can be a single nm, laser cutters generate incisions a couple of 100nm wide and cause severing that is barely 700nm deep. More recently, efficient light microscopes require only nano watts of power to induce a reasonable fluorescence emission. However, all technologies are mainly applied in 2D cellular systems, in a cellular context of hard and flat surfaces.

Obtaining relevant information requires the geometry, the mechanical properties, the media flux and the biochemistry of a context found in living tissues. Such a physiological context excludes cells on cover slips. It is found in complex three-dimensional cell structures, which grow on patterned surfaces, in a three-dimensional environment and, naturally, in model embryos or in live tissue.

The observation and optical manipulation of thick and optically dense biological specimens suffer from two problems: a) specimens tend to scatter and absorb and the delivery of the probing light as well as the extraction of the signal light become quite inefficient; b) many biochemical compounds absorb light and suffer from degradation of some sort, inducing malfunction or even death.

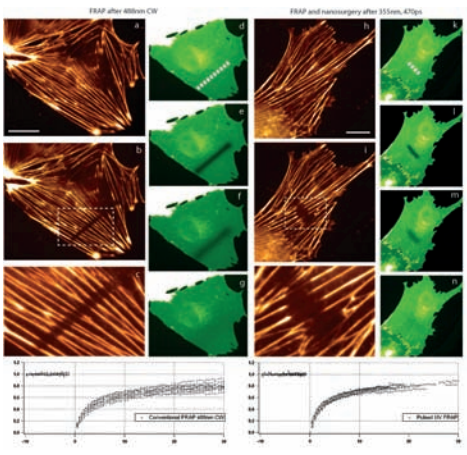
Not by chance, our group developed technologies for the dynamic observation of large and complex 3D biological specimens. They are based on the optical light sheet, which is fed into the specimen sideways and observed at an angle of 90° to the illumination axis. The focal volumes of the detection lens and the light sheet overlap. Optical sectioning and no photo damage outside the common focal plane are intrinsic properties. EMBL's implementations are the single plane illumination microscope (SPIM) and its more refined version (DSLM), taking advantage of modern camera technology and combinable with many contrasts and specimen manipulation tools found in modern microscopes and operating in three dimensions.

Future projects and goals

It is our medium-term goal to integrate the optical nanotechnologies developed during the past years into our novel microscopes and apply them to complex biological objects.

We developed a technological basis that integrates SPIM with perfusion cell culturing units. Time-lapse imaging of cell cultures for several days under controlled medium and temperature conditions are possible and provide model systems for studying organ morphogenesis.

The optical path in SPIM is designed to allow high flexibility and modularity. We successfully integrated our nanoscalpel and devised a toolbox of photonic nanotools. We will investigate the influence of localised mechanical forces on cell function by inducing perturbations in cellular systems. Typical relaxation experiments include cutting Actin fibres and microtubules, optical ablation of cells contacts, manipulation of submicrometer particles and stimulation of selected compartments with optically trapped probes.



Intracellular surgery. High reactivity of cytoskeleton versus membrane. A comparison of Fluorescence Recovery After Photobleaching (FRAP) affecting the membrane of Ptk-2 cells with two different lasers. Regular FRAP with a continuous Argon laser line at 488nm is achieved in a) to g). FRAP with a nanosurgery pulsed laser unit (470ps pulses, 355nm) is achieved in h) to n). Results show that fluorescence recovery after nanosurgery is 15% faster than for the conventional FRAP, demonstrating that diffusion is not prevented by a nanosurgical severing. Scale bars: 20 µm. Prepared by Julien Colombelli.

Selected references

Taxis, C., Maeder, C., Reber, S., Rathfelder, N., Miura, K., Greger, K., Stelzer, E.H. & Knop, M. (2006). Dynamic organization of the actin cytoskeleton during meiosis and spore formation in budding yeast. *Traffic*, 7, 1628-1642

Keller, P.J., Pampaloni, F. & Stelzer, E.H.K. (2006). Life sciences require the third dimension. *Curr. Opin. Cell Biol.*, 18, 117-124

Colombelli, J., Reynaud, E.G., Rietdorf, J., Pepperkok, R. & Stelzer, E.H.K. (2005). In vivo selective cytoskeleton dynamics quantification in interphase cells induced by pulsed ultraviolet laser nanosurgery. *Traffic*, 6, 1093-1102

Huisken, J., Swoger, J., Del Bene, F., Wittbrodt, J. & Stelzer, E.H.K. (2004). Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science*, 305, 1007-1009

Thomas Surrey

PhD 1995, Eberhard-Karls University, Tübingen, Germany.
Postdoctoral research at Princeton University, USA and EMBL.
Staff Scientist 2001-2002.
Team Leader at EMBL since 2002. Group Leader since 2006.

Physical biology of molecular motors involved in intracellular organisation

Previous and current research

Motor proteins are key determinants for the spatial organisation of eukaryotic cells. They are thermodynamic non-equilibrium machines playing a crucial role for the dynamic nature of cellular order. In fact, they provide a paradigm for the concept of intracellular order depending on molecular dynamics. How exactly the collective behaviour of various motors with different kinetic properties drives the organisation of the cytoskeleton is not understood.

Presently, we are following several lines of research to determine how motors contribute to intracellular order. We work on: (1) motile properties of purified motors *in vitro*, single molecule and ensemble measurements; (2) the effect of molecular motors on microtubule dynamics *in vitro*; (3) self-organisation of motors and microtubules *in vitro*; (4) development of novel assays to measure kinetic properties of motors inside cells and in cell extracts; and (5) regulation of motors by kinases *in vitro* and in cell extracts.

The techniques we use include advanced light microscopy, biochemistry, cell biology and molecular biology.

Future projects and goals

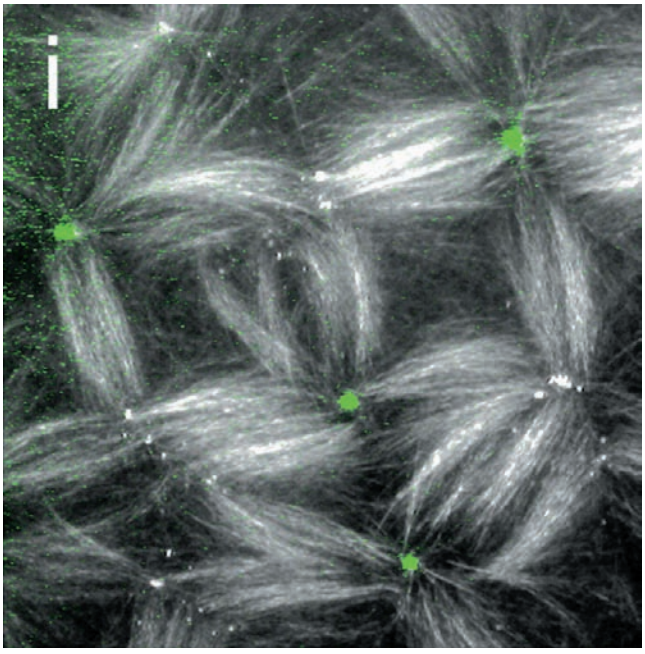
Our goal is to understand how the biochemical and physical properties of microtubule-dependent motors determine their specific activity in a cell. We would like to understand how the biological 'function' of a whole set of motors and microtubules for a given process (such as mitotic spindle assembly) is generated from their coordinated and regulated interactions. We are interested in searching for design principles underlying intracellular dynamics and organisation. Understanding the regulation of motor protein properties will be of crucial importance. Therefore, we will develop tools that will allow us monitor and to manipulate the spatio-temporal regulation of motor protein activities in cells and cell extracts using modern light microscopy techniques.

Selected references

Bringmann, H., Skiniotis, G., Spilker, A., Kandels-Lewis, S., Vernos, I. & Surrey, T. (2004). A kinesin-like motor inhibits microtubule dynamic instability. *Science,* 303, 1519-1522

Seitz, A. & Surrey, T. (2006). Processive movement of single kinesins on crowded microtubules visualized using quantum dots. *EMBO J,.* 25, 267-277

Surrey, T., Nédélec, F., Leibler, S. & Karsenti, E. (2001). Physical properties determining self-organization of motors and microtubules. *Science,* 292, 1167-1171



Network of microtubules and two kinds of motor proteins created by self-organisation in vitro. *Microtubules are visualised by darkfield microscopy; one of the two motor species is visualised by fluorescence microscopy.*

Developmental Biology Unit

Groups in the Developmental Biology (DB) unit uses a variety of organisms to address fundamental problems in animal development. A common theme is the concern with biological processes at the organismal level. Within that context, our interests range from the level of individual cells, e.g. subcellular localisation of RNA, polarised cellular responses to signalling gradients, to the organismal, e.g. control of tissue growth and metabolism.

The molecular basis for cell fate specification and differentiation are core interests of developmental biologists. An emerging part of this picture involves whole genome expression analysis. Various groups in the DB unit are engaged in generating data-sets and databases that permit the use of expression pattern data as a computable resource. Quantitative data from expression profiling is one part of the picture. Another is qualitative assessment of the spatial and temporal patterns of gene expression that are rendered useful for computation by describing them with a controlled, hierarchical annotation vocabulary, similar to GO terms. Computational analysis based on spatial expression data has already proven useful in understanding microRNA function. Expression data from zebrafish, medaka and *Platynereis* are being integrated in a project coordinated in the context of the Computational Biology Centre and with the participation of the Wittbrodt, Furlong, Steinmetz and Arendt labs. An important part of this effort involves collaboration with colleagues in the SCB unit in Heidelberg and at the EBI to develop new tools for analysis of these large data sets. Integration of large-scale expression data with analysis of transcription factor binding through whole genome 'chromatin immunoprecipitation' and with computational tools to predict gene-regulatory elements (developed in a collaborative effort of the Wittbrodt and Birney labs) will be important in developing a system-level view of the control of cell identity in animal development. The Arendt lab has begun to build on this type of thinking to view evolution from a cell-type perspective: suites of genes define a cell type. Considering evolution at the level of cell-type-specific gene expression 'signatures', not only at the level of molecules, is emerging as a powerful new paradigm.

One highlight from the DB unit in 2006 has been the systematic analysis of muscle gene expression carried out by the Furlong group. One of the aims of Developmental Biology is to understand the regulatory cascades – hierarchies of gene expression choices – that control developmental decisions. The Furlong lab has over the past few years built up a complete picture of muscle gene expression by expression profiling normal and mutant animals and extended this to chromatin immunoprecipitation for muscle specific transcription factors. This has yielded a nearly complete picture of which genes are regulated by which transcription factor in the developing muscle (Sandmann *et al.*, 2006, 2007). This data will provide the opportunity to undertake ambitious longer-term systems biology goals, such as modelling transcriptional regulatory units, and more ambitiously, transcriptional regulatory networks controlling morphogenesis and organogenesis.

Several groups in the DB unit have interests in disease mechanism. Recent work in the Treier lab has generated new mouse models for endocrine cancer, premature ovarian failure, polycystic kidney disease and obesity. Expression profiling has begun to provide important insights into the molecular basis of these diseases. Studies in other organisms on control of cell proliferation, epithelial mesenchymal transition, apoptosis and metabolism show promise of leading to the development of new mouse disease models in years to come.

Stephen Cohen
Head of Developmental Biology Unit

Stephen Cohen

PhD 1983, Princeton University, USA. Postdoctoral research at the Whitehead Institute, MIT, Boston, USA, and Max Planck Institute for Developmental Biology, Tübingen, Germany.
Assistant Professor, Howard Hughes Medical Institute, Baylor College of Medicine, Houston, USA.
Group Leader at EMBL since 1993. Head of Unit since 1996.

MicroRNAs, growth and metabolism

Previous and current research

Gradients of secreted signalling proteins organise spatial pattern and control tissue growth during animal development. The signals are produced by small groups of cells at defined positions. Once secreted, these proteins function as morphogens to instruct cells about their fate as a function of the local concentration of the signalling protein, hence they can be said to convey positional information. We have taken molecular and genetic approaches to understanding how these morphogen gradients are established and how they control tissue growth during development.

Growth of tissues and organs during animal development involves careful coordination of the rates of cell proliferation and cell death. The connections between control of cell proliferation and apoptosis in normal development and in cancer are not yet well understood. Among the new genes we've identified are a protein kinase that controls cell survival in response to proliferative cues and a microRNA. microRNAs are a class of short 21-23 nucleotide RNA molecules implicated in the control of gene expression. To date few miRNAs have been assigned functions, but those that are understood regulate post-transcriptional gene expression. With Alex Stark and Rob Russell (page 47), we have developed bioinformatic tools to identify the target mRNAs regulated by miRNAs and find that the average miRNA regulates hundreds of genes. Hundreds of miRNAs have been identified leading us to estimate that miRNAs regulate at least a third of all genes. Combining target prediction with experimental analysis of miRNA expression has begun open the door to a broader understanding of the roles that miRNAs play in evolution and development.

The mechanisms that control growth of cells and tissues during embryonic life are closely linked to control of metabolism. Recent work in the lab has identified a number of genes implicated in control of metabolism as well as tissue growth. Among these are novel modulators of the Insulin/TOR signalling pathway. Ongoing work is aimed toward assessing their functions in mammalian systems.

Future projects and goals

Our goal is to understand the cellular and molecular mechanisms controlling growth and metabolism in animal development.

Selected references

Teleman, A.A. & Cohen, S.M. (2006). *Drosophila* lacking microRNA miR-278 are defective in energy homeostasis. *Genes Dev.*, 20, 417-422

Thompson, B.J. & Cohen, S.M. (2006). The Hippo pathway regulates the bantam microRNA to control cell proliferation and apoptosis in *Drosophila*. *Cell*, 126, 767-774

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

Teleman, A.A., Chen, Y.W. & Cohen, S.M. (2005). 4E-BP functions as a metabolic brake used under stress conditions but not during normal growth. *Genes Dev.*, 19, 1844-1848

Detlev Arendt

PhD 1999, Albert-Ludwigs-Universität, Freiburg, Germany.
Postdoctoral research at EMBL.
Team Leader at EMBL since 2002.
Group Leader since 2006.

Evolution of the central nervous system in Bilateria

Previous and current research

We are intrigued by one of the remaining great mysteries in animal evolution: how did our central nervous system come into existence? What did it look like at first and how did it function? We are especially interested in the CNS of an extinct animal known as Urbilateria, the last common ancestor of humans, flies and most other 'higher' animals that live today. We know that these ancient creatures lived some 600 million years ago in the ocean.

We have therefore chosen to work on a 'living fossil', the marine annelid *Platynereis dumerilii*, that we keep in laboratory culture. This species exhibits many ancient features in its lifestyle, anatomy and development. In bioinformatics comparisons we found that *Platynereis* also shows an ancestral gene inventory and gene structure.

We combine morphological and molecular approaches in a novel evo-devo approach, the molecular comparison of cell types. Animal nervous systems are made up of different sorts of sensory neurons, motor- and interneurons. Each type displays a characteristic 'molecular fingerprint', a unique combination of specifying transcription factors and downstream effector genes such as receptors, transmitters or neuropeptides. The comparison of molecular fingerprints allows the tracing of cell types through animal evolution. For example, in the *Platynereis* brain we have characterised a special type of photoreceptor cell, a 'ciliary photoreceptor' that by molecular fingerprint comparison relates to the rods and cones, the visual photoreceptors of the vertebrate retina. This has led to the fascinating hypothesis that the vertebrate eye has evolved from within the Urbilaterian brain.

Besides ciliary photoreceptors, the *Platynereis* brain harbours several neuron types that have a dual function: they are both sensory and neurosecretory. The ongoing molecular characterisation of these cell types again revealed striking parallels to vertebrate cell types, mostly situated in the hypothalamus. Finally, we have also characterised the molecular architecture of the *Platynereis* trunk central nervous system and discovered striking parallels to the molecular architecture of the vertebrate neural tube. Basically, it appears that the vertebrate neural tube has evolved by the infolding of a pre-existing central nervous system that was in place already in the bilaterian ancestors.

Future projects and goals

It is now clear that our molecular fingerprint comparisons between annelid, vertebrate and insect have the potential to unravel the origin of the bilaterian central nervous system. We are excited by the prospect of further deciphering the evolution of photoreceptor cells and of the diverse eye types that exist in animals. Also, we want to know the evolutionary origin of the most advanced brain part that ever evolved, the telencephalon. We have discovered neurons in *Platynereis* related to telencephalic neuron types by molecular fingerprint, and started to investigate them further.

The clear picture is emerging that the *Platynereis* brain harbours many cell types so far known only for the vertebrates, but in a much more simple, very different overall arrangement. This makes it an attractive goal to elucidate the functioning of these cell types in the ancient marine environment in order to gain insight into the evolutionary origins of the vertebrate brain. To this end we are establishing neurobiological assay systems for larval swimming and for adult learning, and we are heading towards computer modelling of these and of other complex behavioural traits.

Platynereis dumerilii *(Polychaeta, Annelida, Lophotrochozoa)*

Selected references

Purschke, G., Arendt, D., Hausen, H. & Müller, M.C.M. (2006). Photoreceptor cells and eyes in Annelida. *Arthropod Structure & Development*, 35, 211-230

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

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

Arendt, D., Tessmar-Raible, K., Snyman, H., Dorresteijn, A.W. & Wittbrodt, J. (2004). Ciliary photoreceptors with a vertebrate-type opsin in an invertebrate brain. *Science*, 306, 869-871

Anne Ephrussi

PhD 1985, Massachusetts Institute of Technology, USA.
Postdoctoral work at Harvard University and the Whitehead Institute, Cambridge, Massachusetts, USA.
Group Leader at EMBL since 1992.
Coordinator of EICAT since 2005. Dean of Graduate Studies since 2005.

Cell polarity and RNA localisation

Previous and current research

Polarity is a main feature of most eukaryotic cells and underlies many fundamental cell functions and basic developmental processes. Polarisation involves the asymmetric distribution of cytoskeletal structures, organelles, and molecules within the cell.

In *Drosophila*, asymmetrically localised cell fate determinants in the oocyte specify the body axes and basic patterning of the future embryo. The posterior determinant, *oskar*, directs abdomen and germline formation, and is localised as an mRNA at the oocyte posterior pole. As ectopic *oskar* activity causes severe developmental defects, its tight restriction to the posterior is critical and is achieved by mRNA localisation-dependent translation: *oskar* translation is repressed during transport and is activated when the mRNA reaches the posterior pole. Once translated, Oskar protein is maintained at the posterior pole by mechanisms ensuring its tight anchoring at the oocyte cortex.

Assembly of the *oskar* mRNA transport complex begins in the nucleus and involves splicing-dependent deposition of Exon Junction Complex (EJC) proteins on the mRNA, at the first exon-exon junction. Assembled into an mRNP containing the EJC and proteins bound to its 3'UTR, *oskar* mRNA is transported to the posterior pole in large mRNP particles, by a mechanism involving kinesin heavy chain and the polarised oocyte microtubule cytoskeleton. *oskar* mRNA translational repression is achieved by dual mechanisms, one cap-dependent and one cap-independent, that inhibit small ribosomal subunit recruitment. The mechanistically novel, cap-independent mechanism involves formation of large RNP particles ('silencing particles') containing known translational repressors of *oskar*, several of which are also P-body components. It remains unclear how *oskar* mRNA translation is activated once the mRNA reaches the posterior pole.

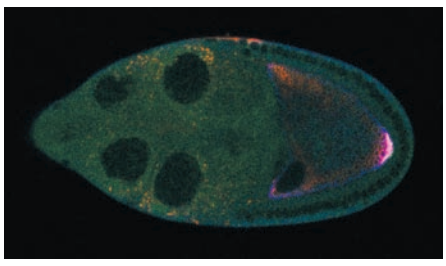
Oocyte polarity is critical for correct *oskar* mRNA localisation and subsequent embryonic development. The *Drosophila* oocyte is ideally suited to genetic, molecular, and cell biological investigation of the processes of cell polarisation, mRNA localisation and translational control. We make use of this model system to study (1) the mechanisms underlying establishment and maintenance of cell polarity; (2) assembly of the *oskar* mRNP complex and the mechanisms of *oskar* mRNA localisation, translational regulation and post-translational control; and (3) Oskar protein anchoring at the oocyte cortex.

Future projects and goals

Using genetics, proteomics, biochemistry, and molecular and cell biological approaches, we are investigating:

- Oskar protein and how it nucleates formation of polar granules, the germline granules of *Drosophila*;
- the mechanisms underlying oocyte determination and polarisation;
- the role of the cytoskeleton and motors in *oskar* mRNA localisation;
- the architecture of the *oskar* localisation mRNP: cis acting RNA elements, the EJC and other interacting proteins, and how they assemble to form a functional localisation mRNP;
- the mechanisms coupling *oskar* mRNA localisation and translational control;
- the relationship between oskar mRNA 'silencing particles' and P-bodies.

Our goal is to understand how oocyte polarity is established and translated into a correctly patterned embryo.



A Drosophila *egg-chamber, showing colocalisation of oskar mRNA, Staufen protein and a microtubule polarity marker at the posterior of the oocyte.*

Selected references

Coutelis, J.-B., & Ephrussi, A. (2007). Rab6 mediates membrane organisation and determinant localization during *Drosophila* oogenesis. *Development*, 134, 1419-1430

Chekulaeva, M., Hentze, M.W. & Ephrussi, A. (2006). Bruno acts as a dual repressor of *oskar* translation, promoting mRNA oligomerization and formation of Silencing Particles. *Cell*, 124, 521-533

Jenny, A., Hachet, O., Závorszky P., Cyrklaff A., Weston M.D.J., St Johnston D. Erdélyi M. and Ephrussi A. A translation-independent role of oskar RNA in early *Drosophila* oogenesis (2006). *Development*, 133, 2827-2833

Vaccari, T., Rabouille, C. and Ephrussi, A. (2005). The *Drosophila* PAR-1 spacer domain is required for lateral membrane association and for polarization of follicular epithelial cells. *Current Biology*, 15, 255-261

Hachet O. and Ephrussi A. (2004). Splicing of *oskar* RNA in the nucleus is coupled to its cytoplasmic localization. *Nature*, 428, 959-963.

Eileen Furlong

PhD 1996, University College Dublin, Ireland.
Postdoctoral research at Stanford University, USA.
Group Leader at EMBL since September 2002. Joint appointment with the Gene Expression Unit.
Coordinator, High-throughput Functional Genomics Centre.

Dissecting the logics of developmental regulatory networks

Previous and current research

One of the central questions in developmental biology is how pluripotential fields of cells become progressively restricted to specific cell fates. Embryonic development occurs through the establishment of complex spatio-temporal patterns of gene expression, which are dynamically modified as development proceeds. These expression states are regulated by the integration of signalling and transcriptional networks converging on cis-regulatory modules (CRMs). Genetics studies have been very successful in identifying key transcription factors (TFs) required for different developmental events. However, due to the complexity of mutant phenotypes and pleotropic effects of many TFs, it is often not understood how these regulators function at a molecular level. For the majority of key developmental regulators only a handful of direct target genes are known. Moreover, the limited number of characterised CRMs has severely hampered our ability to make accurate predictions about the logics of their regulation. Our work aims to bridge this gap by combining genomic, computational and genetic approaches to investigate both the topology and function of developmental networks (Fig. 1). For these studies we are using muscle specification in *Drosophila* as a model system.

Muscle development is a well-established paradigm for transcriptional regulation. A number of essential families of TFs play conserved role in myoblast specification and differentiation from flies to man (Fig. 2), including members of the MyoD, Mef2, Nkx, Tbox and GATA families. We are using these as entry points to seed a global network describing key stages of mesoderm and muscle development. By combining Chromatin ImmunoPrecipitation followed by microarray analysis (ChIP-on-chip) and expression profiling in defined developmental windows we are starting to unravel how these factors drive the myogenic program. This work is ongoing, but some exciting patterns are starting to emerge. Subgroups of TFs selectively co-bind to specific sets of enhancers in a temporally regulated manner. We have also identified a number of new muscle specific TFs, which will further enrich the network. Through the integration of our data we have constructed an initial transcriptional network describing the early stages of mesoderm subdivision (Fig. 1). The overall topology of the network was unexpected and makes some clear predictions of cooperative regulation by different groups of TFs, which we are currently testing using *Drosophila* genetics.

Future projects and goals

A number of experimental (including genetics, genomics, immuno-staining, biochemistry, tissue culture) and computational approaches will be used to address the following:

1. Dissect the regulatory code that provides precise temporal and spatial expression of overlapping subsets of genes during specific stages of myogenesis
2. Identify new functions for these well characterised transcription factors, which may have been masked by the complexity of their mutant phenotypes.
3. Further enrich our network by integrating additional ChIP-on-chip data.
4. Examine functional redundancy and robustness within the network.
5. Use both the topology and dynamics of the network to decipher the logic of how these integrated circuits regulate specific aspects of tissue development.

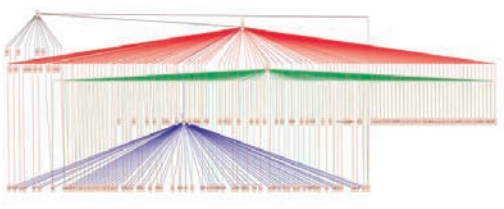


Fig. 1: An initial transcriptional network for early mesoderm development. All connections represent in vivo *co-binding to the same CRM at the same stages of development.*

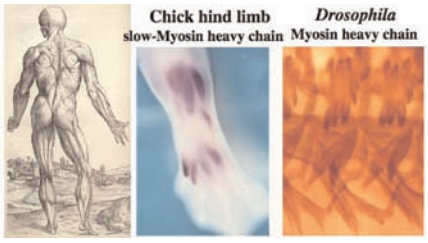


Fig. 2: The genetic cascades governing muscle development are highly conserved from flies to man

Selected references

Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V. & Furlong, E.E. (2007). A core transcriptional network for early mesoderm development in *Drosophila melanogaster*. *Genes Dev.*, 21, 436-49

Girardot, C., Sklyar, O., Grosz, S., Huber, W. & Furlong, E.E. (2007). CoCo: A web application to display, store and curate ChIP-on-chip data integrated with diverse types of gene expression data. *Bioinformatics*, Jan 17 (Epub)

Sandmann, T., Jensen, L.J., Jakobsen, J.S., Karzynski, M.M., Eichenlaub, M.P., Bork, P. & Furlong, E.E. (2006). A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. *Dev. Cell*, 10, 797-807

Spitz, F. & Furlong, E.E. (2006). Genomics and development: Taking developmental biology to new heights. *Dev. Cell*, 11, 451-457

Furlong, E.E. (2005). A functional genomics approach to identify new regulators of Wnt signaling. *Dev. Cell*, 8, 624-626

Furlong, E.E. (2004). Integrating transcriptional and signalling networks during muscle development. *Curr. Opin. Genet. Dev.*, 14, 343-350

Carl Neumann

PhD 1997, EMBL.
Postdoctoral research at the Max Planck Institute for Developmental Biology, Tübingen, Germany.
Group Leader at EMBL since 2001.

The genetic control of vertebrate organogenesis – eye and limb development in zebrafish

Previous and current research

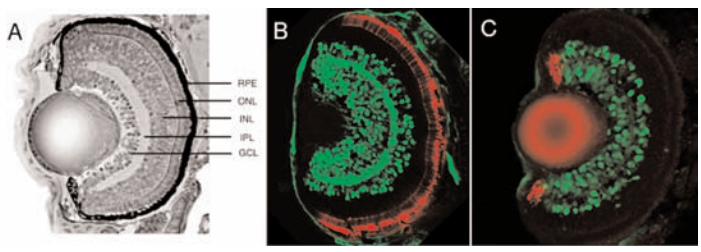
Organs are specialised tissues required for the survival and integrity of multi-cellular organisms. The first step in organogenesis is the specification of a primordium – the establishment of a group of founder cells committed to form an organ. This step is critical, since it determines the location of the organ within the embryo. Secondly, the cells in the organ primordium proliferate to an appropriate degree, and positional information is generated within the developing tissue, leading to distinct cellular identities in different regions. Finally, in order to complete organogenesis, precursor cells must undergo cell cycle exit at the correct time and location, and differentiate as cell types specialised for their context.

We are studying the genetic control of several steps in organogenesis using the zebrafish, *Danio rerio*, and focusing mainly on two organ systems: the paired fins and the eyes. The paired fins of zebrafish are homologous to the limbs of land vertebrates, and just like these, they develop from buds in the lateral plate mesoderm. In the context of fin/limb development, we are interested in the earliest steps of organogenesis, and have analysed a group of genes important for initiation of limb development. These include genes affecting retinoic acid, Wnt and Fgf signalling, as well as transcription factors such as Tbx5 and Prdm1. Our analysis of these genes has clarified their regulatory hierarchy, and has revealed how positional information present in the trunk of the embryo is transferred to the lateral plate mesoderm to trigger limb formation.

The vertebrate retina develops from an evagination of the neural tube. It is initially a non-neuronal, un-patterned epithelium, which develops into a highly organised 3D array of neurons. Differentiation in the retina proceeds in a wave from the centre to the periphery, and is preceded by cell-cycle withdrawal of retinal precursor cells. In the context of the retina, our aim is to understand how cell-cycle exit is coordinated with differentiation, and how cell-cycle exit and cell fate specification are controlled by signalling between cells. We have previously shown that the Sonic hedgehog signalling protein functions as a short-range signal to trigger cell-cycle exit and differentiation in the zebrafish retina, and that this effect of Sonic hedgehog is mediated by transcriptional activation of the cyclin-kinase inhibitor p57Kip2. p57Kip2 inhibits activity of CDKs that drive G1 progression, and thereby promotes cell-cycle exit. These results shed light on the mechanism whereby cellcycle exit is controlled by cell-cell signalling during eye development.

Future projects and goals

In order to systematically characterise the genetic network controlling limb induction, we are using high-thoughput expression profiling experiments on zebrafish mutants affecting early limb development. We are also using bioinformatics approaches to identify transcriptional regulators of known genes important for limb induction. To gain further insight into the mechanisms guiding cell-cycle progression and cell-cycle exit during development, we are analysing the role of several signalling pathways in directing the cell-cycle during organogenesis. Finally, since nearly all aspects of organogenesis are controlled by cell-cell signalling, we are also examining the role of heparan sulphate proteoglycans in regulating signalling proteins in the extracellular space.



Development of structure and cellular diversity in the zebrafish retina. (A) Methylene bluestained sections of a 3 day zebrafish eye. (B,C) Cryosections of 3 day zebrafish embryos carrying a shhGFP transgene immunostained for GFP (in green) and rod-photoreceptors (red) in panel B, or atonal homologue 5 (red) in panel C. Note that shh-GFP expression is present in the neurons of the inner retina, while photoreceptors are present in the outermost layer, and atonal homologue 5 is expressed in regions of the margin where neurogenesis continues in the mature retina. GCL, ganglion cell layer; INL, inner nuclear layer; IPL, inner plexiform layer; ONL, outer nuclear layer; RPE, retinal pigmented epithelium.

Selected references

Mercader, N., Fischer, S. & Neumann, C.J. (2006). Prdm1 acts downstream of a sequential RA, Wnt and Fgf signaling cascade during zebrafish forelimb induction. *Development*, 133, 2805-2815

Norton, W.H., Ledin, J., Grandel, H. & Neumann, C.J. (2005). HSPG synthesis by zebrafish Ext2 and Extl3 is required for Fgf10 signalling during limb development. *Development*, 132, 4963-4973

Shkumatava, A. & Neumann, C.J. (2005). Shh directs cell-cycle exit by activating p57Kip2 in the zebrafish retina. *EMBO Rep.*, 6, 563-569

Shkumatava, A., Fischer, S., Muller, F., Strahle, U. & Neumann, C.J. (2004). Sonic hedgehog, secreted by amacrine cells, acts as a short-range signal to direct differentiation and lamination in the zebrafish retina. *Development*, 131, 3849-3858

Pernille Rørth

PhD 1993, University of Copenhagen, Denmark.
Staff Associate at the Carnegie Institute of Washington, USA.
Group Leader at EMBL since 1998; Senior Scientist since 2003.
Joint appointment with the Cell Biology and Biophysics Unit.

Regulation of cell migration

Previous and current research

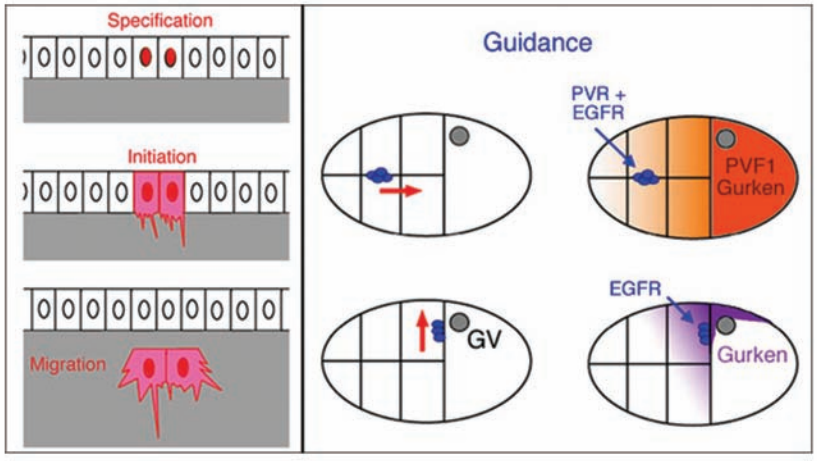
During animal development, many cells migrate from one place to another in order to perform their biological functions. Cell migration is a complex process that involves dynamic interactions between migrating cells and the tissue through which they migrate. In order to migrate, cells change their shape and adhesion properties and become invasive and motile. They must also read guidance cues provided by the target tissue that tell them where to go and when to stop. To analyse these events at the molecular and cellular level, we are studying a specific migration event in *Drosophila*: the migration of border cells in the ovary. We use genetics to identify genes important for cell migration *in vivo* and investigate the molecular action of these genes in order to understand how cell migration is controlled.

We have identified the guidance cues that direct border cell migration and the receptors that interpret this information. The receptors are two receptor tyrosine kinases, PVR and EGFR. We are now investigating the signalling pathway downstream of the receptors responsible for guidance, as well as how this signal is localised properly within the cell and its relationship to cell polarisation. We are also analysing how border cells control their adhesion to, and traction on, the substratum as well as the cellular mechanics of migration. In parallel with this, we are carrying out new genetic screens to identify other critical players in the control of migration.

A number of transcription factors, including Slbo, are specifically required in border cells for initiation of migration. To investigate how cells become migratory and the role of specific transcription factors, we are using DNA-microarrays. We investigate gene expression regulated by Slbo as well as more generally gene expression changes associated with initiation of cell migration. The MAL/SRF transcription factor complex has a different and intriguing role in migrating cells. MAL/SRF target genes and the molecular mechanism of MAL regulation are being explored.

Future projects and goals

Although our emphasis is presently on the migration of border cells, we are also exploring other examples of cell migration during development. As part of our analysis, we have initiated real-time imaging of migrating cells *in vivo* and plan to look at real-time dynamics of key molecules. We are interested in further exploring the relationship between signalling and cytoskeletal forces.



Selected references

Borghese, L., Fletcher, G., Mathieu, J., Atzberger, A., Eades, W.C., Cagan, R.L. & Rørth, P. (2006). Systematic analysis of the transcriptional switch inducing migration of border cells. *Dev. Cell*, 10, 497-508

Jekely, G., Sung, H.H., Luque, C.M. & Rorth, P. (2005). Regulators of endocytosis maintain localized receptor tyrosine kinase signaling in guided migration. *Dev. Cell*, 9, 197-207

Duchek, P., Somogyi, K., Jekely, G., Beccari, S. & Rørth, P. (2001). Guidance of cell migration by the *Drosophila* PDGF/VEGF receptor. *Cell*, 107, 17-26

Duchek, P. & Rørth, P. (2001). Guidance of cell migration by EGF receptor signaling during *Drosophila* oogenesis. *Science*, 291, 131-133

François Spitz

PhD 1997, Institut Cochin de Génétique Moléculaire, Paris.
Postdoctoral research at the University of Geneva.
Group Leader at EMBL since 2006.

Genome architecture, vertebrate development and evolution

Previous and current research

Redeployment or modification of ancestral genes by adding or altering a cis-regulatory element have been fruitful ways to develop evolutionary novelties. This evolutionary 'tinkering' probably accounts for the pleitotropic functions of most 'developmental genes' as well as for the highly intricate and interdigitated distribution of genes and cis-regulatory elements along chromosomes. It is likely that this complex organisation contributed to the preservation of chromosomal syntenies between different animals, since rearrangements occurring within these conserved arrays of genes would disrupt their proper regulation, leading to consequences that are dramatically illustrated by several human genetic disorders.

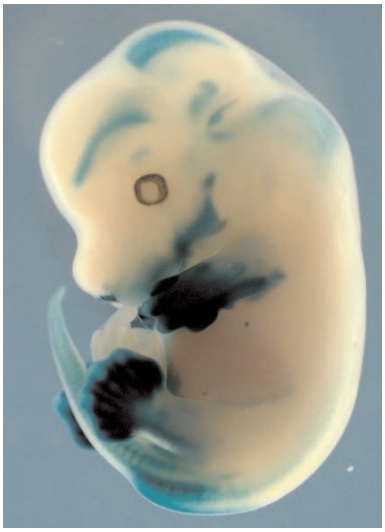
There is therefore a tight link between the architecture of chromosomal loci and the expression/function of the genes lying within. However, we still know very little about the underlying mechanisms, i.e. those which promote recruitment of genes by emerging or distant cis-elements, enable remote enhancers to control their target genes without acting on other genes or, on the contrary, make them act promiscuously to define global regulatory landscapes. We are aiming to explore these issues, focusing on regions involved in developmental processes (either containing 'developmental genes' or where chromosomal rearrangements have been associated to developmental defects). Such approaches should lead not only to a better understanding of the regulatory machinery that control gene expression and shape our chromosomes, but also provide new insights in the genes involved in morphogenesis of the vertebrate embryo and in the mechanisms promoting evolutionary innovations.

We are currently interested in a region where rearrangements such as tandem duplications or insertion of retro-transposons have been linked to severe limb malformations in human and mice respectively. These structural alterations of this otherwise highly conserved locus should disrupt and re-shuffle the complex and global mechanisms that control the expression programs of the genes lying within, leading to a dramatic change in limb morphology. We are aiming 1) to identify the gene(s) involved, so as to get insights into the genetic programme that controls the growth and patterning of our limbs; 2) to understand the molecular consequences of these chromosomal rearrangements so as to tackle the general mechanisms that link genome architecture to gene expression.

Future projects and goals

Using mouse genetics/transgenesis as well as a combination of molecular, biochemical and some computational approaches, we will focus on:

- the reorganisation of the mouse genome through chromosomal engineering strategies to study the resulting consequences (phenotypes, changes in gene expression, chromatin structure, etc.);
- the evolution of cis-regulatory modules around developmental genes;
- the identification of the genes/molecular mechanisms involved in long-range enhancer-promoter interactions.



Selected references

Spitz, F. & Furlong, E.E. (2006). Genomics and development: Taking developmental biology to new heights. *Dev. Cell*, 11, 451-457

Spitz, F., Herkenne, C., Morris, M.A. & Duboule, D. (2005). Inversion-induced disruption of the HoxD cluster leads to the partition of regulatory landscapes. *Nature Genetics* 37, 889-93

Zuniga, A., Michos, O., Spitz, F. *et al.* (2004). Mouse limb deformity mutations disrupt a global control region within the large regulatory landscape required for Gremlin expression. *Genes Dev.*, 18, 1553-64

Spitz, F., Gonzalez, F. & Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell,* 113, 405-417

Mathias Treier

PhD 1994, University of Heidelberg, Germany.
Postdoctoral research at the University of California, San Diego, USA.
Group Leader at EMBL since 2000.
Joint appointment with EMBL Monterotondo.

Mammalian organogenesis and physiology

Previous and current research

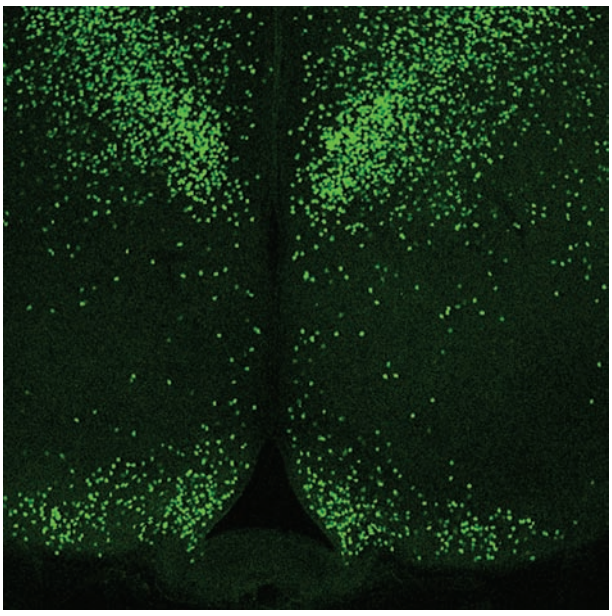
The development of specific cell types during mammalian organogenesis has been studied intensively in the last decade. The challenge now is to understand how these different cell types function in the context of a whole organ, and the ultimate goal is to determine how all the different organs function in a concerted action to create a whole mammalian organism.

We are employing mouse genetics to study several steps of mammalian organogenesis with special focus on the hypothalamic-pituitary axis and the kidney. The hypothalamic-pituitary (HP)-axis serves critical homeostatic functions by regulating key peripheral endocrine organs; the thyroid gland, the adrenal gland and the kidney. In addition, the HP-axis plays a central role in the sexual maturation of the reproductive organs. In particular, we are interested in how the HP-axis regulates mammalian energy homeostasis. Within the DIABESITY FP6 consortium we are studying the development and function of hypothalamic circuitries that are implicated in this process (www.eurodiabesity.org).

Another major research direction in the lab is concerned with the understanding of organ growth control. Hardly anything is known about why an organ stops growing after it has reached a certain size. We are mainly modelling two human diseases in the mouse to gain further insights into this problem. Firstly, we are investigating the molecular mechanisms that underlie the development of Multiple Endocrine Neoplasia type 1 (MEN1) that leads to a general benign overgrowth of endocrine organs *i.e.* pituitary gland and pancreas. Secondly, we are interested in the abnormal growth regulation during the development of Polycystic Kidney Disease (PKD). PKD is the most common genetic, life-threatening disease affecting an estimated 12.5 million people worldwide – regardless of sex, age, race or ethnic origin (www.pkdcure.org).

Future projects and goals

We have established a series of mouse knockout strains for different transcription factors that serve us now as model systems to gain novel insights into the above described problems. Furthermore, we have been able to identify putative stem cell niches in our organs of interest. This will allow us in the future to genetically manipulate the proliferation state of these stem cell pools in the adult organism to devise strategies to delay degenerative processes within these organs.



Genetic labeling of hypothalamic neuronal populations with H2bGFP.

Selected references

Elling, U., Klasen, C., Eisenberger, T., Anlag, K. & Treier, M. (2006). Murine inner cell mass-derived lineages depend on Sall4 function. *Proc. Natl. Acad. Sci. USA*, 103, 16319-16324

Muller, T., Anlag, K., Wildner, H., Britsch, S., Treier, M. & Birchmeier, C. (2005). The bHLH factor Olig3 coordinates the specification of dorsal neurons in the spinal cord. *Genes Dev.*, 19, 733-743

Schmidt, D., Ovitt, C.E., Anlag, K., Fehsenfeld, S., Gredsted, L., Treier, A.C. & Treier, M. (2004). The murine winged-helix transcription factor Foxl2 is required for granulosa cell differentiation and ovary maintenance. *Development*, 131, 933-942

Treier, M., O'Connell, S., Gleiberman, A., Price, J., Szeto, D.P., Burgess, R., Chuang, P.T., McMahon, A.P. & Rosenfeld, M.G. (2001). Hedgehog signaling is required for pituitary gland development. *Development*, 128, 377-386

Jochen Wittbrodt

PhD 1991, University of Munich, Germany.
Postdoctoral research at the Biocenter, University of Basel, Switzerland.
Junior Group Leader at the Max Planck Institute for Biophysical Chemistry, Göttingen, Germany.
Group Leader at EMBL since 1999.

Brain patterning and eye development

Previous and current research

The vertebrate eye emerges from an epithelial Anlage by inductive interactions beginning at late gastrula stages. Under the influence of midline signalling during neurulation, the single retina Anlage is split into two retinal primordia localised as pre-evaginated domains in the lateral wall of the prosencephalon.

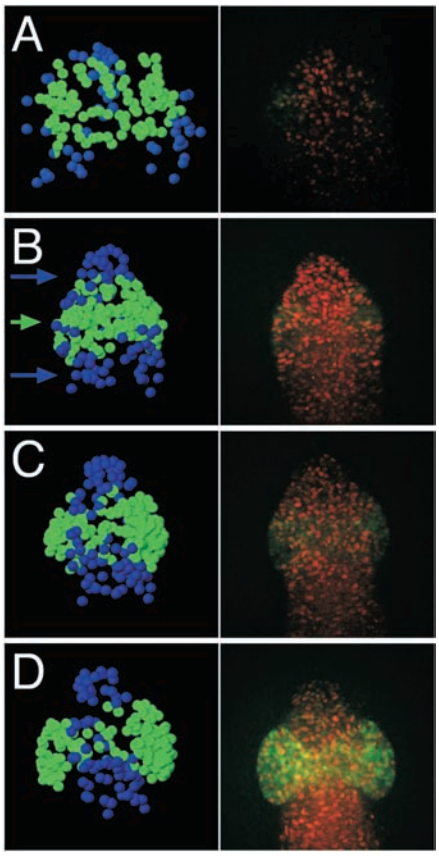
The subsequent evagination of the primordia is driven by the migration of individual cells into the forming optic vesicle. It results in the formation of optic vesicles that differentiate to the seven cell types of the neural retina, the retinal pigmented epithelium and the optic stalk respectively. In anamniotes (fish, amphibia), the ciliary margin of the neural retina contains a stem cell population that gives rise to all retinal cell types and facilitates lifelong growth of the eye.

We investigate vertebrate (medaka, zebrafish) eye development following three complementary experimental strategies: functional studies involving large scale mutagenesis screens and mutant analysis; functional analyses involving gain- and loss-of-function analysis by ectopic expressing of transgenes in the developing eye or by blocking gene function using the morpholino knock-down technique; and *in vivo* analysis of retinal morphogenesis and differentiation using 4D microscopy on transgenic lines stably expressing green fluorescent protein (GFP) in different substructures of the retina.

Future projects and goals

For the study of retinal differentiation, regeneration and retino-tectal projection we incorporate bioinformatics as a novel tool to exploit the resources provided with the publication of the medaka genome. We plan to take advantage of the situation in fish that exhibit life-long growth of the retina and correspondingly, the optic tectum. This requires a close control of the balance between proliferation and differentiation. Addressing and understanding that in the context of fish will shed light onto the situation in amniotes, where retinal stem cells are not found in distinct domains. The maintenance of the topographic retino-tectal projection of a growing retina and tectum requires re-connectivity in the optic tectum. We plan to set up subtle genetic screens to address this question that again is of major bio-medical relevance.

Mutants and other tools established in the lab are currently analysed with the aim of understanding the path that neural progenitors cells take towards terminal differentiation under conditions of growth and regeneration.



Morphogenesis of the optic vesicle. (A) At the neurula stage the eye-field is highlighted by the expression of rx3:GFP (green). Upon formation of the neural keel (B) rx3 positive retinal progenitor cells are modulated in their approach towards the midline, priming the evagination site. (C) Epithelialisation forms a neck posterior to the future optic vesicles and spreads towards posterior. (D) Optic vesicles evaginate due to the mirgration of individual cells that delaminate from central positions and intercalate into the forming optic bud.

Selected references

Rembold, M., Loosli, F., Adams, R.J. & Wittbrodt, J. (2006). Individual cell migration serves as the driving force for optic vesicle evagination. *Science*, 313, 1130-1134

Wichmann, O., Wittbrodt, J. & Schultz, C. (2006). A small-molecule FRET probe to monitor phospholipase A2 activity in cells and organisms. *Angew Chem. Int. Ed. Engl.*, 45, 508-512

Ettwiller, L., Paten, B., Souren, M., Loosli, F., Wittbrodt, J. & Birney, E. (2005). The discovery, positioning and verification of a set of transcription-associated motifs in vertebrates. *Genome Biol.*, 6, R104

Martinez-Morales, J.R., Del Bene, F., Nica, G., Hammerschmidt, M., Bovolenta, P. & Wittbrodt, J. (2005). Differentiation of the vertebrate retina is coordinated by an FGF signaling center. *Dev. Cell*, 8, 565-574

Gene Expression Unit

The 'central dogma' of molecular biology states that 'DNA makes RNA makes protein'. This is a summary of the flow of information from the genetic material (DNA) to the effector molecules (RNA and protein) responsible for the implementation of the genetic blueprint which ultimately gives rise to phenotype. Each step of the pathway of gene expression is complex and, furthermore, subject to regulation. The goals of this Unit are to study the molecular details of the mechanisms of gene expression, its control in eukaryotes and its implications in phenotypic variation. The approaches employed include genetics, biochemistry, proteomics, functional genomics, chemical biology, cell biology and advanced light microscopy. This powerful combination enables the dissection of many of the complex processes on the expression pathway. Within the Unit, different groups study gene expression at different levels.

Genes are packaged into chromatin which in turn is organised into chromosomes, and the template for gene transcription is a complex of DNA with many proteins, most prominently histones. We actively study the composition and structure of chromosomes and of specific regions such as centromeres. In addition, we study how chromatin affects transcription and helps to provide both dynamic and stable patterns of gene expression. A particular focus are chromatin-modifying and -remodelling protein complexes and their role in developmental and sex chromosome specific transcription regulation.

In eukaryotes, many steps of gene expression, such as transcription and RNA processing take place in the structurally complex environment of the nucleus and involve transport of RNA and protein between the nucleus and the cytoplasm through the pore complexes embedded in the nuclear envelope. The principles and mechanisms of assembling these structures are also topics of research in the Unit.

A control point, frequently used in eukaryotic cells, is the cytoplasmic regulation of gene expression. This means regulation either at the level of translation or of messenger RNA stability and both are currently under active study in the Unit.

To understand the genetic principles underlying variations in phenotypes such as complex heritable traits, the Unit develops and employs functional genomics approaches and high throughput methods that query genetic variation and gene expression status systematically and at multiple levels.

Finally, it is important to observe the outcome of gene expression regulation on the protein level. Here the Unit focuses on advanced proteomic approaches and the direct visualisation of major regulatory and signal transduction pathways that feed back on transcriptional control and are relevant to human disease.

In summary, the Unit is equipped for the study of gene expression at several steps of the expression pathway and at different scales of biological organisation. In eukaryotes the expression of many genes is controlled at more than one step. In order to study gene regulation in its entirety, many approaches must therefore be utilised in parallel.

Jan Ellenberg
Head of Gene Expression Unit

Jan Ellenberg

PhD 1998, postdoctoral research at the Cell Biology and Metabolism Branch, NICHD, NIH, Bethesda, USA.
Group Leader at EMBL since 1999.
Head of Unit since 2006.

Functional dynamics of nuclear structure during the cell cycle

Previous and current research

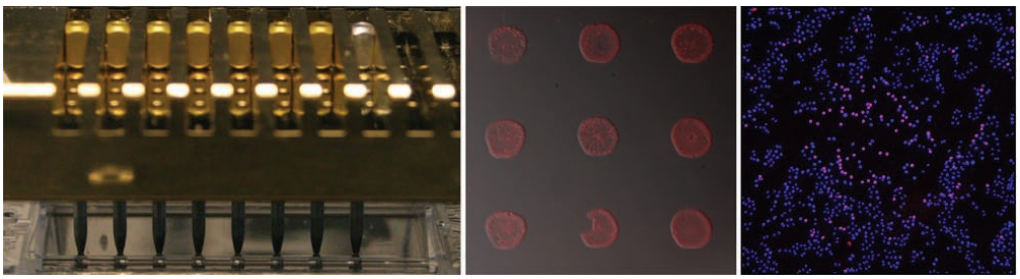
The genome of eukaryotic cells is compartmentalised inside the nucleus, delimited by the nuclear envelope (NE) whose membranes are continuous with the ER and stabilised by the nuclear lamina filament meshwork. The double nuclear membrane is perforated by nuclear pore complexes (NPCs), which allow selective traffic between nucleus and cytoplasm. Attached to the nuclear periphery are heterochromatic regions of chromosomes. These four major structural components of the nucleus are stably connected by many molecular interactions between their resident proteins. In M-phase, most metazoan cells dismantle the highly ordered structure of the NE. Nuclear membranes that surround chromatin in interphase are replaced by cytoplasmic spindle microtubules, which segregate the condensed chromosomes in an 'open' division. After chromosome segregation the nucleus rapidly reassembles.

The overall aim of our research is to elucidate the mechanisms underlying cell cycle remodelling of the nucleus in live cells. Breakdown and reassembly of the nucleus and the formation and correct movement of compact mitotic chromosomes are essential but poorly understood processes. To study them, we are assaying fluorescently tagged nuclear structural proteins and their regulators. 4D confocal microscopy is used to directly observe structural dynamics while laser photobleaching/photoactivation methods serve to analyse binding interactions and diffusion of proteins. Because of the complexity of kinetics and geometry in live cells, we then use computer simulations to extract biophysical parameters and build mechanistic models.

In the past, we showed that NPCs and lamins form a stably interlinked protein network in interphase. We could define the ER as the reservoir and means of partitioning for nuclear membrane proteins in mitosis and found that nuclear breakdown is facilitated by microtubule mediated tearing of the nuclear lamina. During meiotic maturation of starfish oocytes, we demonstrated that partial NPC disassembly is the earliest event of nuclear disassembly and that subsequent long-range chromosome motion is driven by actin. In addition, we have analysed mitotic chromosome dynamics and showed that their overall arrangement is transmitted through mitosis.

Future projects and goals

Objective of our future work is to gain further mechanistic insight into nuclear remodelling in live cells. In particular, we are focusing on the mechanism of nuclear growth in interphase, nuclear disassembly and reformation as well as chromosome condensation and positioning in somatic cells and microtubule-independent chromosome motion in oocytes. To rapidly obtain quantitative data from intact cells, we aim to automate and standardise advanced fluorescence microscopy assays as much as possible. This enables us to apply them in higher throughput to all relevant proteins and achieve a systems level understanding of the transformations in nuclear structure during cell division. Because for many of these processes not all required proteins are known, we are also using high-throughput live cell imaging to identify novel genes that function in the above mitotic processes by RNAi screening.



Production and analysis of siRNA microarrays for microscopy-based cellular assays.

Selected references

Neumann, B., Held, M., Liebel, U., Erfle, H., Rogers, P., Pepperkok, R. & Ellenberg, J. (2006). High-throughput RNAi screening by time-lapse imaging of live human cells. *Nat. Methods*, 3, 385-390

Lenart, P., Bacher, C.P., Daigle, N., Hand, A.R., Eils, R., Terasaki, M. & Ellenberg, J. (2005). A contractile nuclear actin network drives chromosome congression in oocytes. *Nature*, 436, 812-818

Rabut, G., Doye, V. & Ellenberg, J. (2004). Mapping the dynamic organization of the nuclear pore complex inside single living cells. *Nat. Cell Biol.*, 6, 1114-1121

Gerlich, D., Beaudouin, J., Kalbfuss, B., Daigle, N., Eils, R. & Ellenberg, J. (2003). Global chromosome positions are transmitted through mitosis in mammalian cells. *Cell*, 112, 751-764

Asifa Akhtar

PhD 1998, Imperial Cancer Research Fund, London, UK.
Postdoctoral research at EMBL and
Adolf-Butenandt-Institut, Munich, Germany.
Group Leader at EMBL since 2001.

Mechanisms of transcription regulation through chromatin

Previous and current research

DNA tightly packed together with histones into nucleosomes is not easily accessible to the enzymes that use it as a template for transcription or replication. Consequently, remodelling of chromatin structure may play an essential role in the regulation of gene expression. Structural changes in chromatin may also form the basis for dosage compensation mechanisms that have evolved to equalise levels of X-linked gene products between males and females. In humans, one of the two X chromosomes in females is randomly inactivated by condensation of the chromosome into a Barr body, a process known as X-inactivation. In contrast, in *Drosophila* this is achieved by a two fold hyper-transcription of the genes on the male X chromosome. Genetic studies have identified a number of factors that are important for dosage compensation in *Drosophila*, including five proteins (MSL1, MSL2, MSL3, MLE, MOF) and two non-coding RNAs (roX1 and roX2). The hyperactive X is also specifically hyper-acetylated at histone H4, acetylation which is achieved by the MOF histone acetyl transferase.

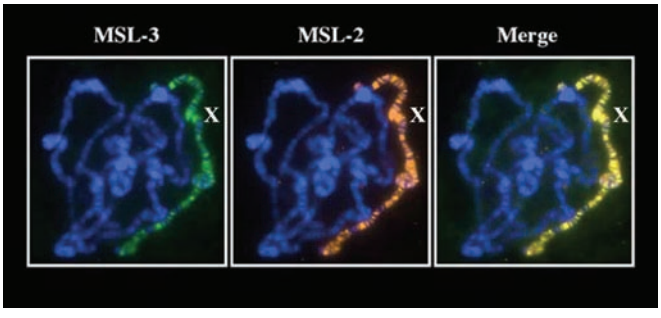
Our major goal is to study the epigenetic mechanisms underlying X-chromosome specific gene regulation using *Drosophila* dosage compensation as a model system. More specifically, we are interested in addressing how the dosage compensation complex, composed of RNA and proteins (the MSL complex), gets targeted to the X chromosome. In addition, we are studying the mechanism by which the MSL complex modulates X chromosomal transcriptional output.

Future projects and goals

The role of nuclear periphery in X chromosomal regulation. We have recently discovered the involvement of nuclear pore components in the regulation of dosage compensation in *Drosophila*. This work has raised several interesting questions about the role of genome organisation and gene regulation, which we will continue to actively address in the future. In addition to using functional genomic approaches, we plan to study in detail the mechanism of nuclear pore/X chromosomal interaction by employing detail cell biology and biochemical chromatin based strategies. This multifaceted approach will be instrumental in future studies to decipher the mechanism of X chromosomal regulation by the MSL complex.

The role of non-coding RNA in dosage compensation. The involvement of non-coding RNAs as potential targeting molecules adds another level of complexity to chromatin regulation. Interestingly, the dosage compensation complex includes two non-coding roX RNAs. However, the mechanism by which these RNAs function is unknown. One of our future aims will be to elucidate how these interactions influence transcription activation of the X-linked genes.

The function of the mammalian MSL complex. There is a remarkable evolutionary conservation of all the known *Drosophila* dosage compensation complex members in mammals. In fact, we have recently purified the *Drosophila* and mammalian MSL complexes and shown that there is a high degree of conservation also at the biochemical level, implying a functional role for the mammalian MSL complex in gene regulation which we will continue to study.



Immunostaining of polytene chromosomes from salivary glands of male Drosophila *using antibodies directed against members of the dosage compensation complex (DCC). The figure shows that MSL-3 and MSL-2 co-localise specifically on hundreds of sites on the male X chromosome. All the chromosomes are also stained with Hoechst to show staining of DNA. The position of the X chromosome is indicated by X.*

Selected references

Buscaino, A., Legube, G. & Akhtar, A. (2006). X-chromosome targeting and dosage compensation are mediated by distinct domains in MSL-3. *EMBO Rep.*, 7, 531-538

Legube, G., McWeeney, S.K., Lercher, M.J. & Akhtar, A. (2006). X-chromosome-wide profiling of MSL-1 distribution and dosage compensation in *Drosophila*. *Genes Dev.*, 20, 871-883

Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Mueller, J., Wilm, M., Stunnenberg, H.G., Saumweber, H. & Akhtar, A. (2006). Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Mol. Cell*, 21, 811-823

Taipale, M., Rea, S., Richter, K., Vilar, A., Lichter, P., Imhof, A. & Akhtar, A. (2005). hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol. Cell Biol.*, 25, 6798-6810

Andreas Ladurner

PhD 2000, Cambridge University, UK.
Postdoctoral research at the Howard Hughes Medical Institute, University of California, Berkeley, USA.
Group Leader at EMBL since 2003.
Joint appointment with the Structural and Computational Biology Unit.

Plasticity and novel chromatin functions in genome regulation

Previous and current research

Chromatin packages our DNA, protects it from damage and regulates gene expression. The smallest unit of chromatin is the nucleosome, a tightly-knit, dynamic and versatile assembly of histones and DNA. Our lab has discovered novel paradigms of gene regulation at the nucleosome level. In particular, we study how post-translational modifications, cell metabolites and chromatin remodelling factors contribute to chromatin plasticity during genome regulation and inheritance.

We focus our attention on two important chromosomal structures, the centromere and the inactive mammalian X chromosome. The first is key to genetic stability, as accurate chromosome segregation is vital. The second provides the molecular basis for gene dosage compensation between males and females (females have two X chromosomes, so they need to shut down gene expression on the second). We address how genes are silenced and how silencing is maintained, determine the role of specialised histones and identify cellular metabolites that regulate chromatin structure.

We use an interdisciplinary approach that includes genetics, biochemistry, cell biology and structural biology to answer fundamental biological questions and to dissect the role of our newly-identified molecular mechanisms in chromatin plasticity. Below is a summary of ongoing research.

All living entities are propagated by cell division, and the proteins that make up centromeric chromatin are essential in this complex and dynamic process. But we do not know how these proteins decide to locate to a specific chromosomal location. On this front, we have successfully used genetic, biochemical and cell biology approaches to isolate a novel heterochromatic protein complex in the fission yeast *S. pombe*. Secondly, we have identified a set of proteins that universally interact with the RNAi-component Argonaute and that are required for RNAi-dependent centromeric silencing. Thirdly, we have investigated the structure of human centromeres using high-resolution human genomic microarrays.

We are also exploring the physiological connections between cellular metabolism and chromatin structure. A human, heterochromatin-enriched histone variant binds NAD metabolites through its macro domain. An alternative splice variant cannot bind the metabolite, and structural data explains this high level of discrimination. Intriguingly, there appears to be a link between these histone isoforms and cell proliferation, suggesting that NAD metabolism may be involved in regulating gene expression. We complement these studies in mammalian cells with a systems-approach of the poly-ADP-ribose pathway in the fruit fly *Drosophila*.

Future project and goals

- Regulation of cellular plasticity, chromatin structure and mammalian gene expression by NAD metabolites.
- Determination of the molecular structure and function of eukaryotic centromeres.
- Role of the RNAi pathway in transcriptional gene silencing at centromeres.

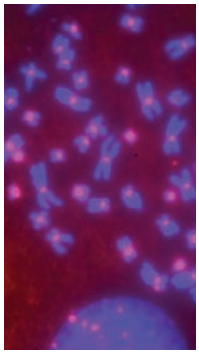


Fig. 1: Accurate genome inheritance relies on the formation of a protein complex on centromeric chromatin. We are studying this process using a highly interdisciplinary approach.

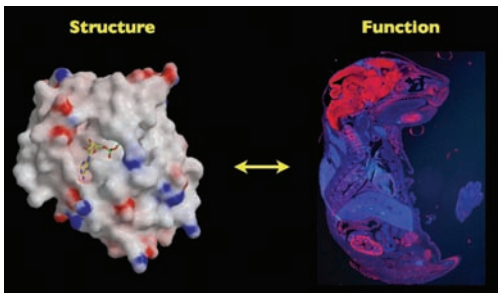


Fig. 2: We have pioneered the discovery of metabolites binding to human chromatin (left) and seek to understand the physiological role of these interactions during gene regulation.

Selected references

Ladurner, A.G. (2006). Rheostat control of gene expression by metabolites. *Mol. Cell*, 24, 1-11

Karras, G.I., Kustatscher, G., Buhecha, H.R., Allen, M.D., Pugieux, C., Sait, F., Bycroft, M. & Ladurner, A.G. (2005). The macro domain is an ADP-ribose binding module. *EMBO J.*, 24, 1911-1920

Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K. & Ladurner, A.G. (2005). Splicing regulates NAD metabolite binding to histone macroH2A. *Nat. Struct. Mol. Biol.*, 12, 624-625

Ladurner, A.G. (2003). Inactivating chromosomes: a macro domain that minimizes transcription. *Mol. Cell*, 12, 1-3

Jürg Müller

PhD 1991, University of Zürich, Switzerland.
Postdoctoral research at the MRC Laboratory of Molecular Biology, Cambridge, UK.
Junior Group Leader at the Max-Planck-Institute for Developmental Biology in Tübingen, Germany.
Team Leader at EMBL since 2001.

Mechanisms of transcriptional regulation in development

Previous and current research

Our laboratory studies the molecular mechanisms by which trithorax group (trxG) and Polycomb group (PcG) proteins maintain transcriptional ON and OFF states of target genes. PcG and trxG proteins are conserved in both animals and plants and are essential for diverse developmentally regulated processes in these organisms. For example, the PcG/trxG system maintains expression patterns of developmental control genes in animals, is required for X-chromosome inactivation and maintenance of stem cell pluripotency in mammals and is needed for the control of seed development and flowering time in plants.

We use *Drosophila* as model system and our aim is to understand the molecular mechanisms by which PcG and trxG regulators maintain expression patterns of HOX and other genes during development. PcG and trxG proteins control gene expression at the level of chromatin in ways that are still poorly understood. To date, we purified several different PcG protein complexes from *Drosophila*. A central part of our work is to characterise these complexes *in vitro*, using biochemical and biophysical assays. Our studies revealed that PcG protein complexes possess particular chromatin-modifying and chromatin-binding activities. This *in vitro* work is complemented by approaches that allow us to study the role of PcG and trxG complex activities *in vivo*. In particular, we analyse the binding of PcG and trxG protein complexes to target genes in *Drosophila* and we monitor how the binding of these complexes affects chromatin modifications at these target genes. Structure/function analyses of PcG and trxG proteins in *Drosophila* provide additional insight into the role of their chromatin-modifying and -binding activities. The combination of these different approaches has provided insight into the mechanisms by which PcG protein complexes are targeted to the genes they regulate and into the role of their chromatin-modifying activities. In addition, analyses of the genome-wide binding of PcG proteins in *Drosophila* and validation that the identified target genes are indeed subject to PcG control, has provided a more comprehensive view of the genes and processes that are controlled by the different PcG complexes. Analysis of these newly identified target genes will allow us identify the basic principles of the mechanism by which the PcG/trxG system regulates transcription of target genes.

Future projects and goals

The regulation of HOX and other target genes in *Drosophila* will continue to provide the framework to gain mechanistic insights into PcG and trxG protein function *in vivo*. Biochemical purification and forward genetic screening strategies to identify novel components of the PcG/trxG system, combined with in-depth *in vitro* and *in vivo* analyses of newly identified proteins are central to this part of our work. A second focus will be to analyse the interaction of PcG protein(s) complexes with recombinant chromatin of defined states using biophysical methods.

Our long-term goal is to understand how transcriptional ON and OFF states controlled by the PcG/trxG system are propagated through replication and cell division.

Selected references

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

Müller, J. & Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr. Opin. Genet. Dev.*, 16, 476-484

Papp, B. & Müller, J. (2006). Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.*, 20, 2041-2054

Nekrasov, M., Wild, B. & Muller, J. (2005). Nucleosome binding and histone methyltransferase activity of *Drosophila* PRC2. *EMBO Rep.*, 6, 348-353

Carsten Schultz

PhD 1989, University of Bremen, Germany.
Postdoctoral research at the University of California, San Diego, USA.
Habilitation 1997, Organic Chemistry, University of Bremen, Germany.
Group Leader, MPI for Molecular Physiology, Dortmund, Germany. Group Leader at EMBL since 2001.

Bio-organic chemistry of signalling molecules

Previous and current research

In the past, our research focused on finding novel ways to stimulate chloride and water secretion of epithelial cells due to their crucial role in the genetic disease cystic fibrosis (CF). Our compounds helped to investigate some of the underlying intracellular signalling pathways and provided drug candidates that effectively increased chloride secretion and reduced water uptake of nasal epithelia from CF patients *ex vivo*. We have developed chemical methods to convert highly polar signalling molecules like cyclic nucleotides, inositol phosphates and phosphoinositides to membrane-permeant, bioactivatable derivatives. These compounds ('prodrugs' or 'Trojan horse compounds') help to deliver polar compounds to the cytosol. With the resulting tools we were able to elevate the concentration of a signalling molecule of interest inside living cells without disrupting the plasma membrane. This technology was successfully extended to drug candidates based on these signalling molecules. As a member of the Molecular Medicine Partnership Unit (MMPU) of EMBL and the University of Heidelberg, we are joining forces with Marcus Mall at the Medical School to test these compounds in CF mouse models.

A central goal is to visualise the signalling pathways involved in chloride secretion by fluorescent probes. By doing so, we hope to provide a more complete picture of the signalling network and to create novel ways to screen compounds that might be beneficial for CF patients. The function of the probes is based on fluorescent resonance energy transfer (FRET) or translocation and is suitable for ratio imaging and confocal microscopy. They are currently combined in an approach called Multiparameter Imaging, where 5-6 cellular events are monitored simultaneously. This work is funded by the Molecular Imaging IP in FP6 and is a joint effort with Dorus Gadella (Amsterdam).

Small molecule fluorescent FRET probes are prepared to study intracellular enzyme activities with a focus on phospholipases (Wichmann *et al.*, 2006). Recently we prepared a very effective probe to monitor phospholipase A2 activity in cells and small organisms. Studies on the development of fish embryos were performed in collaboration with Jochen Wittbrodt (page 28).

To examine the interaction of phospholipids with lipid-binding proteins we are preparing fluorescently labelled phospholipids that serve as FRET partners for fluorescent fusion proteins. The labelled lipids are membrane-permeant and allow lipid translocation events to be studied. Furthermore, the effect of phospholipids on vesicle trafficking is investigated in collaboration with the group of Rainer Pepperkok (page 16). In a similar approach we are investigating the formation of enzyme-substrate complexes alongside the group of Philippe Bastiaens (visiting).

A third approach is based on conformational changes of double fluorescently labelled proteins that lead to changes in FRET. With these reporter probes we monitor several phosphorylation and aggregation events in living cells (Brumbaugh *et al.*, 2006; Piljic & Schultz, 2006). These projects are performed in collaboration with the groups of Michael Sattler (visiting), Frank Gannon (page 53) and Mathias Gautel (London).

Finally, we are interested in switchable nucleotide molecules and novel ways to introduce them to the cytosol. In collaboration with Roland Krämer of the Institute for Inorganic Chemistry at the University of Heidelberg, PNA molecules were prepared that carried a chelating group targeted at transition metals. Upon binding of zinc ions an increase in permeability could be observed thereby delivering the interfering oligonucleotide mimic into cells (Füssl *et al.*, 2006).

Most projects rely on organic chemistry to produce tools or parts of the tools described above. The symbiosis of chemistry, biochemistry, and cell biology opens new doors and grants novel insights into how cells are working.

Selected references

Wichmann, O., Wittbrodt, J. & Schultz, C. (2006). A small molecule FRET probe to monitor PLA2 activity in cells and organisms. *Angew. Chem. Int. Ed. Engl.*, 45, 508-512

Brumbaugh, J., Schleifenbaum, A., Gasch, A., Sattler, M. & Schultz, C. (2006). A dual parameter FRET probe for measuring PKC and PKA activity in living cells. *J. Am. Chem. Soc.*, 128, 24-25

Füssl, A., Schleifenbaum, A., Goeritz, M., Ridell, A., Schultz, C. & Kraemer, R. (2006). Zinc-enhanced cellular uptake of PNA-terpyridine conjugates. *J. Am. Chem. Soc.*, 128, 5986-5987

Piljic, A. & Schultz, C. (2006). Annexin A4 self-association modulates general membrane protein mobility in living cells. *Mol. Biol. Cell*, 17, 3318-3328

Lars M. Steinmetz

PhD 2001, Stanford University, USA.
Postdoctoral research at the Stanford Genome Technology Center, USA.
Group Leader at EMBL since 2003.

Functional genomics of complex traits and pathways

Previous and current research

Most heritable traits are genetically complex, involving the interaction of several genes with the environment. Little is known about the genetic factors underlying complex traits in humans and in other organisms, yet most natural phenotypes and diseases belong to this class. We use and develop functional genomic approaches and high-throughput methods in yeast to identify the genetic variants that underlie complex traits and pathways, and apply our findings to humans.

Using oligonucleotide microarrays we can detect and genotype single nucleotide polymorphisms at high-density over the entire yeast genome. We have used this technology to map at high-resolution the complex, quantitative trait loci (QTL) for virulence traits of high-temperature resistant pathogenic *Saccharomyces cerevisiae*. We have developed a new technology termed Reciprocal Hemizygosity Analysis (RHA) that allows the contribution to phenotype to be determined for all alleles from the genomes of two independent strains. Using our newly designed oligonucleotide tiling array for the yeast genome, which has probes densely tiling both strands of the complete genome, we have undertaken an unbiased investigation of the transcriptional activity across it. We find complex transcription architectures, including hundreds of novel transcripts and are actively pursuing study of the functional significance of this transcription. These technologies will allow us to explore genetic contributions to quantitative traits under a variety of environmental conditions and to investigate ecology at the genome level.

We have a strong interest in mitochondrial biology because of the high conservation from yeast to humans, the complexity of mitochondrial functions, and the large range of human diseases implicated. We have used the recently completed collection of yeast single gene deletion strains to characterise gene function on a genomic scale. In each strain, one gene is deleted and replaced with a cassette containing a kanamycin resistance gene for selection purposes and two molecular barcode tags. The tags allow strain detection with a high-density oligonucleotide microarray containing sequences complementary to each tag, which enables the entire collection of deletions strains to be grown as a pool in a single culture under different environmental conditions. We have used this collection to identify genes involved in mitochondrial function. By identifying human orthologs, we determined new candidate genes for human putative-mitochondrial disorders. We are expanding our understanding of the proteins involved in mitochondria using a combination of computational and functional genomic methods in yeast, including whole organelle proteome and transcriptome analyses. By integrating this data with the deletion data and other genomic datasets, including computational prediction of mitochondrial proteins, we have generated a combined list of mitochondrial proteins. We have taken a network interaction approach to study this list of protein in detail and to place all proteins into a functional context. These approaches allow the study of the mitochondrial organelle at a systems level and speed the discovery of disease-related genes in humans.

Future projects and goals

We are interested in continuing to explore ways to unravel the genetic complexity of quantitative traits in yeast and to bridge the gap between model systems biology and medicine. In particular, we will focus on applying our yeast mitochondrial studies to human proteins and testing mitochondrial candidate genes in patients. We will also continue to explore ways to apply our functional genomic dataset to fundamental questions in evolutionary biology. We have begun studies to dissect the genetic basis of sensitivity and resistance to malaria parasites in the mosquito, *Anopheles gambiae*, with a goal to apply the high-throughput methods developed in yeast to this medically relevant system.

Selected references

Perocchi, F., Jensen, L.J., Gagneur, J., Ahting, U., von Mering, C., Bork, P., Prokisch, H. & Steinmetz, L.M. (2006). Assessing systems properties of yeast mitochondria through an interaction map of the organelle. *PLoS Genet.*, 2, e170

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W. & Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. USA*, 103, 5320-5325

Steinmetz, L.M., Scharfe, C., Deutschbauer, A.M., Mokranjac, D., Herman, Z.S., Jones, T., Chu, A.M., Giaever, G., Prokisch, H., Oefner, P.J. & Davis, R.W. (2002). Systematic screen for human disease genes in yeast. *Nat. Genet.*, 31, 400-404

Steinmetz, L.M., Sinha, H., Richards, D.R., Spiegelman, J.I., Oefner, P.J., McCusker, J.H. & Davis, R.W. (2002). Dissecting the architecture of a quantitative trait locus in yeast. *Nature*, 416, 326-330

Matthias Wilm

MD 1991, University of Münster, Germany.
PhD 1996, Physics, EMBL and the University of Münster, Germany.
Postdoctoral research at the University in Düsseldorf, Germany and at EMBL.
Group Leader at EMBL since 1998.

Quantitative proteomics

Previous and current research

Every field of research offers areas with an high degree of importance. Visionary research is characterised by seeing this importance early and recognising that its targets can be met. To this end, we were at the origins of functional proteomics and systems biology when we supported the 'Tandem Affinity Purification' (TAP) method with all our resources. TAP, in conjunction with high throughput mass spectrometric protein characterisation techniques, opened the way to place all the proteins whose existence were revealed by the human genome project into a functional context. Before the TAP project, functional assignment of all the new proteins appeared to be an insurmountable task. Only five years later we see the proteome as a complex network of interconnected protein complexes. This project was so rich in content that it served as a building ground for the possibly most capable proteomics company of the world, CellZome (www.cellzome.de). We continued to support functional proteomics projects at EMBL and improved the TAP technique for use in higher eukaryotes (the iTAP project).

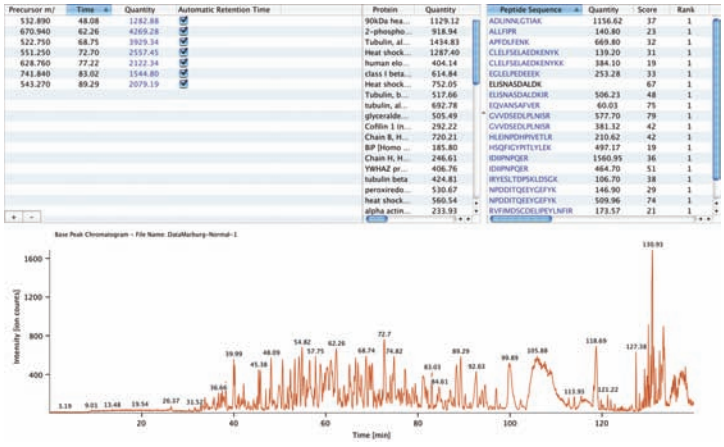
Our current research is focused on turning quantitative proteomics into reality, as quantitative analysis is crucial for research in systems biology. With the increasing sensitivity of mass spectrometric protein sequencing, functionally relevant proteins can only be recognised by seeing their quantitative change in comparison to an experimental control. Systems biology can not be thought of without a quantitative description of the development of a biological system over time. In protein based research, mass spectrometric techniques will be the optimum choice to generate these experimental data. We focused our activity on extracting quantities directly from our electrospray tandem MS investigations as proposed by Waters in 2004 (see www.waters.com/WatersDivision/Contentd.asp?watersit=PSTD-5ZBAUY).

For this purpose, in 2004 we began to develop a Macintosh application, Arcade, that will be able to handle large amounts of data. We chose to base it on CoreData libraries published by Apple in 2004 which support SQLite data formats. This format makes it relatively straightforward to write multithreaded code to take full advantage of the multi-core processor chip architecture and to handle large volumes of data in an interactive application. The current state of the development is illustrated in the figure below. We can now extract quantitative information from every ion that was ever measured on our electrospray mass spectrometers.

Future projects and goals

The next steps in the development towards quantitative proteomics include the ability to take up several chromatograms within one file, standardisation and, ultimately, application in research projects.

As well as being a program to extract quantitative information, Arcade is also a platform for experimental data analysis in proteomics – to evaluate different experimental approaches in research to display and understand biological contexts. To reach its full potential, all the Application Programming Interfaces (APIs) have to be documented so that other people can build on them.



Current status of the Arcade program. The lower panel shows the mass spectrometry chromatogram, the upper right panel all the proteins and their peptides that could be identified using the peptide fragment spectra within the chromatographic run. For nearly all peptides quantitative information had been extracted from the data set. The upper left panel shows individual precursor masses whose quantities were extracted from the data set.

Selected references

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, 368-379

Koffa, M.D., Casanova, C.M., Santarella, R., Kocher, T., Wilm, M. & Mattaj, I.W. (2006). HURP is part of a ran-dependent complex involved in spindle formation. *Curr. Biol.*, 16, 743-754

Mendjan, S., Taipale, M., Kind, J., Holz, H., Gebhardt, P., Schelder, M., Vermeulen, M., Buscaino, A., Duncan, K., Müller, J., Wilm, M., Stunnenberg, H.G., Saumweber, H. & Akhtar, A. (2006). Nuclear pore components are involved in the transcriptional regulation of dosage compensation in *Drosophila*. *Mol. Cell*, 21, 811-823

Niggeweg, R., Kocher, T., Gentzel, M., Buscaino, A., Taipale, M., Akhtar, A. & Wilm, M. (2006). A general precursor ion-like scanning mode on quadrupole-TOF instruments compatible with chromatographic separation. *Proteomics*, 6, 41-53

Structural and Computational Biology Unit

Currently, the Structural and Computational Biology Unit consists of twelve research groups with a wide range of methodological experience. It covers Electron Microscopy (three groups), X-ray crystallography (two groups, one team), NMR (one group), Chemical Biology (one group) and Computational Biology (two groups, two teams). In addition, two groups based in different units have shared appointments with the SCB unit (the Ladurner group from Gene Expression (page 32) and the Nédélec team from Cell Biology (page 15)).

The unit is very well equipped for experimental and computational work. Experimental facilities include image plate detectors for the collection of X-ray diffraction data, a 500 MHz and 600 MHz NMR spectrometer, transmission electron microscopes and scanning microdensitometers; there are also facilities for electron cryo-microscopy, cryo-3D tomography, automated crystallisation experiments, CD and fluorescence spectroscopy and analytical ultracentrifugation, as well as for large scale growth of prokaryotes and eukaryotes. The whole computing environment of central clusters and separate workstations is conveniently networked.

There is a continuing interplay between the different groups with expertise in different methodologies, reflecting the belief that a combination of structural and functional studies is the most rewarding route to an understanding of the molecular basis of biological function, and that computational biology is essential to integrate the variety of tools and heterogeneous data.

A particular strength of the unit is its ability to tackle problems at different ranges of resolution connecting atomic structures and dynamic information obtained by X-ray crystallography and NMR with medium range resolution from single particle electron microscopy and cellular imaging obtained by EM tomography and light microscopy. Biochemistry, chemical biology and computational biology complement the structural biology activities and integrate them into a comprehensive description of biological function.

In this way, the groups in the Structural and Computational Biology Unit pursue a few common large projects that require the input of the different skill sets. Examples are the structural determination or modelling of a large number of protein complexes in yeast (in the context of a large EU grant) and obtaining the structure of an entire cell at almost molecular resolution by mapping protein complexes to tomograms of an entire cell. First proof of principle results have recently been obtained.

Peer Bork and Christoph Müller
Heads of Structural and Computational Biology Unit

Peer Bork

PhD 1990, University of Leipzig, Germany.
Habilitation 1995, Humboldt University, Berlin, Germany.
At EMBL since 1991.
Joint Head of Unit since 2001.

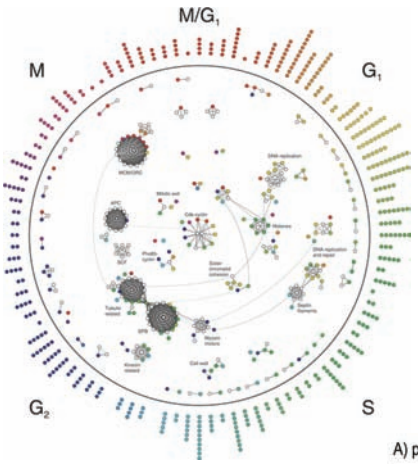
Deciphering function and evolution of biological systems

Previous and current research

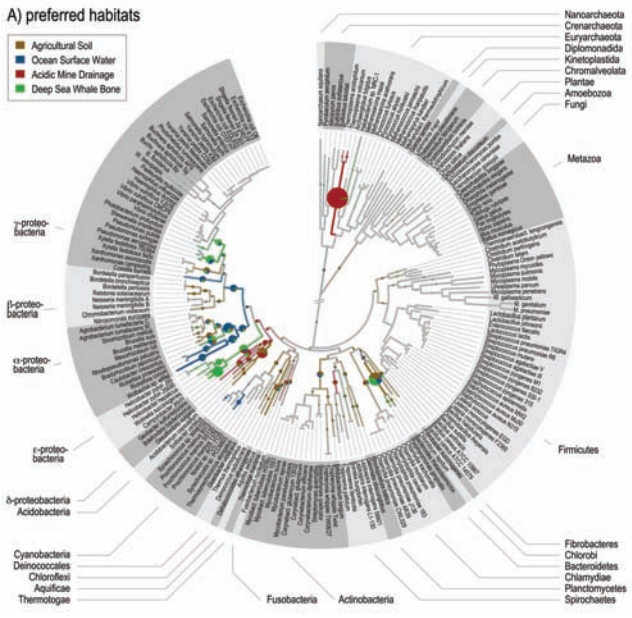
The main focus of this Computational Biology Group is to predict function and to gain insights into evolution by comparative analysis of complex molecular data. The group currently works on three different scales:

- genes and proteins;
- protein networks and cellular processes;
- phenotypes and environments.

They require both tool development and applications. Some selected projects include the analysis of small molecule-protein interactions in the context of networks, the study of temporal and spatial protein network aspects and comparative metagenomics of environments. All are geared towards the bridging of genotype and phenotype through a better understanding of molecular and cellular processes. The group is partially associated with the Max Delbrück Center for Molecular Medicine (MDC) in Berlin.



Left: Temporary interaction networks and dynamic complex formation during yeast cell cycle. 600 cell cycle regulated proteins in yeast (shaded) dots as identified from microarray data interact with noncyclic scaffolding proteins (white). The temporal cell cycle regulation can evolve quickly (see Jensen et al.*, 2006,* Nature, *443 and refs therein).*



Right: Quantitative phylogenetic assessment of microbial communities in four environmental (metagenomic) samples by mapping marker genes onto the tree of life (see von Mering et al.*, 2007,* Science *and references therein).*

Selected references

Jensen, L.J., Jensen, T.S., de Lichtenberg, U., Brunak, S. & Bork, P. (2007). Co-evolution of transcriptional and post-translational cell-cycle regulation. *Nature*, 443, 594-7

von Mering, C., Hugenholtz, P., Raes, J., Tringe, S.G., Doerks, T., Jensen, L.J., Ward, N. & Bork, P. (2007). Quantitative phylogenetic assessment of microbial communities in diverse environments. *Science*, 315, 1126-30

Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B. & Bork, P. (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science*, 311, 1283-1287

Jensen, L.J., Saric, J. & Bork, P. (2006). Literature mining for the biologist: from information retrieval to biological discovery. *Nat. Rev. Genet.*, 7, 119-129

Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., Bork, P., Hugenholtz, P. & Rubin, E.M. (2005). Comparative metagenomics of microbial communities. *Science*, 308, 554-557

Christoph W. Müller

PhD 1991, University of Freiburg, Germany.
Postdoctoral work at Harvard University, Cambridge, USA.
At EMBL Grenoble since 1995.
Joint Head of Unit at EMBL Heidelberg since 2007. Joint appointment with the Gene Expression Unit.

Structural biology of transcriptional regulation

Previous and current research

Our group is interested in the molecular mechanisms of transcriptional regulation. We use structural information mostly obtained by X-ray crystallography but also electron microscopy combined with other biophysical and biochemical techniques to gain insight into these complicated cellular processes.

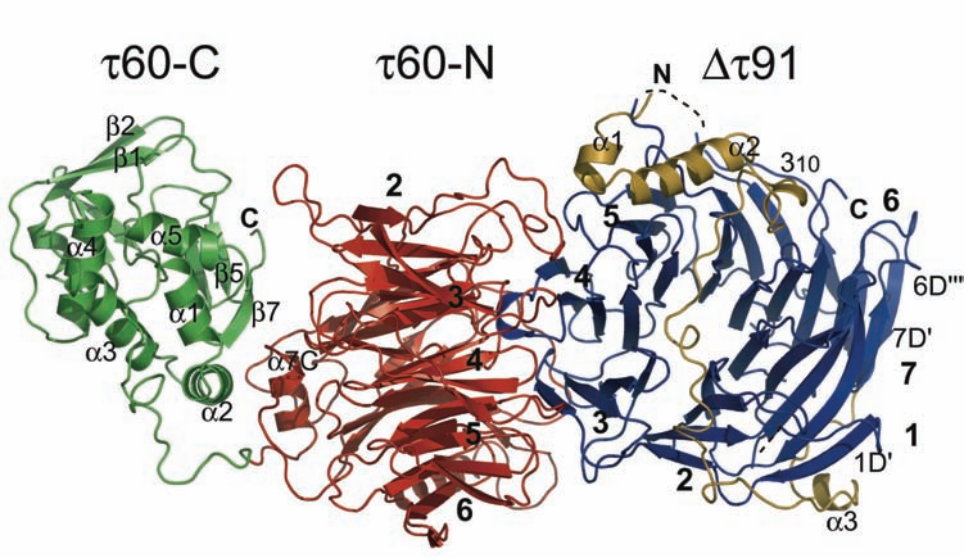
In the last years our research has been increasingly moving from the analysis of individual transcription factors and their interactions with DNA target sites towards the structural and functional analysis of multi-protein complexes involved in eukaryotic transcriptional regulation. Systems currently under investigation include yeast RNA polymerase III, the *Drosophila* nucleosome remodelling complex CHRAC and other multi-protein complexes involved in chromatin targeting and histone modifications.

RNA polymerase III consists of 17 subunits and is responsible for the transcription of small RNAs like tRNA and 5S RNA. Recruitment of the enzyme requires binding of the general transcription factor TFIIIC, composed of six subunits, to internal promoter sites followed by the binding of TFIIIB composed of three-subunits. Our research aims to understand the overall architecture and dynamics of RNA polymerase III, TFIIIC and TFIIIB and how they interact during transcription initiation, elongation and termination.

The accessibility of chromatin in eukaryotes is regulated by ATP-dependent chromatin remodelling factors and histone-modifying enzymes. Both classes of enzymes use similar domains like bromodomains, chromodomains, PHD fingers and SANT domains for the controlled access to defined genomic regions. Many of these enzymes also do not work in isolation but rather in the context of larger multi-subunit assemblies. We try to understand the molecular architecture of chromatin modifying complexes, how they interact with chromatin templates and how their activities are regulated.

Future projects and goals

We want to use structural information on multi-protein complexes and protein-DNA complexes involved in eukaryotic transcriptional regulation to better understand their functions.



Crystal structure of the τ60/Δτ91 subcomplex of yeast general transcription factor IIIC.

Selected references

Fernández-Tornero, C., Böttcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R.W.H., Sentenac, A., Müller, C.W. & Schoehn, G. (2007). Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. *Mol. Cell*, 25, 813-823

Mylona, A., Fernandez-Tornero, C., Legrand, P., Haupt, M., Sentenac, A., Acker, J. & Muller, C.W. (2006). Structure of the tau60/Delta tau91 subcomplex of yeast transcription factor IIIC: insights into preinitiation complex assembly. *Mol Cell*, 24, 221-32

Petosa, C., Morand, P., Baudin, F., Moulin, M., Artero, J.B. & Muller, C.W. (2006). Structural basis of lytic cycle activation by the Epstein-Barr virus ZEBRA protein. *Mol. Cell*, 21, 565-572

Petosa, C., Schoehn, G., Askjaer, P., Bauer, U., Moulin, M., Steuerwald, U., Soler-Lopez, M., Baudin, F., Mattaj, I.W. & Muller, C.W. (2004). Architecture of CRM1/Exportin1 suggests how cooperativity is achieved during formation of a nuclear export complex. *Mol. Cell*, 16, 761-775

Bettina Böttcher

PhD 1991, University of Stuttgart, Germany.
Postdoctoral research at the MRC Laboratory of Molecular Biology, Cambridge, UK.
Assistant Professor at the Albert-Ludwigs-Universität, Freiburg, Germany.
Group Leader at EMBL since 1999.

Electron microscopy and image reconstruction of biological macromolecules

Previous and current research

Electron microscopy and image processing is one of the most powerful tools in the structural investigation of large biological complexes. Insights into the structural organisation of the complexes can be gained without the necessity of crystallising the complex.

We exploit these methods to investigate a variety of different complexes within 3D-Repertoire (EU NOE). These investigations are aimed at developing a comprehensive picture of the stable complexes within a yeast cell and provide detailed structural information which can be used as a spatial scaffold on which protein protein interactions can be modelled. Many of the yeast complexes have very low abundance, which make structural investigations extremely challenging and leave electron microscopy often as the only option for obtaining structural information.

Alternative to the purification of the native complexes, we build complexes from over-expressed protein either by co-expressing subunits or mixing them after expression. These *in vitro* complexes can usually be purified in large quantities, enabling crystallography as well as detailed characterisation of interaction by different biophysical methods providing exact knowledge on shape, stoichiometry, secondary structure and binding constants. Besides providing a basis for structural modelling this information is also vital for simulating protein-protein interactions in Systems Biology.

Another fascinating aspect of our work is focused on the investigation of structure and dynamics of virus capsids by electron microscopy and image processing. These capsids are often regarded as rigid containers for a genome. However, this perception does not at all describe the versatile role of a capsid. Our research is aimed on demonstrating the structural dynamics of virus capsids by observing capsids in different states of their life cycle or actively challenging the capsid structure by introducing limited conformational stress. So far, we have observed a variety of conformational response mechanisms, which provide structural insight into underlying signalling mechanisms.

Future projects and goals

Our general goal is to understand the structural organisation of protein complexes and their interplay within the cell. Therefore the projects are either aimed at obtaining as detailed structural information of a single complex as possible or at elucidating the interaction between complexes in a native environment and their impact on cell morphology. Here our focus is mainly on membrane bound complexes, which often form super complexes that have a significant impact on membrane morphology.

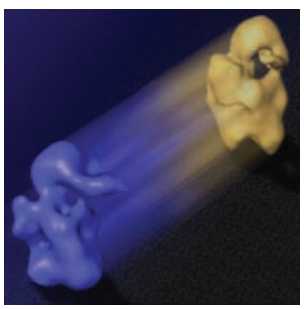


Fig.1: Electron microscopy is a superb method for structure determination of dynamic complexes. Here, the pre-40S ribosome (yellow) and the mature 40S-ribosome (blue) are shown (Schäfer et al.*, 2006).*

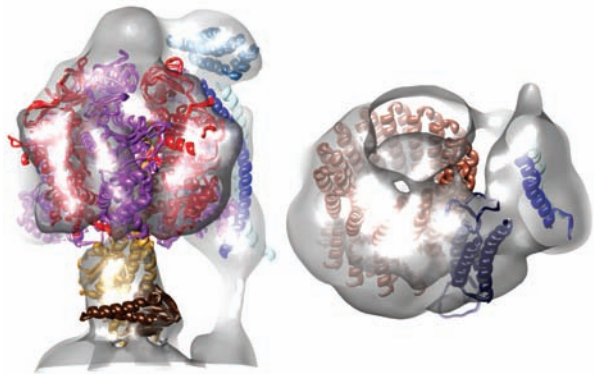


Fig. 2: Image reconstructions derived from electron micrographs in combination with atomic models of sub-complexes can be combined to pseudo atomic models of large complexes. As an example the soluble (left) and the membrane part (right) of the ATPsynthase from chloroplasts are shown (Mellwig & Böttcher, 2003).

Selected references

Böttcher, B., Vogel, M., Ploss, M. & Nassal, M. (2006). High plasticity of the hepatitis B virus capsid revealed by conformational stress. *J. Mol. Biol.*, 356, 812-822

Schafer, T., Maco, B., Petfalski, E., Tollervey, D., Bottcher, B., Aebi, U. & Hurt, E. (2006). Hrr25-dependent phosphorylation state regulates organization of the pre-40S subunit. *Nature*, 441, 651-655

Venzke, D., Domgall, I., Kocher, T., Fethiere, J., Fischer, S. & Böttcher, B. (2005). Elucidation of the stator organization in the V-ATPase of *Neurospora crassa*. *J. Mol. Biol.*, 349, 659-669

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

John Briggs

PhD, Oxford University, UK.
Postdoctoral research at the
University of Munich, Germany.
Group Leader at EMBL since 2006.

How proteins manipulate membranes – cryo-electron microscopy and tomography

Previous and current research

A cell's control over the shape and dynamics of its membrane systems is fundamental to its function. We are interested in how proteins can define and manipulate the shapes of membranes. To explore this question we are studying a range of different cellular and viral specimens using cryo-electron microscopy and tomography.

Cryo-electron microscopy techniques are particularly appropriate to this area of research because they allow membrane topology to be observed in the native state, while maintaining information about the structure and arrangement of associated proteins. Computational image processing and three-dimensional reconstructions are used to extract and interpret this information. We take a step-by-step approach to understanding the native structure. Three dimensional reconstructions can be obtained using cellular cryo-electron tomography of the biological system in its native state. These reconstructions can be better interpreted by comparison with data collected from in-vitro reconstituted systems. A detailed view is obtained by fitting these reconstructions with high resolution structures obtained from images of purified complexes.

A particular emphasis of our research is the structure and life-cycle of asymmetric membrane viruses such as HIV. The assembly of the virus particles and their subsequent fusion with target cells offer insights into general features of vesicle budding and membrane fusion.

Future projects and goals

Our goal is to understand the interplay between protein assemblies and membrane shape. How do proteins induce the distortion of cellular membranes into vesicles of different dimensions? What are the similarities and differences between the variety of cellular and viral budding events? What is the role and arrangement of the cytoskeleton during membrane distortions. What membrane topologies are involved in fusion of vesicles with target membranes? How does the curvature of a membrane influence its interaction with particular proteins? We will develop and apply microscopy and image processing approaches to such questions.

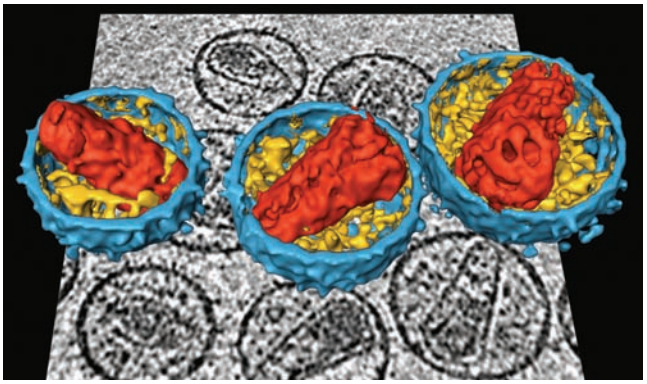


Fig. 1: 3D reconstruction of HIV-1 virions using cryo-electron microscopy.

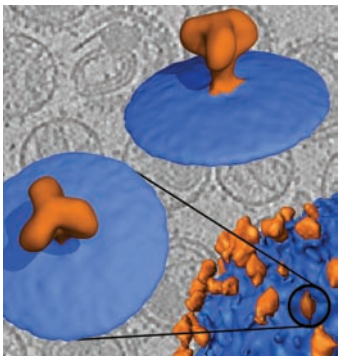


Fig. 2: 3D reconstruction of the SIV glycoprotein spike, generated by averaging sub-tomograms extracted from whole virus tomograms. (Zanetti et al., 2006)

Selected references

Briggs, J.A., Johnson, M.C., Simon, M.N., Fuller, S.D. & Vogt, V.M. (2006). Cryo-electron microscopy reveals conserved and divergent features of gag packing in immature particles of *Rous sarcoma* virus and human immunodeficiency virus. *J. Mol. Biol.,* 355, 157-68

Briggs, J.A., Grunewald, K., Glass, B., Forster, F., Krausslich, H.G. & Fuller, S.D. (2006). The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions. *Structure*, 14, 15-20

Zanetti, G., Briggs, J.A., Grunewald, K., Sattentau, Q.J. & Fuller, S.D. (2006). Cryo-electron tomographic structure of an immunodeficiency virus envelope complex in situ. *PLoS Pathog.*, 2, e83

Elena Conti

PhD 1996, Imperial College, London, UK.
Postdoctoral research at the Rockefeller University, New York, USA.
Group Leader at EMBL since 1999.
Joint appointment with the Gene Expression Unit.

Structural studies of nucleo-cytoplasmic transport and mRNA metabolism

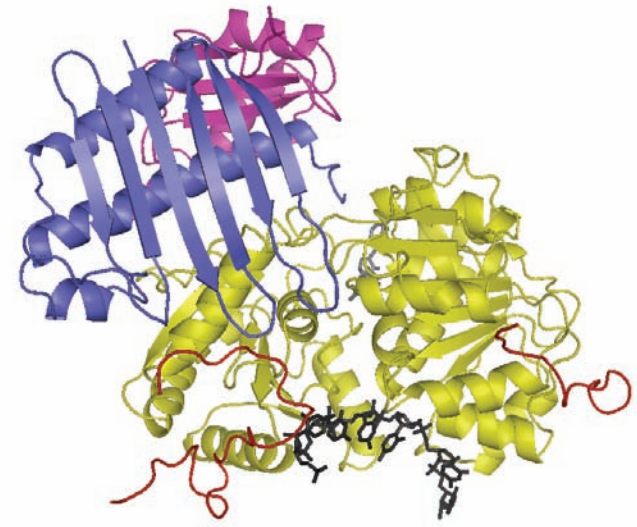
Previous and current research

Our laboratory is interested in the molecular mechanisms that govern the transport of nuclear proteins and RNAs from their site of synthesis to their site of function. In eukaryotes this process allows communication to occur between cytosolic translation and nuclear replication and transcription. We are also interested in the connections of nuclear transport to upstream and downstream processes. In the case of nuclear export, for example, there is increasing evidence that proteins involved in mRNA transport are linked to mRNA processing and surveillance. To obtain molecular insights into these cellular processes, we use X-ray crystallography in combination with other biophysical and biochemical methods.

Nucleo-cytoplasmic transport relies on the recognition of targeting signals by a set of transport factors. The bulk of nuclear import/export is mediated by the karyopherin family of proteins (for example Cse1). These proteins can carry both proteins and RNA cargoes such as tRNAs or pre-miRNAs. The transport of mRNA is instead mediated by a different system unrelated to the karyopherins. Nuclear export of mRNAs is connected to a complex series of events ensuring that only correct and properly spliced mRNAs are translated into proteins. The exon-exon junction complex (EJC) plays a crucial role in this context. In humans, the EJC is deposited on mRNA upon splicing and functions together with SMG proteins to mediate nonsense-mediated mRNA decay (NMD). NMD is a surveillance mechanism that recognises mRNAs with premature stop codons and targets them for rapid degradation, thus avoiding the synthesis of truncated and potentially harmful protein products. Degradation of nonsense transcripts is mediated by the same cellular machinery that participates in general mRNA turnover. In particular, degradation of mRNAs in the 3' to 5' direction is catalysed by the exosome, a protein complex that is also involved in the maturation of structural RNAs in the nucleus.

Future projects and goals

Our goal is to understand the mechanisms by which macromolecular complexes in the nucleus and the cytosol interact with each other, ensuring the continuous flow of information that is essential for eukaryotic life. In particular, we will study the mRNA decay machinery, both in terms of premature stop codon recognition and targeting (i.e. the exon junction complex and the SMG proteins) and in terms of RNA degradation (i.e. the exosome complex). We also plan to continue our studies on export mediators, in particular on the transport factors that export RNAs, and more generally on the connection of nuclear transport to mitotic entry.



Structure of the exon junction complex, an assembly of four proteins that forms in the presence of ATP and RNA, and that regulates many aspects of mRNA fate in the cytoplasm.

Selected references

Bono, F., Ebert, J., Lorentzen, E. & Conti, E. (2006). The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. *Cell*, 126, 713-725

Glavan, F., Behm-Ansmant, I., Izaurralde, E. & Conti E. (2006). Structures of the PIN domains of SMG6 and SMG5 reveal a nuclease within the mRNA surveillance complex. *EMBO J.*, 25, 5117-5125

Jinek, M., Chen, Y.W., Clausen, H., Cohen, S.M., & Conti, E. (2006). Structural insights into the Notch-modifying glycosyltransferase Fringe. *Nat Struct Mol Biol.*, 13, 945-946

Lorentzen, E. & Conti, E. (2006). The exosome and the proteasome: nano-compartments for degradation. *Cell*, 125, 651-654

Achilleas S. Frangakis

PhD 2001, Max Planck Institute of Biochemistry, Martinsried, Germany.
Postdoctoral research at Caltech, Pasadena, California, USA.
Group Leader at EMBL since 2003.

Electron tomography

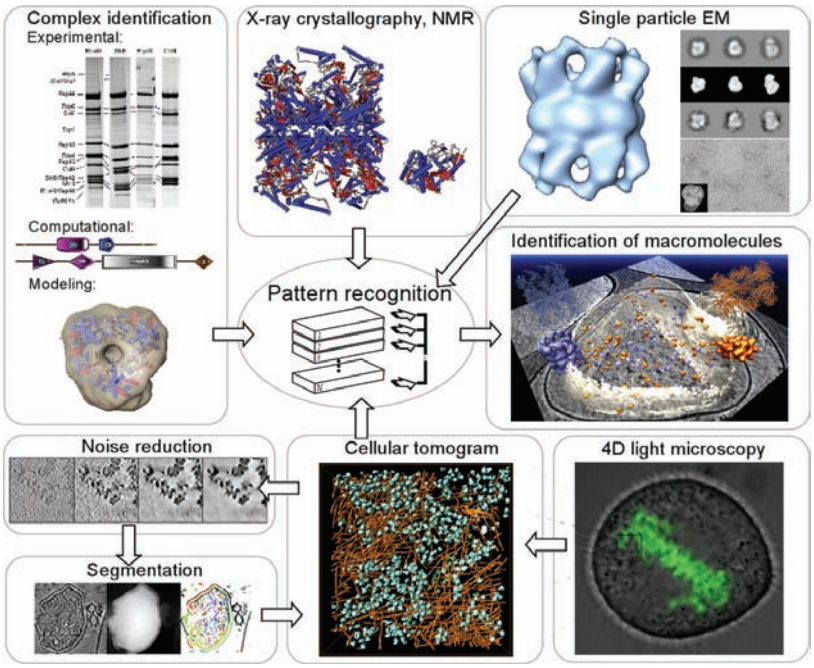
Previous and current research

The main focus of the group is to reveal the macromolecular organisation of living cells by means of cryo-electron tomography. Cryo-electron tomography is the only technique that can obtain molecular resolution images of intact cells in a quasi-native environment. The tomograms contain an imposing amount of information; they are essentially a three-dimensional map of the cellular proteome and depict the whole network of macromolecular interactions. Information mining algorithms exploit structural data from various techniques, identify distinct macromolecules and computationally fit atomic resolution structures in the cellular tomograms, thereby bridging the resolution gap.

A multitude of biological questions can be answered by electron tomography; visualisation of the cellular structure at molecular resolution is largely uncharted territory. The group works with a wide spectrum of specimens, including prokaryotic and eukaryotic cells, but also model systems. Prokaryotic cells are smaller and can therefore be easily penetrated by the electrons. Eukaryotic cells have compartments in which the protein density is lower, facilitating pattern recognition techniques. Model systems, on the other hand, are particularly helpful in improving computational algorithms and in providing solutions for cell systems that are too complex to be investigated by electron tomography.

Future projects and goals

Achieving these goals will enable us to visualise macromolecules in an unperturbed cellular environment and to chart the network of interactions underlying cellular functions. This aim of the group is to prove that a cell is not an envelope of freely diffusing enzymes and substrates, but rather a highly organised and coordinated machine.



The scheme visualises the ultimate goal of the group and the true power of cryo-electron tomography. We are using cryo-electron tomography in conjunction with pattern recognition techniques in order to match atomically resolved structures in the context of living cells. Practically, we integrate the information from X-ray crystallography, structural genomics and single-particle electron microscopy in order to computationally search for macromolecular complexes in the three-dimensional cryo-electron tomograms.

Selected references

Kurner, J., Frangakis, A.S. & Baumeister, W. (2005). Cryo-electron tomography reveals the cytoskeletal structure of Spiroplasma melliferum. *Science*, 307, 436-438

Frangakis, A.S. & Forster, F. (2004). Computational exploration of structural information from cryo-electron tomograms. *Curr. Opin. Struct. Biol.*, 14, 325-331

Frangakis, A.S., Bohm, J., Forster, F., Nickell, S., Nicastro, D., Typke, D., Hegerl, R. & Baumeister, W. (2002). Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. *Proc. Natl. Acad. Sci. USA*, 99, 14153-14158

Bohm, J., Frangakis, A.S., Hegerl, R., Nickell, S., Typke, D. & Baumeister, W. (2000). Toward detecting and identifying macromolecules in a cellular context: template matching applied to electron tomograms. *Proc. Natl. Acad. Sci. USA*, 97, 14245-14250

Anne-Claude Gavin

PhD 1992, University of Geneva, Switzerland.
Postdoctoral research at EMBL.
Director, Molecular and Cell Biology, Cellzome AG, Heidelberg.
Group Leader at EMBL since 2005.

Biochemical and chemical approaches to biomolecular networks

Previous and current research

How is biological matter organised? Can the protein and chemical worlds be matched to understand the cell's inner works? As our knowledge on the basic building blocks of eukaryotic genomes, proteomes and metabolomes grows, the challenge resides in understanding how these parts relate to each other. At cellular levels, gene products very rarely act alone. The orchestration of complex biological functions is the result of networks of molecules. Traditional approaches have typically focused on a few, selected gene products and their interactions in a particular physiological context. We are proponents and pioneers of more general strategies aiming at understanding complex biological systems. We are interested in understanding the principles that govern the assembly of these networks and follow three main lines of research.

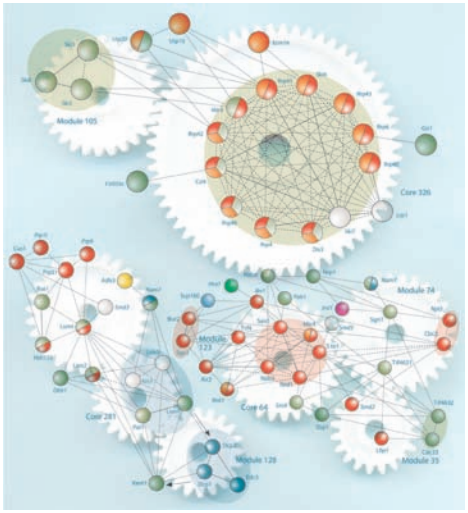
The charting of protein-protein interaction networks: Our knowledge on protein-protein interaction is still anecdotic; current estimations reveal that probably less than 10% have been characterised so far. We adopted the Tandem-Affinity purification/Mass Spectrometry (TAP/MS) technology to perform a genome-wide analysis of protein complexes in the yeast *S. cerevisiae*. More than 400 different protein complexes, more than half entirely novel, were characterised. The approach was particularly successful in further extensive collaborations within the program. These collaborations aimed at the structural characterisation of protein complexes through integration of electron microscopy data and *in silico* approximations.

The study of protein complexes and network order of assembly and dynamics: generally, the use of protein interaction networks to predict the behaviour of whole systems has been relatively limited. Protein networks usually fail to capture the dynamic aspect of protein interactions that is essential for the whole cell functioning. The charting and modelling of the highly dynamic assembly and reorganisation of protein complexes following cell perturbation represents one of the major current research interests of the group.

The extension of interaction networks from proteins to other cell's building blocks; metabolites-on-proteomes networks: Metabolites account for about half of the cell's volume and represent important class of biomolecules. They have long been considered simple building blocks for the assembly of more complex macromolecules. It is however becoming evident that the interactions between the metabolites' and the proteins' worlds are not limited to substrate/product relationships. Metabolites can have well known signalling functions and many proteins are allosterically modulated by metabolites. These bindings are sometimes mediated by a variety of specialised domains. Every time it has been possible to chart such interactions they turned out to have profound functional implications. The interactions taking place between the cell's chemical world and proteomes are still poorly defined and have certainly not yet been studied in a comprehensive way. It represents the second major research interest of the group.

Future projects and goals

- Analysis of the order of assembly and dynamic nature of yeast protein complexes, in a pathway oriented approach.
- Further development and improvement of existing chemical biology methods, based on affinity purification ('metabolite pull-down') to monitor protein-metabolites interaction.
- Global screen aiming at the systematic charting of the interactions between the proteome and the metabolome in *Saccharomyces cerevisiae*.
- Develop new and existing collaborations with computational and structural biology groups at EMBL and outside to tackle the structural and functional aspects of biomolecular recognition.



Graphic: Petra Riedinger

Selected references

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B. *et al.* (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440, 631-636

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

Gavin, A.C. & Superti-Furga, G. (2003). Protein complexes and proteome organization from yeast to man. *Curr. Opin. Chem. Biol.*, 7, 21-27

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., Remor, M., Hofert, C. *et al.* (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 415, 141-147

Toby Gibson

PhD 1984, Cambridge University, UK.
Postdoctoral research at the Laboratory of Molecular Biology, Cambridge, UK.
Team Leader at EMBL since 1986.

Biological sequence analysis

Previous and current research

The group seeks to gain insight through the computational analysis of biological molecules, particularly at the protein sequence level. To this end, we deploy many sequence analysis methods and look to develop new tools as the need arises. Where possible, we contribute to multidisciplinary projects involving structural and experimental groups at EMBL.

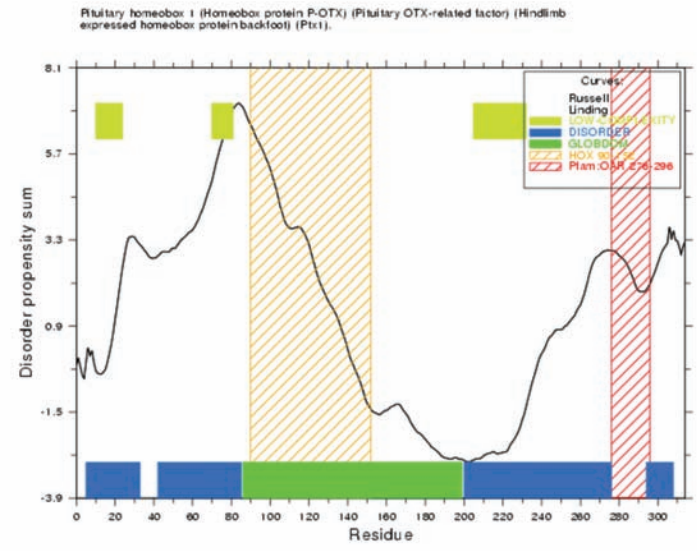
We are probably best known for our involvement with the Clustal W and Clustal X programs that are widely used for multiple sequence alignment. We work closely with Julie Thompson (Strasbourg) and Des Higgins (Dublin) to maintain and develop these programs. We also maintain several public web servers at EMBL, including ELM, the protein linear motif resource; Phospho.ELM, a collection of >13,000 reported phosphorylation sites; GlobPlot, a tool for exploring protein disorder; and Gene2EST, a BLAST server specialised for mapping ESTs to gene sequences.

A major focus recently has been to develop and deploy tools for protein architecture analysis. Our group coordinated the EU-funded ELM consortium that developed the Eukaryotic Linear Motif resource (http://elm.eu.org/) to help users find functional sites in modular protein sequences. Short functional sites are used for the dynamic regulation of large cellular protein complexes and their characterisation is essential for understanding cell signalling. Such sites are most often in IUP (intrinsically unstructured protein segments) and we have developed tools such as GlobPlot to locate these segments as a prerequisite of ELM motif hunting.

Future projects and goals

We apply computers in molecular biology in the hope of gaining new biological insight that may inform experimental strategies. For example, we have proposed new vaccine targets after a combined phylogenetic analysis and proteome survey revealed that bacteria have acquired α2-macroglobulin genes found in metazoa. We will continue to survey individual gene families in depth and will undertake proteome surveys when we have specific questions to answer. Molecular evolution is one of the group's interests, especially when it has practical applications.

With our collaborators, we will look to build up the protein architecture tools, especially the unique ELM resource, taking them to a new level of power and applicability. We will apply the tools in the investigation of modular protein function and may deploy them in proteome analysis pipelines. Our links to experimental and structural groups should ensure that bioinformatics results feed into experimental analyses of signalling interactions and descriptions of the structures of modular proteins and their complexes, with one focus being regulatory chromatin proteins.



GlobPlot of the Ptx1 transcription factor. Ascending slope indicates disorder preference and descending slope a preference for order. The known Homeobox domain lies within the assigned globular preference while the remaining sequence is unstructured and is expected to contain multiple regulatory motifs interacting with other chromatin proteins.

Selected references

Taussig, M.J., Stoevesandt, O., Borrebaeck, C.A., Bradbury, A.R., Cahill, D., Cambillau, C., de Daruvar, A., Dubel, S., Eichler, J., Frank, R., Gibson, T.J., Gloriam, D., Gold, L., Herberg, F.W., Hermjakob, H., Hoheisel, J.D., Joos, T.O., Kallioniemi, O., Koegl, M., Konthur, Z., Korn, B., Kremmer, E., Krobitsch, S., Landegren, U., van der Maarel, S., McCafferty, J., Muyldermans, S., Nygren, P.A., Palcy, S., Pluckthun, A., Polic, B., Przybylski, M., Saviranta, P., Sawyer, A., Sherman, D.J., Skerra, A., Templin, M., Ueffing, M. & Uhlen, M. (2007). ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nat. Methods*, 4, 13-7

Budd, A., Blandin, S., Levashina, E.A. & Gibson, T.J. (2004). Bacterial α2-macroglobulins: colonization factors acquired by horizontal gene transfer from the metazoan genome? *Genome Biol.* 5, R38

Diella, F., Cameron, S., Gemund, C., Linding, R., Via, A., Kuster, B., Sicheritz-Ponten, T., Blom, N. & Gibson, T.J. (2004). Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. BMC *Bioinformatics* 5, 79

Linding, R., Russell, R.B., Neduva, V. & Gibson, T.J. (2003). GlobPlot: Exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* 31, 3701-3708

Rob Russell

PhD 1993, Oxford University, UK.
Postdoctoral research at Imperial Cancer Resarch Fund, London, UK.
Senior Investigator at SmithKline Beecham, London, UK.
Group Leader at EMBL since 2001. Joint appointment with EMBL-EBI.

Structural bioinformatics

Previous and current research

Our research is focused broadly on the relationships between protein three-dimensional (3D) structure, function and evolution. A major interest over the past four years has been to develop methods to understand and predict the molecular details for macromolecular interactions. These efforts can be currently classified into three main areas.

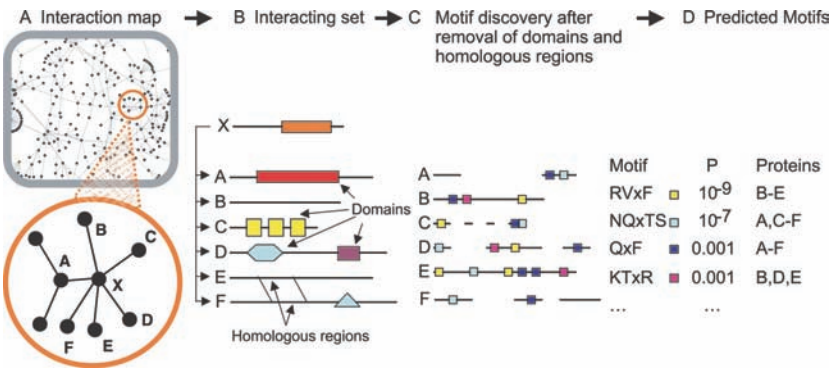
Protein interactions and complexes: Protein interaction networks are central to any understanding of cellular processes, and though many thousands are now known, few initiatives to uncover them pay much attention to one of the best sources of data available: complexes of known 3D structure. We thus study protein interactions by considering known 3D structures. We use 3D complexes to interrogate interactions identified by other methods (e.g. yeast two-hybrids) and to predict specific interactions within protein families. A major initiative in the group is related building as complete models as possible for all interacting proteins and complexes in a whole cell.

Protein and DNA sequence motifs mediating interactions: A major current challenge in biology is to discover and understand short protein or nucleic acid stretches that mediate functional interactions. We currently search for new protein-peptide and microRNA target sequences in genomes using a variety of techniques. Both methods already make fascinating predictions of biological phenomena and provide a wealth of information for people working with such sequences experimentally. We are also actively involved in experimental efforts to test or validate our predictions, either doing them ourselves or working in close collaboration with experimental groups.

Chemical biology: linking chemicals to pathways: We are now actively developing a database of chemicals linked to proteins in order to interrogate the likely effects of disrupting pathways with chemicals. This will be complemented by new and existing methods for the design of small-molecule or peptide inhibitors with the aim of working closely with experimentalists.

Future projects and goals

- To develop the 'one-stop-shop' for structural information on currently available information (predicted or experimental) on large protein complexes. This will involve development and application of new methods for complex structure prediction.
- To devise techniques to predict peptide binding sites on protein surfaces to complement our predictions of interacting peptides.
- To derive methods to predict the outcome of chemical treatment on particular pathways by cross-referencing chemicals with pathway components.



Schematic outlining our approach for finding protein linear motifs that mediate protein-protein interactions. Sets of proteins (A-F) sharing an interaction partner (X) are grouped and domains and homologous sequences are removed. We then search for 3-8 residue motifs that are over-represented in the remaining sequence, and score these by a binomial probability to give a ranked set of candidate motifs mediated the interaction with protein X.

Selected references

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M., Marzioch, M., Rau, C., Jensen, L.J., Bastuck, S., Dumpelfeld, B., Edelmann, A., Heurtier, M.A., Hoffman, V., Hoefert, C., Klein, K., Hudak, M., Michon, A.M., Schelder, M., Schirle, M., Remor, M., Rudi, T., Hooper, S., Bauer, A., Bouwmeester, T., Casari, G., Drewes, G., Neubauer, G., Rick, J.M., Kuster, B., Bork, P., Russell, R.B. & Superti-Furga, G. (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440, 631-636

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

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

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

Klaus Scheffzek

PhD, Max-Planck-Institut für Medizinische Forschung, Heidelberg.
Postdoctoral research at the Max-Planck-Institut für Molekulare Physiologie, Dortmund, Germany.
Team Leader at EMBL since 1999.

Signal transduction – disease proteins

Previous and current research

Defects in signalling pathways are often associated with the occurrence of severe diseases, with cancer being a very common example. We are interested in understanding the mechanisms of pathogenesis associated with cancer-related diseases. In previous work we have characterised the regulation of Ras, a GTP binding protein mutated in 30% of human tumours, and the related Rho proteins. Ras functions like a binary molecular switch cycling between GTP-bound 'ON'- and GDP-bound 'OFF'-states; Ras mediated GTP hydrolysis turns the switch off. This intrinsically slow process is enhanced by so-called GTPase activating proteins (GAPs). Oncogenic Ras mutants are permanently activated and are not sensitive to GAPs. In earlier studies we have elucidated the chemical mechanism of GTPase activation and explained why oncogenic Ras mutants are not GAP sensitive.

Currently our major focus is on neurofibromatosis type 1 (NF1), a genetic disease with an incidence of 1 in 3,500 worldwide. NF1 patients have an elevated tumour risk, may show a variety of developmental defects and frequently show learning disabilities. The NF1 gene encodes a huge protein of about 2,800 amino acids, termed neurofibromin, and when mutated is responsible for the pathogenesis of the disease. Neurofibromin acts as a Ras specific GAP, and in some tumour types lacking the protein, Ras is indeed upregulated. In previous studies we have characterised the GTPase activating domain and investigated potential effects of patient mutations.

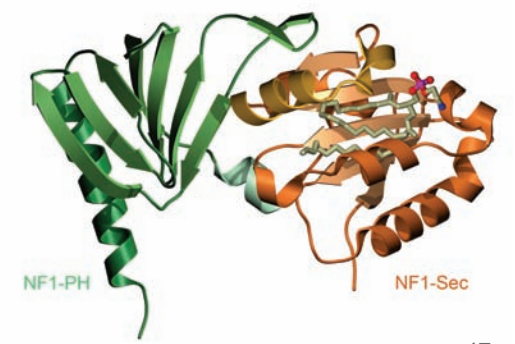
The GAP activity of neurofibromin resides in a segment which represents only 10% of the protein. After more than ten years of research following the discovery of the NF1 gene, no other biochemical function of neurofibromin has been clearly defined. We are following a structural proteomics approach to explore possible functions of the remaining 90% of the protein. The idea is to identify neurofibromin segments that can be expressed as soluble proteins, determine the structures of such fragments, and by comparison with known protein structures or by bound ligands obtain ideas for functional/biochemical experiments. Work on this project offers the opportunity to contribute to a challenging and physiologically exciting research topic. Our main technique is X-ray crystallography, with other methods of protein characterisation being increasingly employed.

Using the approach described above we have recently discovered a novel bipartite module containing a lipid binding Sec14-homology (NF1-Sec) and a previously undetected pleckstrin homology (NF1-PH)-like domain, the function of which we are now investigating. We have found that the Sec14-portion binds cellular glycerophospholipids, which represent major membrane components in cells. Our structural proteomics approach has been successful with other proteins and is becoming increasingly important in the functional analyses of genomic sequences coding for hypothetical proteins of hitherto unknown functions.

Future projects and goals

A major goal is to arrive at a 3D model of neurofibromin. This is a challenging task, given the size and the expected complexity of the molecule. Therefore we adopt a 'divide and conquer' strategy, which will enable us to obtain structures of at least some segments. We are increasingly including automated strategies to identify soluble protein fragments that are accessible to biochemical/structural analysis. In addition, we will continue searching for interaction partners of neurofibromin and investigate their role for the function of the protein. Studying Sec14 like domains in the context of other signal regulatory proteins such as RhoGAPs, RhoGEFs and PTPases will be an important direction in the future.

Other projects include signalling by eukaryotic and prokaryotic protein kinases, further aspects in structural neurobiology and regulation of transcription.



Structure of a bipartite module from neurofibromin composed of a Sec14 homologous (NF1-Sec) and a pleckstrin homology (NF1-PH) like domain bound to a cellular glycerophospholipid.

Selected references

Welti, S., Fraterman, S., D'Angelo, I., Wilm, M. & Scheffzek, K. (2007). The sec14 homology module of neurofibromin binds cellular glycerophospholipids: mass spectrometry and structure of a lipid complex. *J. Mol. Biol.*, 366, 551-562

D'Angelo, I., Welti, S., Bonneau, F. & Scheffzek, K. (2006). A novel bipartite phospholipid-binding module in the neurofibromatosis type 1 protein. *EMBO Rep.*, 7, 174-179

Kustatscher, G., Hothorn, M., Pugieux, C., Scheffzek, K. & Ladurner, A.G. (2005). Splicing regulates NAD metabolite binding to histone macroH2A. *Nat. Struct. Mol. Biol.*, 12, 624-625

Ahmadian, M.R., Kiel, C., Stege, P. & Scheffzek, K. (2003). Structural fingerprints of the Ras-GTPase activating proteins neurofibromin and p120GAP. *J. Mol. Biol.,* 329, 699-710

Reinhard Schneider

PhD 1994, University of Heidelberg, Germany. Postdoctoral research at EMBL.
Co-founder and Chief Information Officer at LION bioscience AG, Germany.
Chief Executive Officer at LION bioscience Research Inc., Cambridge, MA, USA.
Team Leader at EMBL since 2004.

Data integration and knowledge management

Previous and current research

Nowadays it is widely recognised that a comprehensive integration of data can be one of the key factors to improve productivity and efficiency in the biopharmaceutical research and development process. Successful data integration helps researchers to discover relationships that enable them to make better and faster decisions, thus considerably saving time and money. In a logical extension to these arguments one can apply the ideas and technologies used in industry in a basic research environment. An additional challenge in an academic environment is the even less structured 'process chain' with completely new data types and fast changing requirements from the end-users.

Over the last 20 years, biological research has seen a very strong proliferation of data sources. Each research group and each new experimental technique generates a source of valuable data. The creation, use, integration and warehousing of biological data is central to large-scale efforts in understanding biological systems. These tasks pose significant challenges from the standpoint of data storage, indexing, retrieval and system scalability over disparate types of data.

The principal aim of the group is to capture and centralise the knowledge generated by the scientists in the several divisions, and to organise that knowledge such that it can be easily mined, browsed and navigated. By providing access to all scientists in the organisation, it will foster collaboration between researchers in different cross-functional groups.

The group is involved in the following areas:

- data schema design and technical implementation;
- metadata annotation with respect to experimental data;
- design and implementation of a scientific data portal;
- providing access to and further developing data-mining tools (e.g. text-mining);
- visualisation of heterogeneous data mining results.

Future projects and goals

Our goal is to develop a knowledge platform for the life sciences. We will first focus on the biology-driven research areas, but will also extend into chemistry-related fields, primarily by collaborating with groups within EMBL. Other research areas will include advanced data-mining and visualisation techniques, data integration and knowledge management.

Selected references

Kremer, A., Schneider, R. & Terstappen, G.C. (2005). A bioinformatics perspective on proteomics: data storage, analysis, and integration. *Biosci. Rep.* 25, 95-106

Ofran, Y., Punta, M., Schneider, R. & Rost, B. (2005). Beyond annotation transfer by homology: novel protein-function prediction methods to assist drug discovery. *Drug Discov. Today* 10, 1475-1482

Dietrich Suck

PhD 1971, the Max Planck Institute for Experimental Medicine, Göttingen, Germany.
Postdoctoral research at the MPI, Göttingen, Germany,
and Purdue University, Indiana, USA.
Group Leader at EMBL since 1982.

Structural basis of protein-nucleic acid interactions

Previous and current research

We study protein-nucleic acid interactions with a combination of crystallographic and biochemical techniques to understand the structural basis of their widely varying specificities, the formation ofmulti-component protein-nucleic acid complexes and the catalytic mechanisms involved in nucleicacid modifications.

A continuing theme of our research has been the substrate recognition by structure-selective nucleases. The level of their structural specificity ranges from simple discrimination between single- and double-stranded substrates (nucleases P1 and S1), the recognition of DNA groove geometry and flexibility (DNase I), to the recognition of specialised structures, such as flap DNA (T5 5'-nuclease) or Holliday junctions (HJ) (T4 endonuclease VII, Cre recombinase). Currently we are working on HJ resolvases from phage T4 endonuclease VII (Endo VII), whose intrinsic conformational flexibility is thought to be of functional significance for its broad substrate specificity, archaeal Hjc's and yeast Cce1, which in contrast to Endo VII do specifically recognise and cleave 4-way DNA junctions. Crystals of several HJ complexes have been obtained.

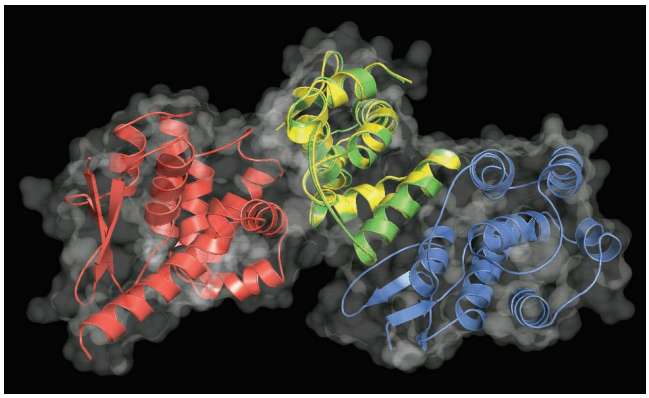
A major focus of our current work are the multi-component Arc1p and Ccr4-Not complexes from *S. cerevisiae*. Arc1p forms a stable complex with Met and Glu aminoacyl-tRNA synthetases (MetRS,GluRS) and stabilises their interaction with the cognate tRNAs. The yeast Arc1p complex functionally links tRNA nuclear export and aminoacylation and represents an evolutionary intermediate of the mammalian, multisubunit aminoacyl-tRNA synthetase complex. X-ray structures of the interacting N-terminal domains of Arc1p, GluRS and MetRS and their binary complexes reveal that all three domains adopt a GST-like fold, and that simultaneous interaction of Arc1p with GluRS and MetRS is mediated by a novel interface (between Arc1p/GluRS) in addition to a classical GST homodimer interface. Our data provide a structural model for ternary complex formation between Arc1p and the two synthetases

The Ccr4-Not core complex, a global gene expression regulator playing a major role in mRNA deadenylation, has a MW of ~1MDa and consists of 9 proteins. It contains two nucleases, Pop2 (also known as Caf1) and Ccr4, both of which are essential for deadenylation *in vivo*. We have solved the crystal structure of the Pop2p non-canonical nulease domain revealing its structural homology with the 3'-5'exonuclease domains of E. coli DNA polymerases I and III. Several constructs of the Ccr4p, Pop2p and Not1p proteins are currently being expressed/co-expressed to study their interactions.

Another project deals with Sm-like proteins from archaea and bacteria. Our X-ray structures of the *E. coli* Hfq protein and RNA complexes of Sm-related proteins from archaea (*A. fulgidus*, *P. abyssi* and *S. solfataricus*) have provided a high resolution picture of RNA-binding in an Sm core domain and revealed the close evolutionary relationship to eukaryotic snRNPs. Based on immunoprecipitation experiments we proposed a possible role of the A. fulgidus Sm proteins in tRNA maturation. To learn more about the function of archaeal Sm proteins, we are trying to identify their *in vivo* partners by expressing tagged proteins in *S. solfataricus*.

Future projects and goals

Future projects include the X-ray structure determinations of substrate complexes of Endo VII and other resolvases, of the multicomponent Arc1p-tRNA synthetase and Ccr4-Not complexes and structure-function studies of Sm-related proteins. By determining the crystal structures of these complexes, we want to reveal the structural basis for their formation, their selectivity and function.



Model of ternary complex formation of Arc1p (green/yellow) with Glu- (red) and Met-(blue) tRNA synthetases derived from the X-ray structures of binary sub-complexes of their N-terminal interacting domains.

Selected references

Simader, H., Hothorn, M., Kohler, C., Basquin, J., Simos, G. & Suck, D. (2006). Structural basis of yeast aminoacyl-tRNA synthetase complex formation revealed by crystal structures of two binary sub-complexes. *Nucleic Acids Res.*, 34, 3968-79

Thore, S., Mauxion, F., Seraphin, B. & Suck, D. (2003). X-ray structure and activity of the yeast Pop2 protein: a nuclease subunit of the mRNA deadenylase complex. *EMBO Rep.*, 4, 1150-1155

Sauter, C., Basquin, J. & Suck, D. (2003). Sm-like proteins in Eubacteria: the crystal structure of the Hfq protein from *Escherichia coli*. *Nucleic Acids Res.*, 31, 4091-4098

Thore, S., Mayer, C., Sauter, C., Weeks, S. & Suck, D. (2003). Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. Common features of RNA binding in archaea and eukarya. *J. Biol. Chem.*, 278, 1239-1247

Directors' Research

Directors' Research is unlike other EMBL units in that it covers three independent research groups without an overall coordinator. These groups are headed by the Director General and Associate Director of EMBL and the Executive Director of EMBO. As these appointments involve people with substantial management responsibilities for all the units of EMBL or for EMBO respectively, it was logical to define the groups as a separate unit rather than placing them under the control of a specific unit coordinator.

The Mattaj Group studies diverse processes that are under the control of the Ran GTPase. During mitosis, Ran is needed for both mitotic spindle assembly and nuclear envelope (NE) formation. Remarkably, their studies suggest that Ran's mitotic functions occur by the same mechanism as nucleocytoplasmic transport, i.e. via regulation of interactions between transport receptors and factors involved in spindle or NE assembly. Using *in vitro* and *in vivo* methods, they are identifying the factors that mediate these processes and finding out how they function and how Ran controls their activity. NE assembly is a multi-stage process. Both the membranes that give rise to the NE and the proteins that form nuclear pore complexes (NPCs), through which transport across the NE occurs, associate with the chromatin surface early in NE assembly. Membrane fusion events and NPC assembly proceed in concert, and have to be regulated in an integrated way. The group has begun to understand how Ran controls NPC assembly, but has little information on how the NE membranes assemble, or how NPC insertion into the membranes takes place. In addition, although it is known that Ran regulates where NE assembly occurs in the cell, they do not know how the process is temporally regulated, i.e. why it occurs in telophase rather than at other times during mitosis. Understanding the spatial regulation of mitotic events by Ran and their temporal regulation during the cell cycle is an ambitious long-term goal.

The Hentze Group combines interests in the posttranscriptional regulation of gene expression and in mammalian iron metabolism with research on diseases that result from disturbances in both areas. Their work on posttranscriptional control mainly addresses the regulation of protein synthesis, examining the mechanisms of action of regulatory RNA-binding proteins and/or miRNAs on the translational apparatus. In the context of the Molecular Medicine Partnership Unit (MMPU), they also investigate (jointly with Andreas Kulozik from Heidelberg University) nonsense-mediated RNA decay and 3' end processing as aspects of mRNA metabolism that give rise to common hematological disorders. The use of mouse models has become central to the exploration of the IRE/IRP network in mammalian iron homeostatis. The group studies the importance of this regulatory network for physiological cell and organ functions as well as its involvement in human disorders. Together with Martina Muckenthaler of Heidelberg University, the group also undertakes research in the MMPU on the regulation of the iron hormone hepcidin and its involvement in iron overload and deficiency diseases.

The Gannon Group focuses on the regulation of expression of genes regulated by the steroid hormone estrogen. The physiological effects of estrogen are transduced through specific nuclear proteins, the Estrogen Receptors, which are estrogen-dependent transcription factors. They have defined the promoter regions of ER-α from human, chicken and mouse, and described multiple start points of transcription. In addition, they have demonstrated that the 3'-untranslated region of the human oestrogen receptor has a role in destabilising the ER-α messenger RNA. More recently, they have comprehensively portrayed the sequence of events, instigated by Estrogen Receptor, that achieve and then limit transcription of estrogen responsive promoters. Ultimately, the group's aim is to understand how estrogen regulates a number of diverse physiological processes, such as embryonic development, sex determination and reproduction. This information is essential to understanding the role of steroid hormone receptors in the onset and progression of a variety of pathological conditions such as cancer, osteoporosis and Alzheimer's disease.

Iain Mattaj

PhD 1979, University of Leeds, UK.
Postdoctoral work at the Friedrich Miescher Institute and the Biocenter, Basel, Switzerland.
Group Leader at EMBL since 1985. Programme Coordinator since 1990.
Scientific Director of EMBL 1999–2005. Director General since 2005.

The RanGTPase as a spatial regulator

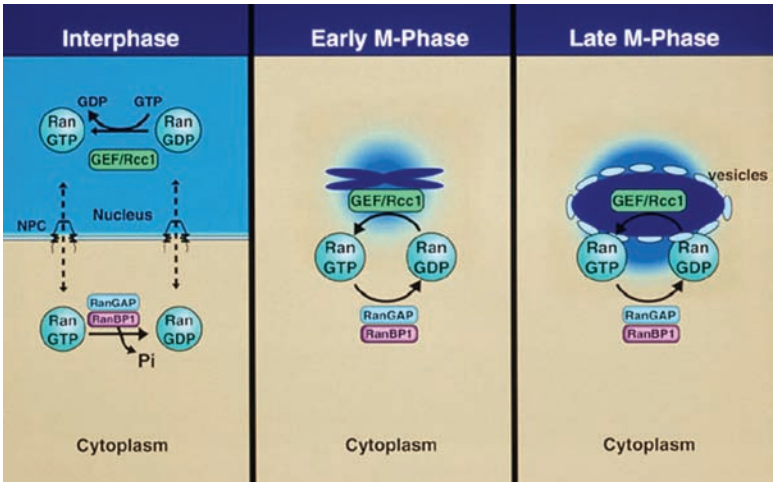
Previous and current research

The research in our group is centred on diverse processes that are under the control of the Ran GTPase. Ran requires regulators for its activity. These are a GEF (Guanine Nucleotide Exchange Factor) that loads Ran with GTP, and a GAP (GTPase Activating Protein) that is required for Ran to hydrolyse GTP to GDP. These two factors are asymmetrically distributed within cells, both in interphase and during mitosis (see figure). As a result, RanGTP is present in interphase at high concentration in the nucleus and at low concentration in the cytoplasm. In mitosis, RanGTP production is locally increased in the vicinity of chromatin because of the concentration of the GEF there. RanGTP interacts with the import and export receptors that mediate nucleo-cytoplasmic transport. In the former case, this interaction results in import cargo release. Thus, import receptors bind cargo in the cytoplasm (low Ran GTP) and release it in the nucleus (high RanGTP). Ran's effect on export receptors is the opposite, they only interact with their cargoes in the presence of RanGTP. These RanGTP-dependent binding events impart directionality to nuclear transport. During mitosis, Ran is needed for both mitotic spindle assembly and nuclear envelope (NE) formation (see figure). Remarkably, our studies suggest that Ran's mitotic functions occur by the same mechanism as nucleo-cytoplasmic transport, i.e. via regulation of interactions between transport receptors and factors involved in spindle or NE assembly.

Future projects and goals

In the case of spindle formation, we know that Ran regulates multiple aspects of spindle assembly; microtubule nucleation, microtubule stability, the production of anti-parallel microtubule arrays, the focusing of the spindle poles, etc. We are using *in vitro* and *in vivo* methods to identify the factors that mediate these processes and to find out how they function and how Ran controls their activity.

NE assembly is a multi-stage process. Both the membranes that give rise to the NE and the proteins that form nuclear pore complexes (NPCs), through which transport across the NE occurs, associate with the chromatin surface early in NE assembly. Membrane fusion events and NPC assembly proceed in concert, and have to be regulated in an integrated way. We have begun to understand how Ran controls NPC assembly but we have essentially no information on how the NE membranes assemble, or how NPC insertion into the membranes takes place. In addition, although we know that Ran regulates where NE assembly occurs in the cell, we do not know how the process is temporally regulated, i.e. why it occurs in telophase rather than in metaphase. Understanding the spatial regulation of mitotic events by Ran and their temporal regulation during the cell cycle is an ambitious long-term goal.



The figure depicts the local production of RanGTP, caused by the non-uniform distribution of the Ran GEF and Ran GAP in relation to the nuclear envelope or the chromosomes at various stages in the cell cycle.

Selected references

Galy, V., Askjaer, P., Franz, C., Lopez-Iglesias, C. & Mattaj, I.W. (2006). MEL-28, a novel nuclear-envelope and kinetochore protein essential for zygotic nuclear-envelope assembly in *C. elegans*. *Curr. Biol.*, 16, 1748-1756

Mansfeld, J., Guttinger, S., Hawryluk-Gara, L.A., Pante, N., Mall, M., Galy, V., Haselmann, U., Muhlhausser, P., Wozniak, R.W., Mattaj, I.W., Kutay, U. & Antonin, W. (2006). The conserved transmembrane nucleoporin NDC1 is required for nuclear pore complex assembly in vertebrate cells. *Mol. Cell*, 22, 93-103

Hetzer, M.W., Walther, T.C. & Mattaj, I.W. (2005). Pushing the envelope: structure, function, and dynamics of the nuclear periphery. *Annu. Rev. Cell Dev. Biol.*, 21, 347-380

Mattaj, I.W. (2004). Sorting out the nuclear envelope from the endoplasmic reticulum. *Nat. Rev. Mol. Cell Biol.*, 5, 65-69

Matthias W. Hentze

MD 1984, University of Münster, Germany.
Postdoctoral training at the NIH, Bethesda, USA.
Group Leader at EMBL since 1989; Senior Scientist since 1998. Associate Director since 2005.
Co-Director of the EMBL/University of Heidelberg Molecular Medicine Partnership Unit since 2002.

Cytoplasmic gene regulation and molecular medicine

Previous and current research

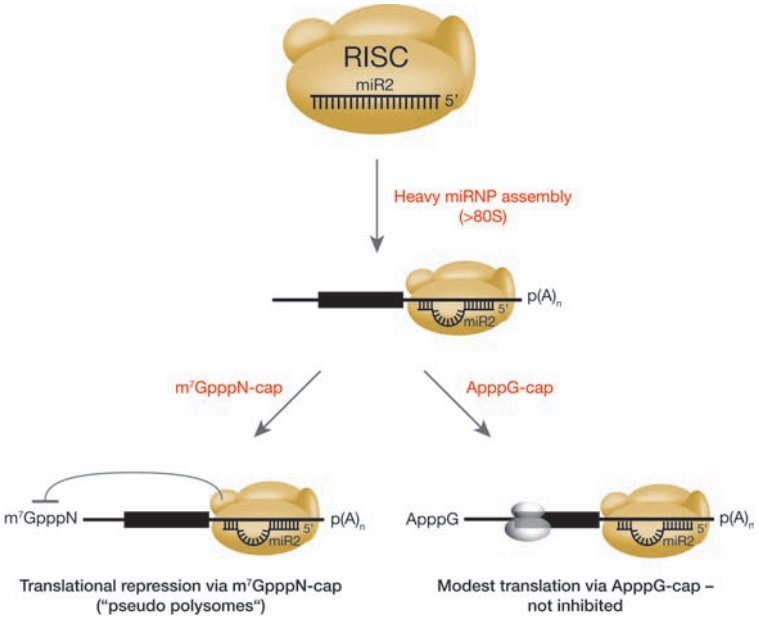
Important steps in the control of gene expression are executed in the cytoplasm: the regulation of mRNA translation and stability. We are elucidating these regulatory mechanisms, including the function of miRNAs, which has become a very active focus of our work (see figure). We use mostly biochemical approaches and mammalian, yeast and *Drosophila* model systems.

Within the Molecular Medicine Partnership Unit (MMPU), we are investigating the mammalian post-transcriptional mRNA quality control mechanism 'nonsense-mediated decay' (NMD) and its importance in genetic diseases (together with Andreas Kulozik). We also study the role of miRNAs in cancer and other diseases (together with Andreas Kulozik and Martina Muckenthaler). For information on research topics and projects within the MMPU, please see www.embl.org/research/partners/mmpu.html.

Our second major interest is the 'systems biology' of mammalian iron metabolism. This work includes the system-wide exploration of the functions of the IRE/IRP regulatory network. Within the MMPU (together with Martina Muckenthaler), we study the molecular basis of genetic and non-genetic diseases of human iron metabolism. Our work employs conditional knockout mouse strains for IRP1 and IRP2 and mouse models of iron metabolism diseases. We also use a unique DNA microarray platform (the IronChip) that we have developed.

Future projects and goals

- To uncover the basic mechanisms underlying protein synthesis and its regulation by miRNAs and RNA-binding proteins in cell metabolism, differentiation and development.
- To help elucidate the role of RNA metabolism in disease, and to develop novel diagnostic and therapeutic strategies based on this knowledge.
- To understand the molecular mechanisms and regulatory circuits to maintain physiological iron homeostasis and its connections to the immune system.
- To contribute to the elucidation of the molecular pathophysiology of common iron overload (haemochromatosis), iron deficiency (anaemia) and iron management (anaemia, Parkinson's disease) disorders.



Micro RNA-mediated translational control.

Selected references

Galy, B., Holter, S.M., Klopstock, T., Ferring, D., Becker, L., Kaden, S., Wurst, W., Grone, H.J. & Hentze, M.W. (2006). Iron homeostasis in the brain: complete iron regulatory protein 2 deficiency without symptomatic neurodegeneration in the mouse. *Nat. Genet.*, 38, 967-969

Chekulaeva, M., Hentze, M.W. & Ephrussi, A. (2006). Bruno acts as a dual repressor of *oskar* translation, promoting mRNA oligomerization and formation of silencing particles. *Cell*, 124, 521-533

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, 368-379

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

Frank Gannon

PhD 1973, University of Leicester, UK. Postdoctoral research at the University of Wisconsin, Madison, USA and the University of Strasbourg, France. Associate Professor and Director of the National Diagnostic Centre, University College Galway, Ireland. Executive Director of EMBO. At EMBL since 1994.

The expression and functional regulation of Estrogen Receptor-α

Previous and current research

A major focus of my laboratory's work is the control of the expression of genes by the Estrogen Receptor-α (ER-α). Although there have been very many studies on this topic, there is still an amazing amount of information that is uncovered every year. In the recent past we have turned our attentions specifically on the detailed description of the mechanism of action of the ER-α. The process that we have defined, in a very precise manner, is the cycling of the ER on the promoter of a target gene. The ER binds to the promoter in the presence or absence of oestradiol, and thereafter follows a sequence of recruitment of co-factors which ultimately results in the binding of polymerase II (in the presence of ligand) or in the non-productive cycle for the receptor when no ligand is present. Almost 50 factors related to transcription were monitored in this ChIP based experiment and the data showed not only that the receptor is cycling, but also that the process had many redundant aspects and that the steps that are normally viewed as being negative (e.g. degradation of the receptor by the proteasome) are in fact an inherent component of the successful action of the estrogen receptor.

Following on the messages from these studies, we are now developing new tools to interfere with this process. The first of these was the use of histone deacetylase inhibitors, but others are following from chemogenomic approaches. We are also studying the process using proteomics and extending the studies to other target genes for the estrogen receptor including those that are down regulated.

Future projects and goals

At all times our studies are linked to human disease and our current focus remains on human breast cancer and osteoporosis. The laboratory, therefore, links very fundamental studies with the potential for an impact on human medicine and in this way fits well into the molecular medicine aspirations of EMBL. In doing so, we continue to collaborate very actively with groups from the Cell Biology and Biophysics, Gene Expression and Structural and Computational Biology Units, as well as with several of the core facilities, particularly with the Genomics and Chemical Biology facilities.

Selected references

Metivier, R., Reid, G. & Gannon, F. (2006). Transcription in four dimensions: nuclear receptor-directed initiation of gene expression. *EMBO Rep.* 7, 161-167

Metivier, R., Penot, G., Carmouche, R.P., Hubner, M.R., Reid, G., Denger, S., Manu, D., Brand, H., Kos, M., Benes, V. & Gannon, F. (2004). Transcriptional complexes engaged by apo-estrogen receptor-α isoforms have divergent outcomes. *EMBO J.* 23, 3653-3666

Metivier, R., Penot, G., Hubner, M.R., Reid, G., Brand, H., Kos, M. & Gannon, F. (2003). Estrogen receptor-α directs ordered, cyclical and combinatorial recruitment of cofactors on a natural target promoter. *Cell* 115, 751-763

Reid, G., Metivier, R., Lin, C.Y., Denger, S., Ibberson, D., Ivacevic, T., Brand, H., Benes, V., Liu, E.T. & Gannon, F. (2005). Multiple mechanisms induce transcriptional silencing of a subset of genes, including estrogen receptor-α, in response to deacetylase inhibition by valproic acid and trichostatin A. *Oncogene* 24, 4894-4907

Reid, G., Hubner, M.R., Metivier, R., Brand, H., Denger, S., Manu, D., Beaudouin, J., Ellenberg, J. & Gannon, F. (2003). Cyclic, proteasome-mediated turnover of unliganded and liganded ER-α on responsive promoters is an integral feature of estrogen signalling. *Mol. Cell* 11, 695-707

Core Facilities

There are three reasons underlying the reorganisation over the past five years of certain EMBL activities into Core Facilities. The first is related to the high cost of equipment and materials required for certain experimental methods that have become a standard part in the repertoire of many EMBL groups, such as microarrays, MS of proteins, high-throughput robotic sample handling and so on.

The second is that EMBL's research groups are typically of modest size, both in terms of the number of people and the size of their budget. Since these groups follow ambitious research programmes that require diverse technologies, it would be impossible for them to attract the capital and personnel investment required to carry out all their planned projects if they had to do so independently.

The third is a result of EMBL's experience with a previous model for providing access to critical, widely used technologies, which was to incorporate them into one of our research groups. This model proved unsatisfactory because groups, and group leaders, were placed in a constant conflict in having to divide their time and energy between, on the one hand, carrying out research and development projects (on whose basis they would largely be judged on leaving EMBL) and, on the other hand, providing service to their colleagues.

Particularly for techniques where supply can virtually never meet demand, such as mass spectrometry or EM, the situation of the mixed service and research service groups was very problematic.

For all these reasons, it was cost effective and efficient for EMBL to launch the Core Facilities to provide access to technologies that are either expensive to set up or maintain, or which require considerable expertise. The establishment of the Core Facilities was started in 2001. Today, the current facilities are Advanced Light Microscopy, Genomics, Proteomics, Protein Expression and Purification, Electron Microscopy, Flow Cytometry, Monoclonal Antibody and Chemical Biology.

Christian Boulin
Head of Core Facilities

Rainer Pepperkok

PhD 1992, University Kaiserslautern, Germany.
Postdoctoral research at University of Geneva, Switzerland.
Lab Head at the Imperial Cancer Research Fund, London, UK.
At EMBL since 1998.

Advanced Light Microscopy Core Facility

The Advanced Light Microscopy Facility (ALMF) offers a collection of state-of-the-art light microscopy equipment and image processing tools. The ALMF was set up as a cooperation between EMBL and industry to improve communication between users and producers of high-end microscopy technology and to support in-house scientists and visitors in using light microscopy methods for their research. The facility also organises regular international courses to teach advanced light microscopy methods.

Major projects and accomplishments

- The ALMF presently manages 17 top-of-the-line microscope systems and four image analysis workstations from leading industrial companies.
- More than 30 visitors per year come to carry out their own experiments in the ALMF or to evaluate microscopy equipment.
- The ALMF was the seed for the European Light Microscopy Initiative (ELMI) that establishes links with light microscopy facilities throughout Europe.
- Usage of the facility has exceeded 30,000 hours per year.

Services provided

Assisting all EMBL groups and visitors through the microscopic imaging process: project planning, sample preparation, staining, microscope selection and use, image processing, presentation, data transfer and storage.

Developing accessory software and microscopy equipment, co-developments with industrial partners, pre-evaluation of commercial equipment.

Technology partners

The ALMF presently has collaborations with the following companies:

- Applied Precision*
- Bitplane*
- Carl Zeiss*
- Cell Biotrading*
- Eppendorf*
- jpk Instruments*
- Leica Microsystems*
- Olympus Europe*
- Olympus BioSystems*
- Perkin Elmer*
- Scientific Volume Imaging
- Visitron*

*The ALMF also serves as a reference site for these companies.

Selected references

Brunet, S., Zimmermann, T., Reynaud, E.G., Vernos, I., Karsenti, E. & Pepperkok, R. (2006). Detection and quantification of protein-microtubules interactions using green fluorescent protein photoconversion. *Traffic*, 7, 1283-9

Graf, R., Rietdorf, J. & Zimmermann, T. (2005). Live cell spinning disk microscopy. *Adv. Biochem. Eng. Biotechnol.*, 95, 57-75

Pepperkok, R., Simpson, J.C., Rietdorf, J., Cetin, C., Liebel, U., Terjung, S. & Zimmermann, T. (2005). Imaging platforms for measurement of membrane trafficking. *Methods Enzymol.*, 404, 8-18

Zimmermann, T., Rietdorf, J. & Pepperkok, R. (2003). Spectral imaging and its applications in live cell microscopy. *FEBS Lett.*, 546, 87-92

Joe Lewis

PhD 1991, Institute of Molecular Pathology, Vienna, Austria.
Postdoctoral research at EMBL.
Group Leader at the Wellcome Trust Centre for Cell Biology, Edinburgh, Scotland.
Group and Global HCV Project Leader at Anadys Pharmaceuticals, Heidelberg, Germany.
Facility Head at EMBL since 2004.

Chemical Biology Core Facility

Small molecules play essential roles in many areas of basic research and are often used to address important biological questions. The aim of our Chemical Biology Core Facility is to enable research groups to address biological questions by identifying and developing 'biotool' compounds against novel targets. We can assist groups in the development of primary and secondary assays for screening against our in-house compound library and guide them through the process of developing tool compounds for their specific target. Chemical optimisation projects can be done in collaboration with our chemistry partners.

The facility is a collaboration between EMBL and the DKFZ (German Cancer Research Center, Heidelberg) to provide the infrastructure and expertise to open up small molecule development to research groups at both institutions.

Major projects and accomplishments

The facility was established at the beginning of 2004. Through a large team effort, it was fully functional in August 2004. We have a very strong pipeline of projects from EMBL and DKFZ, several of which have completed screening and are now in the lead optimisation phase.

Services provided

- Selection of appropriate assay technology platforms.
- Developing assays for medium-throughput screening.
- Small molecule screening library of greater than 50,000 compounds.
- Assisting in the design of secondary specificity assays.
- Compound characterisation.
- Managing compound acquisition through our chemistry partners.

Partners

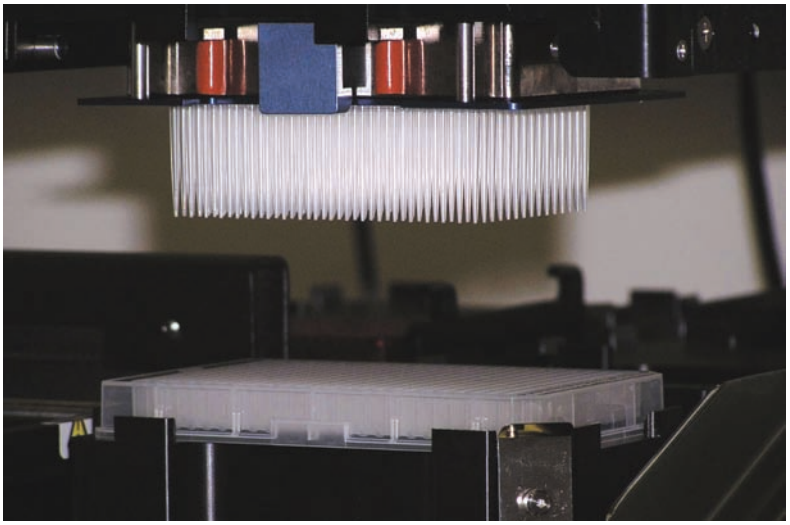
- Technology partners: Perkin Elmer, IDBS.
- Chemistry partners: Tripos Inc. and Tripos Discovery Research Ltd.

Selected references

Boehringer, D., Thermann, R., Ostareck-Lederer, A., Lewis, J. D. & and Stark, H. (2005). Structure of the hepatitis C virus IRES bound to the human 80S ribosome: Remodelling of the HCV IHRES. *Structure* 13, 1695-1706

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

Ostareck-Lederer, A., Clauder-Münster, S., Thermann, R., Polycarpou-Schwartz, M., Gentzel, M., Wilm, M. & Lewis, J.D. (2005). The role of RNA interference in drug target validation: Application to Hepatitis C. In "RNA Interference Technology", Appasani, K. (ed.), Cambridge University Press, 318-330



Parallel pipetting of samples in 384-well format.

Claude Antony

PhD 1984, Université Paris VI, France.
Postdoctoral research at EMBL.
Group Leader at CNRS.
Facility Head and Team Leader at EMBL since 2003.

Electron Microscopy Core Facility

The EMCF gives EMBL scientists access to advanced electron microscopes, relevant sample preparation techniques and specialised instrumentation. Techniques can be applied and adapted to various projects across the units to access EM resolution at cell level. The facility also trains new users to make best use of our advanced equipment. Finally, we develop new approaches and methods in EM applications to cellular and developmental biology.

Major projects and accomplishments

Plastic embedded protein crystals *(collaboration with Raimond Ravelli, EMBL Grenoble, page 66)*. The technique of freeze-substitution has been applied to vitrified protein crystals (lysozyme) by high pressure freezing. (Ravelli *et al.*, 2007).

Setting up correlative microscopy *(collaboration between EMCF and ALMF)*. We are promoting a methodology to perform correlative microscopy with conventionally fixed cells on coverslips. (A methods paper is submitted).

A p53 network built in budding yeast. *Barbara di Ventura (Serrano group, formerly EMBL Heidelberg).*

The mechanisms that regulate and shape newly formed membrane structures in meiotic budding yeast. *Peter Maier (Knop group, page 14).*

The role of the small GTPase Rab6 in the polarisation of the oocyte during oogenesis in *Drosophila*. *Jean-Baptiste Coutelis (Ephrussi group, page 22).*

The nuclear pore complex assembly in an *in vitro* system based on *Xenopus* egg extracts, checking the phenotypes of various depletions on the structure of the nuclear envelope and nuclear pore complexes. *Wolfram Antonin (Mattaj group, page 52).* (Mansfeld *et al.*, 2006).

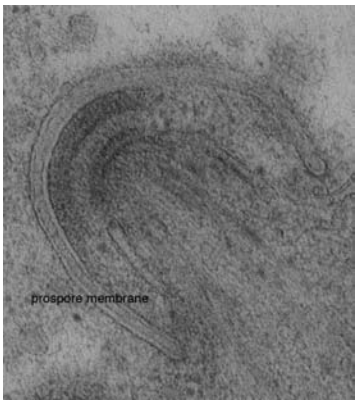
Visualsing the spindle microtubules anchoring at the SPB (spindle pole body) in a mutant context affecting Msd1p, essential in this process. *Mika Toya (CRUK, London; formerly the Brunner group, page 10).* (Toya *et al.*, 2006, in revision).

Services provided

- An up-to-date know-how on EM methods for cell biology, immunocytochemistry, cryosectioning and cryofixation applied to various cell types or organisms.
- Maintaining the electron microscopes and the equipment in the laboratory for sample preparation, microtomy and cryogenic methods.
- Supplying a range of reagents specific for the relevant EM methods and protocoles.
- Assisting users in choosing the right methods and protocols for their project.
- Organizing courses and lectures on EM methods in cell biology.

Technology partners

- FEI Company: Supplier of advanced electron microscopes.
- Leica-microsystems is the constructor of our HPFreezer EMPACT2. The EMPACT2 is a portable machine and has an optional attachment, the Rapid Transfer System (RTS), which permits easy loading of the samples and allows correlative light and electron microscopy.



Spore formation in budding yeast (cryofixed cells). The newly forming prospore membrane is seen connected with the outer face of the meiotic SPB while microtubules from the meiotic spindle are seen attached at the inner face of the SPB.

Selected references

Ravelli, R.B., Haselmann-Weiss, U., McGeehan, J.E., McCarthy, A.A., Marquez, J.A., Antony, C., Frangakis, A.S. & Stranzl, G. (2007). Plastic-embedded protein crystals. *J. Synchrotron Radiat.*, 14, 128-132

Mansfeld, J., Guttinger, S., Hawryluk-Gara, L.A., Pante, N., Mall, M., Galy, V., Haselmann, U., Muhlhausser, P., Wozniak, R.W., Mattaj, I.W., Kutay, U. & Antonin, W. (2006). The conserved transmembrane nucleoporin NDC1 is required for nuclear pore complex assembly in vertebrate cells. *Mol. Cell*, 22, 93-103

Pizon, V., Gerbal, F., Diaz, C.C. & Karsenti, E. (2005). Microtubule-dependent transport and organization of sarcomeric myosin during skeletal muscle differentiation. *EMBO J.*, 24, 3781-3792

Franz, C., Askjaer, P., Antonin, W., Iglesias, C.L., Haselmann, U., Schelder, M., de Marco, A., Wilm, M., Antony, C. & Mattaj, I.W. (2005). Nup155 regulates nuclear envelope and nuclear pore complex formation in nematodes and vertebrates. *EMBO J.*, 24, 3519-31

Andrew Riddell

BSc Hons 1992, Paisley University, Scotland.
PgDip, 1993, Caledonian University, Glasgow, Scotland.
Work at the MRC LMB and CIMR and Hutchison/MRC, Cambridge, UK.
Facility Head at EMBL since 2003.

Flow Cytometry Core Facility

The Flow Cytometry Core Facility offers a range of flow cytometric techniques. The equipment adds flexibility in the preparation and execution of experiments allowing different approaches to research problems. Our facility meets researchers' needs and enables better resolution in terms of analysis and product.

The goal of the facility is to proactively introduce flow cytometric methods into new research areas while supporting and extending current research.

Major projects and accomplishments

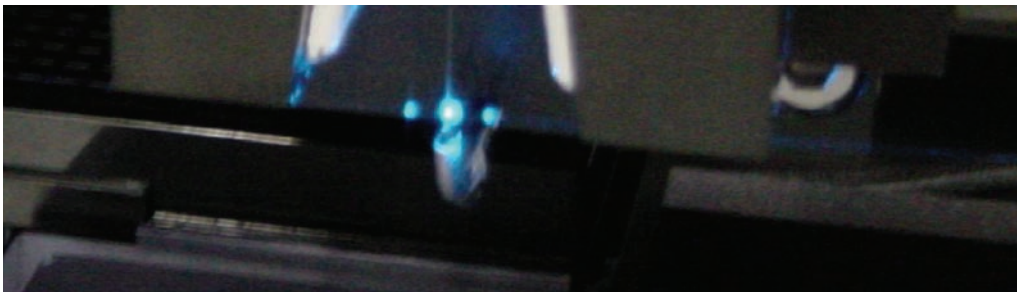
- The analysis of algae life cycle project relied on the intrinsic fluorophores in algae to identify life cycle stages. It required photosaturation of the photosynthetic units in the algae.
- Cell cloning by Darwinian selection required a series of single cell sorts of a target population into 96-well plate in order to select a stable integrated gene into a cell line.
- There is a project investigating a bi-stable state of a reworked bacterial signalling cascades that requires precise and accurate instrument measurements of the bacteria in order for them to be identified.
- The apoptosis project utilises a novel FRET assay in order to identify apoptosis in a cell line.
- An ongoing project, in collaboration with the University of Heidelberg's Chemistry Department, investigates the flow cytometric analysis of cellular uptake of novel synthetically produced probes.

Services provided

- Sorting heterogeneous single cell populations into homogeneous populations for experiments.
- Providing an analysis of single cell populations based on fluorescent probes and light intensities (including light scattering and polarisation).
- Providing expertise in flow cytometric techniques for use in experiments.
- Providing advice in the use of flow cytometry.
- Developing novel flow cytometric techniques for use in the EMBL's scientific activities.

Technology partners

We work with equipment from Cytopeia Inc., DAKO, Becton Dickinson, Union Biometrica and Miltenyi Biotec. We are open to test new technological developments to best serve the needs of the scientific community.



Deflection illumination for calibrating droplet break-off point.

Selected references

Füssl, A., Schleifenbaum, A., Goritz, M., Riddell, A., Schultz, C., Kramer, R. (2006). Cellular uptake of PNA-terpyridine conjugates and its enhancement by Zn2+ ions. *J. Am. Chem. Soc.*, 128, 5986-5987

Corcoran, A.E., Riddell, A., Krooshoop, D. & Venkitaraman, A.R. (1998). Impaired immunoglobulin gene rearrangement in mice lacking the IL-7 receptor. *Nature* 391, 904-907

Fairley, E.A., Riddell, A., Ellis, J.A. & Kendrick-Jones, J. (2002). The cell cycle dependent mislocalisation of emerin may contribute to the Emery-Dreifuss muscular dystrophy phenotype. *J. Cell Sci.* 115, 341-354

Goyenechea, B., Klix, N., Yelamos, J., Williams, G.T., Riddell, A., Neuberger, M.S. & Milstein, C. (1997). Cells strongly expressing Ig(κ) transgenes show clonal recruitment of hypermutation: a role for both MAR and the enhancers. *EMBO J.* 16, 3987-3994

Vladímir Beneš

PhD 1994, Czech Academy of Sciences,
Prague, Czech Republic.
Posdoctoral research at EMBL.
Facility Head since 2001.

Genomics Core Facility (GeneCore)

The availability of whole genome sequence data from many organisms has transformed certain aspects of molecular biology. GeneCore was founded in 2001 to implement the various technological approaches of functional genomics into EMBL's research projects. It is equipped with state-of-the-art hardware required for transcriptome and global location analyses and operated by highly-qualified staff who assist with projects and their troubleshooting in an interactive environment. As training is an inseparable part of GeneCore activities, with staff tutoring individual researchers as well as in organising practical courses on subjects such as quantitative real-time RT-PCR (qPCR) or gene expression profiling and its data analysis.

Services provided

- DNA sequencing, as groundwork for any functional genomics studies;
- DNA microarrays, qPCR and SAGE (serial analysis of gene expression), where knowledge of sequence information is exploited for transcriptome profiling and beyond;
- liquid handling robotics for medium throughput pipetting projects as a basis for large scale projects.

In principle, activities of GeneCore are divided into three modes of operation:

- **sample focused:** DNA sequencing, Bioanalyser; miRNA qPCR profiling;
- **project focused:** microarrays (home-made, commercial), liquid handling robotics and SAGE;
- **access to instruments and complete support:** qPCR, NanoDrop, PCR cyclers, microarray-scanner, high-capacity vacuum concentrator.

In 2006, the DNA Sequencing service team processed more than 44,000 reactions on samples from almost all 'wet lab' groups from EMBL Heidelberg and Grenoble and some from Hamburg. Two 96-capillary MegaBACE sequencers are currently employed to analyse this high number of sequencing reactions. To deal with increasing flow of samples we have developed and launched an electronic sample registration and tracking system, allowing users to register their samples online and track the process with a bar code.

qPCR enables detection of specific target sequence and its quantification with high sensitivity. Three qPCR instruments managed by GeneCore are primarily used for gene transcript quantification to complement and corroborate microarray analyses, as well as for detailed DNA occupancy profiling after chromatin immunoprecipitation (ChIP). EMBL researchers analysed about 110,000 qPCR assayed points in 2006.

During 2006 the DNA microarray team implemented together with the groups of Eileen Furlong (page 23), Asifa Akhtar (page 31) and Jürg Müller (page 34) protocols for genome-wide location analysis on *Drosophila* high-density tiling arrays (ChIP-on-chip). We will use these protocols in the project with Frank Gannon's group (page 53) whose aim is the whole human genome scale analysis of estrogen binding sites. Exon arrays to study alternative splicing of individual transcripts represent another addition to GeneCore's portfolio of provided microarray analyses. These arrays enable exploration of human, mouse and rat transcriptomes into unprecedented depth of individual exons and thus stand for a fundamental tool for determination of their expression pattern.

With 1,100 submitted and processed samples GeneCore's microarray section also experienced an increase in its activities in 2006.

MicroRNAs (miRNAs) are now recognised as an important class of small RNA molecules that play important role in regulation of gene expression. Research into non-coding RNAs and particularly of this category is currently one of the hottest and most challenging areas in molecular biology. Detailed analysis of miRNAs is technically challenging because their mature form is only ~22 frequently highly homologous nucleotides long. In partnership with Matthias Hentze's group (page 32), we have worked closely together on the development of the locked-nucleic-acids based oligonucleotide array called miChip prepared for profiling of human and mouse mature microRNAs. This valuable system is complemented with a complete panel of qPCR TaqMan miRNA assays for human, mouse and *Drosophila*.

A detailed overview of the GeneCore activities can be found at www.genecore.embl.de.

Selected references

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, 913-920

Bustin, S.A., Benes, V., Nolan, T. & Pfaffl, M.W. (2005). Quantitative real-time RT-PCR – a perspective. *J. Mol. Endocrinol.*, 34, 597-601

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

Reid, G., Metivier, R., Lin, C.Y., Denger, S., Ibberson, D., Ivacevic, T., Brand, H., Benes, V., Liu, E.T. & Gannon, F. (2005). Multiple mechanisms induce transcriptional silencing of a subset of genes, including oestrogen receptor alpha, in response to deacetylase inhibition by valproic acid and trichostatin A. *Oncogene*, 24, 4894-4907

Alan Sawyer

BSc 1988, Paddington College, London, UK.
Work at the Dept. of Molecular Oncology,
Harvard Medical School, USA.
Facility Head at EMBL since 2001.

Monoclonal Antibody Core Facility

The Monoclonal Antibody Core Facility (MACF) was established to provide a service for EMBL researchers and EMBO Young Investigators, while continuously advancing the technology. The mission of our facility is to quickly produce high-affinity antibodies in a high-throughput manner while concentrating on quality of product and of service. The MACF manages over 200 projects per year with an annual theoretical capacity of up to 400 projects. Having developed novel high-throughput techniques and screening assays for the production of mouse-derived, high-affinity monoclonal antibodies, we have recently extended our services to external clients. Current commercial partners include large pharmaceutical companies as well as biotechnology companies.

Services provided

- Consulting on peptide and protein design.
- Producing monoclonal antibodies.
- Characterising antibodies by three separate assays.
- Isotyping of the produced antibodies.
- Scaling up antibody production (10-100 mg level).
- Advising on further characterisation of the produced antibodies.

Technology partners

The Monoclonal Antibody Core Facility works closely with Tecan Italia SA and Genomic Solutions, who provide advanced equipment and top-quality supplies.

Abnova Corporation (Taiwan) is a licensee of the facility's technologies.

Selected references

Taussig, M.J., Stoevesandt, O., Borrebaeck, C.A., Bradbury, A.R., Cahill, D., Cambillau, C., de Daruvar, A., Dubel, S., Eichler, J., Frank, R., Gibson, T.J. *et al*. (2007). ProteomeBinders: planning a European resource of affinity reagents for analysis of the human proteome. *Nat. Methods*, 4, 13-17

Yan, X.J., Albesiano, E., Zanesi, N., Yancopoulos, S., Sawyer, A., Romano, E., Petlickovski, A., Efremov, D.G., Croce, C.M. & Chiorazzi, N. (2006). B cell receptors in TCL1 transgenic mice resemble those of aggressive, treatment-resistant human chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. USA*, 103, 11713-11718

Carissimi, C., Saieva, L., Baccon, J., Chiarella, P., Maiolica, A., Sawyer, A., Rappsilber, J. & Pellizzoni, L. (2006). Gemin8 is a novel component of the survival motor neuron complex and functions in small nuclear ribonucleoprotein assembly. *J. Biol. Chem.*, 281, 8126-8134

De Masi, F., Chiarella, P., Wilhelm, H., Massimi, M., Bullard, B., Ansorge, W. & Sawyer, A. (2005). High throughput production of mouse monoclonal antibodies using antigen microarrays. *Proteomics*, 5, 4070-4081

Facility Head Alan Sawyer and his team work on over 200 projects a year in their Monterotondo facility.

Hüseyin Besir

PhD 2001, Max Planck Institute of Biochemistry, Munich.
Postdoctoral research at Roche Diagnostics, Penzberg, Germany
and the Max Planck Institute of Biochemistry, Munich.
Facility Head at EMBL since 2006.

Protein Expression and Purification Core Facility

Our facility produces and purifies proteins from *E. coli*, insect cells and sera using a variety of chromatographic methods. We can perform biophysical analyses to ensure the quality of each purification in terms of correct folding and stability. Our group also develops and evaluates new techniques or advanced protocols for protein production and purification. Considering the hundreds of proteins expressed and purified every year, most of our efforts are dedicated to develop time-saving solutions. Moreover, we are keeping stocks of a large number of expression vectors and bacterial strains for the users as well as preparing a collection of frequently used enzymes which helps to considerably reduce the expenses of our users.

Major projects and accomplishments

The facility was involved in few development projects. The complexity of the protein aggregation during the recombinant expression has been studied using innovative protocols and the use of a 3D geometry gel. Such a device has also been successfully used for parallel SDS-PAGE separation and analysis of 386 protein samples.

The possibility of using chemical chaperones, alone or in combination with molecular chaperones, to improve the protein solubility has been investigated and molecular reporters developed to follow *in vivo* the effect of the expression parameters on the host cell physiology. Cold adapted bacteria have been tested in comparison with *E. coli* to challenge the folding efficiency of recombinant proteins expressed at 4°C.

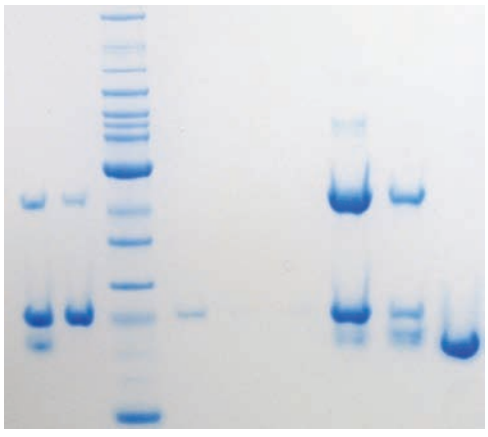
We develope protocols for immortalising polyclonal antibodies, for selecting recombinant antibody expressing cells using permeable membranes, for separating single lines of polyclonals, identifying single epitopes in the antigens and rescuing binders from recombinant antibody libraries.

Services provided

- Expression and purification of proteins in *E. coli* and insect cells;
- Maintaining collections of expression vectors and bacterial strains;
- Producing frequently used enzymes and protein molecular weight marker for general use (TEV protease, 3C protease, Taq, Pfu, T7 RNA pol., LIF, Cre);
- Developing new vectors and protocols;
- Providing access to protocols and vector sequence information on the website;
- Giving scientific and technical advice to users at EMBL and external researchers;
- Caring for equipment for protein production and analysis and cell culture room;
- Providing quality analysis and biophysical characterisation of purified proteins, e.g. analytical ultracentrifugation (AUC) or isothermal titration calorimetry (ITC).

Technology partners

We are open to collaborations with academic or industrial partners to evaluate new products or technological developments that could be helpful for improving the service capabilities of our group. We are frequently approached by company representatives when there are new products to be tested which we try to perform according to requirements.



SDS-PAGE analysis after purification of LIF by ion-exchange chromatography.

Selected references

Huang, H., Liu, J. & de Marco, A. (2006). Induced fit of passenger proteins fused to Archaea maltose binding proteins. *Biochem. Biophys. Res. Commun.*, 344, 25-29

Stegemann, J., Ventzki, R., Martinez, L. & de Marco, A. (2006). Automated protein analysis by online detection of laser-induced fluorescence in slab and 3D-geometry gels. *Electrophoresis*, 27, 3338-48

Schrodel, A., Volz, J. & de Marco, A. (2005). Fusion tags and chaperone co-expression modulate both the solubility and the inclusion body features of the recombinant CLIPB14 serine protease. *J. Biotechnol.*, 120, 2-10

Dummler, A., Lawrence, A.M. & de Marco, A. (2005). Simplified screening for the detection of soluble fusion constructs expressed in *E. coli* using a modular set of vectors. *Microb. Cell Fact.*, 4, 34

Thomas Franz

PhD 1992, Würzburg University, Germany.
Postdoctoral research at Chulabhorn Research Institute, Bangkok, Thailand and at the Laboratory of Organic Chemistry, School of Agriculture, Nagoya University, Japan.
Facility Head at EMBL since 2001.

Proteomic Core Facility

To address the needs of the post genomic era, the Proteomic Core Facility was set up as a cooperation between EMBL and industry. We provide a complete functional proteomic line, including medium-throughput protein and peptide identification, for both in-house service and external visitors.

Major projects and accomplishments

- Completed differential proteomics of wild type and mutant *E. coli* (in cooperation with EMBL Hamburg).
- Identified marker proteins for the rare cerebrovascular disorder Moya Moya, which affects the carotid circulation of the brain (in cooperation with the University Hospital Mannheim).
- Developed new MALDI target coating for direct on-target sample preparation.

Services provided

- Preparative HPLC protein purification.
- IEF with different gradient ranges.
- 2D-gel electrophoresis including colloidal Coomassie, silver or fluorescence staining.
- Laser fluorescence imaging and quantification.
- High-resolution and high-sensitivity imaging densitometer for silver and Coomassie staining.
- PDQuest analysis and evaluation.
- Automatic gel spot excision (fluorescence, silver and Coomassie).
- Full automatic in gel digestion and MALDI target spotting.
- MALDI peptide mass fingerprinting and online database protein identification.
- MS and MS/MS nano electrospray of proteins and peptides.
- Online nano-LC-MS/MS protein identification and CID/ETD analysis of post-translational modifications.

Technology partners

- BIO-RAD – places advanced equipment at the facility's disposal, including the complete Proteome Works System, 2D-gel equipment, spot cutting robot, HPLC, Photometer, Fluorescence Imager and Densitometer.

Selected references

Funk, J., Li, X. & Franz, T. (2005). Threshold values for detergents in protein and peptide samples for mass spectrometry. *Rapid Commun. Mass Spectrom.,* 19, 2986-2988

Li, X., Wilm, M. & Franz, T. (2005). Silicone/graphite coating for on-target desalting and improved peptide mapping performance of matrix-assisted laser desorption/ionization-mass spectrometry targets in proteomic experiments. *Proteomics,* 5, 1460-1471

Schilling, O., Ruggeberg, S., Vogel, A., Rittner, N., Weichert, S., Schmidt, S., Doig, S., Franz, T., Benes, V., Andrews, S.C., Baum, M. & Meyer-Klaucke, W. (2004). Characterization of an *Escherichia coli* elaC deletion mutant. *Biochem. Biophys. Res. Commun.,* 320, 1365-1373

Sola, M., Bavro, V.N., Timmins, J., Franz, T., Ricard-Blum, S., Schoehn, G., Ruigrok, R.W., Paarmann, I., Saiyed, T., O'Sullivan, G.A., Schmitt, B., Betz, H. & Weissenhorn, W. (2004). Structural basis of dynamic glycine receptor clustering by gephyrin. *EMBO J.,* 23, 2510-2519

EMBL Grenoble, France

The EMBL outstation in Grenoble, France, is situated in one of Europe's most beautiful locations, the heart of the French Alps, with a view of snow-covered mountains and the ski slopes. The outstation, a laboratory of about 90 people, shares a campus with the European Synchrotron Radiation Facility (ESRF), which produces some of the world's most intense X-ray beams, and the Institut Laue Langevin (ILL), which provides high-flux neutron beams. The outstation collaborates very closely with these facilities in building and operating beamlines for macromolecular crystallography, in developing the associated instrumentation and techniques and in providing biochemical laboratory facilities and expertise to help external visitors making measurements. The ESRF beamlines are now highly automated and all are equipped with EMBL-designed diffractometers and frozen crystal sample changers.

Within this exciting context, the outstation has a very active in-house research programme in the structural biology of cellular processes, making use of a wide range of techniques including molecular biology, biochemistry, electron microscopy, light scattering, neutron scattering, X-ray crystallography and computing. The availability of such a range of techniques, combined with the neighbouring large-scale facilities, is vital to the success of ambitious projects in modern structural molecular biology.

A strong tradition at the outstation is the study of systems involving protein-nucleic acid complexes and viruses. The structural work on aminoacyl-tRNA synthetases is particularly well known. A number of synthetases were first cloned at EMBL Grenoble and various different synthetase structures have been determined, including several in complex with cognate tRNA. Studies of protein-RNA interactions have been extended to the mammalian signal recognition particle and other proteins and complexes involved in RNA processing, transport and degradation. The analysis of protein-DNA interactions and mechanisms of transcriptional regulation is another important topic. Structural analysis of eukaryotic transcription factor DNA complexes like the first STAT/DNA complex is now moving towards the analysis of larger complexes involved in transcription (e.g. yeast polymerase III) and transcriptional regulation (e.g. chromatin remodelling factors). Recently a group working on the molecular biology of microRNAs has been established.

Another major focus is the study of RNA viruses, such as influenza, rabies and Ebola, with the aim of understanding how they replicate and assemble. In parallel, studies of the structure and function of proteins involved in viral and cellular membrane fusion is actively pursued (e.g. HIV gp41 and proteins and complexes involved in vesicle transport). Some of the projects at the outstation depend on close interactions with colleagues at EMBL Heidelberg and collaborations are underway on proteins involved in nucleocytoplasmic transport, chromatin remodelling and RNA metabolism.

A new development at the outstation is the introduction of automated, high-throughput methods to make structure determination more efficient. This is closely connected to the outstation's involvement in the EU-funded SPINE project (Structural Proteomics in Europe) and its successor (SPINE2-Complexes). A very successful high-throughput robotic system for crystallisation has been implemented, and a novel, high-throughput selection method has been developed for finding soluble protein fragments from otherwise badly expressed or insoluble proteins. These platforms form part of the Partnership for Structural Biology (PSB), which has been established with the neighbouring ESRF, ILL and l'Institut de Biologie Structurale (IBS). The PSB is now housed in a brand new building adjacent to the outstation, together with the CNRS-Grenoble University-EMBL Unit of Virus Host Cell Interactions (UVHCI).

Stephen Cusack
Head of EMBL Grenoble

Stephen Cusack

PhD 1976, Imperial College, London, UK. Postdoctoral work at EMBL Grenoble, France.
Staff scientist at EMBL Grenoble 1980-1989.
Group Leader and Head of Outstation since 1989.
Joint appointment with the Gene Expression Unit.

Structural biology of RNA-protein complexes in gene expression and host-pathogen interactions

Previous and current research

We use X-ray crystallography as a central technique to study the structural biology of protein-RNA complexes involved in RNA metabolism and translation. Additionally we work on the structure of adenovirus capsid proteins involved in host-cell entry, influenza virus polymerase and innate immune system receptors.

In eukaryotic cells, nascent Pol II RNA transcripts (mRNA or snRNA) are rapidly given a m7Gppp cap at the 5' end. The nuclear cap-binding complex (CBC) binds to this tag and mediates interaction with nuclear RNA processing machineries for splicing, poly-adenylation and transport. We determined the structure of human CBC, a 90KDa heterodimeric protein and its complex with a cap analogue and are currently working on structures of several other proteins involved in cap-dependent processes. Once in the cytoplasm, mRNAs are subject to a quality control check to detect premature stop-codons. This process known as nonsense mediated decay (NMD) crucially depends on the three proteins Upf1, Upf2 and Upf3 in all eukaryotic organisms studied, and in mammals, is linked to splicing. We have obtained the first structural information on the interacting domains of these three proteins whose ternary complex formation triggers decay.

Aminoacyl-tRNA synthetases play an essential role in protein synthesis by charging specifically their cognate tRNA(s) with the correct amino acid. We aim to obtain atomic resolution structural information to help us understand the catalytic mechanism of the enzymes and their substrate specificity for ATP, cognate amino acid and tRNA. Most recently we have solved the structures of a class I enzyme, leucyl-tRNA synthetase and a class II enzyme prolyl-tRNA synthetase each with their cognate tRNAs bound. Both these enzymes have the added interest of containing a large inserted editing domain able to recognise and hydrolyse mischarged amino acids; this proof-reading activity is essential for maintaining translational fidelity.

We have been studying the structure of the adenovirus capsid proteins the fibre and penton base which carry respectively primary and secondary receptor binding activity to allow virus entry into cells. We solved the structure of the receptor binding domain and part of the fibrous shaft of Ad2 fibre as well as one domain of the adenovirus receptor (a human cell adhesion protein known as CAR) to which the fibre binds. Recently we have determined structures of complexes of CAR and sialic acid with adenovirus fibres from human Ad37 and canine Cav2. We have determined the structure of the adenovirus penton base, which is at the 12 5-fold vertices of the icosahedral viral capsid, and shown how it interacts with the N-terminal tails of the trimeric fibre. By combining the known crystal structures of the major adenovirus capsid proteins with a 10Å cryo-EM reconstruction of the entire virus, our collaborators and us have obtained a quasi-atomic structure of the whole virus.

Future Projects and Goals

We are currently involved in a number of projects related to RNA metabolism, our goal being to obtain structures of the multi-protein, often transient, complexes involved. These include structural studies on viral capping enzymes, continued work on proteins involved in nonsense mediated decay (NMD) and transport proteins involved in snoRNP assembly and snRNA export. Work is continuing on several aminoacyl-tRNA synthetase systems and their substrate complexes, notably the leucyl- and prolyl-systems, both of which have editing activities. We are also involved in drug design work on synthetases from pathogenic organisms and tropical disease producing nematodes. A major new project has started on the structure of the influenza virus RNA-dependent polymerase. In collaboration with several local groups, we have determined the structure of the C-terminal domain of the PB2 subunit in complex with the human nuclear import receptor, importin alpha 5. We are also working on the structure and mechanism of activation of intracellular pattern recognition receptors of the innate immune system such as the NOD proteins, which respond to fragments of bacterial cell walls and the helicases, which signal interferon production upon detection of viral RNA.

Selected references

Kadlec, J., Guilligay, D., Ravelli, R.B. & Cusack, S. (2006). Crystal structure of the UPF2-interacting domain of nonsense-mediated mRNA decay factor UPF1. *RNA*, 12, 1817-1824

Tukalo, M., Yaremchuk, A., Fukunaga, R., Yokoyama, S. & Cusack, S. (2005). The crystal structure of leucyl-tRNA synthetase complexed with tRNA(Leu) in the post-transfer-editing conformation. *Nat. Struct. Mol. Biol.*, 12, 923-930

Zubieta, C., Schoehn, G., Chroboczek, J. & Cusack, S. (2005). The structure of the human adenovirus 2 penton. *Mol. Cell*, 17, 121-135

Mazza, C., Segref, A., Mattaj, I.W. & Cusack, S. (2002). Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex. *EMBO J.*, 21, 5548-5557

Florent Cipriani

BSc 1974, Physics, University of Grenoble, France.
Senior engineer in nuclear and medical industries.
At EMBL Grenoble since 1991.
Team Leader since 2003.

Diffraction Instrumentation Team

Previous and current research

Our activity is governed by the demands expressed by EMBL crystallographers and the need to maintain the MX beamlines of the EMBL/ESRF Joint Structural Biology Group at the optimum level. In collaboration with the Synchrotron Crystallography Team (page 66), our major themes are the improvement of data quality, the development of instruments for new data collection methods and the development of technologies for beamline automation. We take a particular interest in making our technology available to the scientific community, and most of our developments are licensed to industry in association with EMBLEM. We are especially focused on the user-end of the MX beamlines and develop most of the devices used in experimental hutches: slit-boxes, fast low jitter piezoelectric shutters, diffractometers (MD2x models), air bearing axis, on-beam axis video microscopes, kappa goniometers and automatic sample changers (see figure).

At the ESRF, seven MX beamlines and the BM14 UK CRG beamline have been equipped with MD2x diffractometers and SC3 sample changers, the latter developed and constructed in collaboration with the ESRF and MRC France. The C3D crystal alignment library is permanently upgraded to improve the success rate of automatic crystal centring and to deal with new cryo-loop standards and re-oriented cryo-loops. In addition, a complementary method based on UV induced fluorescence is under evaluation at the ID23-2 and BM14 beamlines.

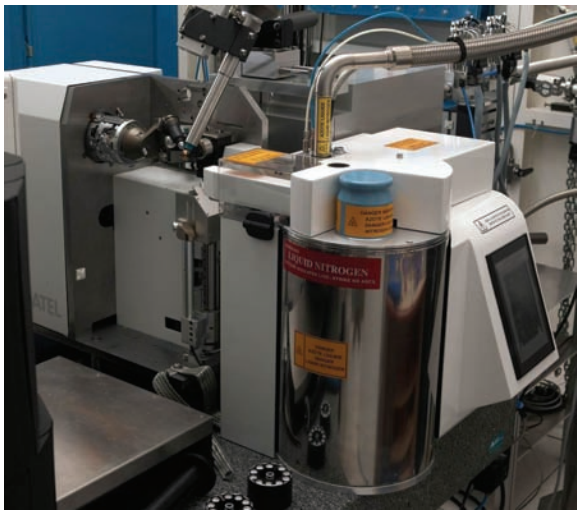
Crystal Reorientation is a method we want to use routinely in a close future to optimise data collection. Several user experiments have been carried on with the four miniKappa goniometer heads installed on the MX MAD beamlines; a new MK3 version with reduced SOC is under evaluation. The STAC software has reached a stage where users can calibrate the device, reorient and re-centre the crystals with minimum effort.

Monitoring crystal radiation damage is essential to guarantee the quality of data collected. Several user experiments have been carried out at the ESRF MX beamline equipped with our integrated micro-spectrophotometer. A single lens and different setups for UV fluorescence are being developed for the MD2 Diffractometer.

Future projects and goals

Our major goals are the improvement of data quality, the automation of the MX beamlines. Several projects will be continued and new ideas explored in collaboration with the Synchrotron Crystallography Team and the Grenoble HTX crystallisation team.

- The improved MK3 version of the miniKappa and associated comprehensive STAC software package will be finalised.
- UV fluorescence crystal centring will be evaluated on the UK CRG ESRF-BM14 beamline. If successful, it will be integrated to C3D
- The UV fluorescence method should be evaluated for automatic detection of crystals in crystallisation trays.
- A project to maximise diffraction from protein crystal in a frozen state or at room temperature will start in 2007. It will be supported by the SPINE 2 European programme.



MD2M diffractometer and SC3 sample changer.

Selected references

Lavault, B., Ravelli, R.B. & Cipriani, F. (2006). C3D: a program for the automated centring of cryocooled crystals. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 1348-1357

Cipriani, F., Felisaz, F., Launer, L., Aksoy, J.S., Caserotto, H., Cusack, S., Dallery, M., di-Chiaro, F., Guijarro, M., Huet, J., Larsen, S., Lentini, M., McCarthy, J., McSweeney, S., Ravelli, R., Renier, M., Taffut, C., Thompson, A., Leonard, G.A. & Walsh, M.A. (2006). Automation of sample mounting for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 1251-1259

Vernede, X., Lavault, B., Ohana, J., Nurizzo, D., Joly, J., Jacquamet, L., Felisaz, F., Cipriani, F. & Bourgeois, D. (2006). UV laser-excited fluorescence as a tool for the visualization of protein crystals mounted in loops. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 253-261

Perrakis, A., Cipriani, F., Castagna, J.C., Claustre, L., Burghammer, M., Riekel, C. & Cusack, S. (1999). Protein microcrystals and the design of a microdiffractometer: current experience and plans at EMBL and ESRF/ID13. *Acta Crystallogr. D Biol. Crystallogr.*, 55 (Pt 10), 1765-1770

Raimond Ravelli

PhD 1998, Chemistry, University of Utrecht, The Netherlands.
At EMBL Grenoble since 1998.
Team Leader since 2003.

Synchrotron Crystallography Team

Previous and current research

Our interests lie in the improvement of synchrotron X-ray macromolecular (MX) data collection techniques to aid structural biologists in their quest for detailed atomic structures used for understanding biological functions of macromolecules. The Synchrotron Crystallography Team forms the Instrumentation Group together with the Diffraction Instrumentation Team (page 65). This group develops hardware, software and novel methodologies for data collection and phasing. Members of the Synchrotron Crystallography Team are involved in collaborations on the cell cytoskeleton, leucine-rich repeat proteins, as well as proteins involved in the invasion of the malaria parasite into human red blood cells.

The instrumentation group works in close collaboration with the MX group (www.esrf.fr/UsersAndScience/Experiments/MX) of the European Synchrotron Radiation Facility (ESRF) on the public MX beamlines. ID23-2 is the world first microfocus beamline fully dedicated to MX, and is operated by David Flot. The undulator MAD beamline ID14-4 is operated by Andrew McCarthy and Raimond Ravelli. The group also provides scientific – through Hassan Belrhali – and technical support for the UK CRG beamline BM14 at the ESRF. The Instrumentation Group includes a number of EMBL Grenoble software engineers, and we coordinate a SciSoft forum that brings them all together.

An ongoing development of methods in crystallography has been the characterisation, mitigation and utilisation of radiation damage in MX. The intense undulator synchrotron radiation of third generation synchrotrons rapidly damages the fragile, cryo-cooled, crystalline macromolecules. Part of our research aims to get an improved understanding as well as a better crystallographic treatment of radiation damage.

Future projects and goals

The present and future work of the group includes:

- Further development of the microfocus beamline ID23-2 with the aim to allow macromolecular crystallographers to collect better data on smaller crystals.
- The optics of ID14-4 will be fully upgraded and commissioned.
- Implementation and exploiting of novel data collection schemes using the inhouse-developed minikappa device.
- Exploring the use of radiation damage for phasing.
- Development and dissemination of spectroscopic techniques complementary to MX.
- Continued contribution to the MAD beamline BM14.

Lighting strikes the ESRF, July 2006, shutting down the ring and most of its instruments. Members of the team stayed till dawn to restart the equipment. Fortunately, most working days are more prosperous, resulting in a continuous flow of novel structures determined by external users and ourselves.

Selected references

Albertini, A.A., Wernimont, A.K., Muziol, T., Ravelli, R.B., Clapier, C.R., Schoehn, G., Weissenhorn, W. & Ruigrok, R.W. (2006). Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science*, 313, 360-363

Cipriani, F., Felisaz, F., Launer, L., Aksoy, J.S., Caserotto, H., Cusack, S., Dallery, M., di-Chiaro, F., Guijarro, M., Huet, J., Larsen, S., Lentini, M., McCarthy, J., McSweeney, S., Ravelli, R., Renier, M., Taffut, C., Thompson, A., Leonard, G.A. & Walsh, M.A. (2006). Automation of sample mounting for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 1251-1259

Nanao, M.H. & Ravelli, R.B.G. (2006). Phasing macromolecular structures with UV-induced structural changes. *Structure*, 14, 791-800

Singh, S.K., Hora, R., Belrhali, H., Chitnis, C.E. & Sharma, A. (2006). Structural basis for Duffy recognition by the malaria parasite Duffy-binding-like domain. *Nature*, 439, 741-744

Darren Hart

PhD 1996, Oxford University, UK.
Postdoctoral research at Cambridge University, UK.
Group Leader at Sense Proteomic Ltd., Cambridge, UK.
Team Leader at EMBL Grenoble since 2003.

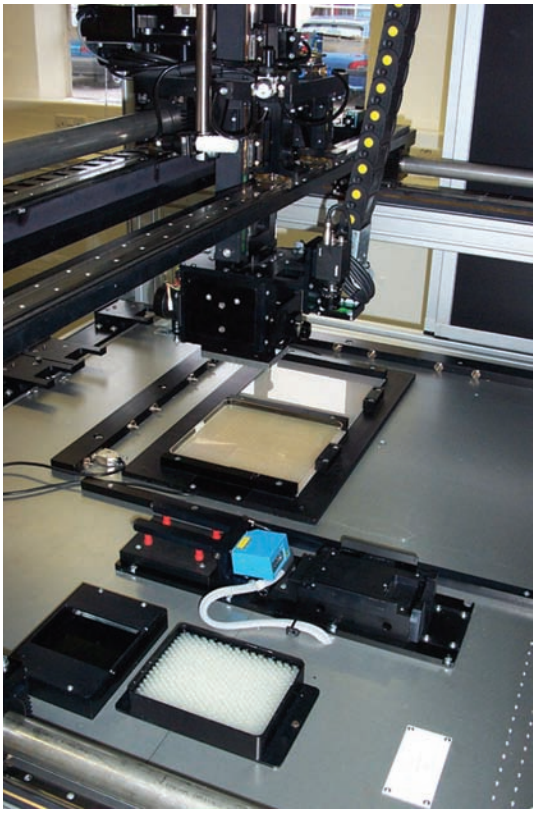
High-throughput protein technologies

Previous and current research

Combinatorial methods in biology (e.g. directed evolution, phage display) are used to address problems that are too complex for rational approaches. First, large random libraries of variants are synthesised in which possible solutions exist at a very low frequency. Second, screening and selection processes are used to isolate the rare hits from the library. Previously, I have used these techniques to explore the function of transcription factors and in the development of functional protein microarrays. At EMBL, we have applied these methods to a common bottleneck of structural biology, that of soluble protein expression. In our high-throughput process, ESPRIT (Expression of Soluble Proteins by Random Incremental Truncation), all truncations of a target protein, are synthesised as a random library. These are then screened using robots to print 'protein arrays' and probed to identify soluble variants. In this way, high value proteins of biological and medicinal importance that have resisted structural analysis due to poor recombinant expression may yield soluble, well-expressed fragments for study by protein crystallography and NMR.

Future projects and goals

We are exploring 'directed evolution' strategies involving gene fragmentation and point mutagenesis for solving the expression problems of previously intractable eukaryotic proteins. As part of the EU Integrated Projects 3D Repertoire and SPINE2, we are adapting the high-throughput robotic methods of structural genomics for the study of protein-protein interactions and applying them to a set of targets. Part of this will involve screening of libraries in eukaryotic systems. As members of the EU project SOUTH, we are using our methods to investigate the targets implicated in lipid metabolism and atherosclerosis, and in FLUPOL, the proteins of influenza. We also work on multidomain human proteins with an industrial collaborator.



A custom-built picking and gridding robot is used to format and analyse large libraries of clones with the aim of identifying members exhibiting improved recombinant protein expression.

Selected references

Tarendeau, F., Boudet, J., Guilligay, D., Mas, P., Bougault, C., Boulo, S., Baudin, F., Ruigrok, R.W.H., Daigle, N., Ellenberg, J., Cusack, S., Simorre, J.-P. & Hart, D.J. (2007). Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat. Struct. Mol. Biol.*, 14, 229-33

Alzari, P.M., Berglund, H., Berrow, N.S., Blagova, E., Busso, D., Cambillau, C., Campanacci, V., Christodoulou, E., Eiler, S., Fogg, M.J. *et al.* (2006). Implementation of semi-automated cloning and prokaryotic expression screening: the impact of SPINE. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 1103-1113

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

Blackburn, J.M. & Hart, D.J. (2005). Fabrication of protein function microarrays for systems-oriented proteomic analysis. *Methods Mol. Biol.*, 310, 197-216

José Márquez

PhD 1997, University of Valencia, Spain.
Postdoctoral research at EMBL.
Staff Scientist at EMBL Grenoble since 2003.
Team Leader since 2005.

The high-throughput crystallisation laboratory at EMBL Grenoble

Previous and current research

Finding conditions in which biological macromolecules would form crystals is recognised as one of the major bottlenecks in structural biology. Once macromolecules are purified they need to be assayed for crystallisation with a collection of precipitants under different chemical environments. This leads to the need to perform hundreds of experiments, consuming large amounts of sample and taking time. At EMBL Grenoble we have established a high-throughput crystallisation platform with the aim to increase the success rate and speed up the process of crystal structure determination. In this platform the whole process of crystallisation screening is automated through the introduction of liquid handling, crystallisation and crystal imaging robots.

The technology introduced allows us to perform experiments using extremely low volumes of sample, which makes it possible to perform extensive screening even when the amount of sample is limited. This platform, which started to operate in September 2003, has now more than two hundred registered users and more one million individual crystallisation experiments have already been performed. The high throughput crystallisation laboratory is not only open to EMBL researchers but also to all the members of the Partnership for Structural Biology (PSB), which includes the ESRF, the ILL, the IBS and the IVMS, and represents one of its core technological platforms.

Future projects and goals

In addition to offering automated crystallisation resources, the HTX lab is actively involved in the development of new methods and concepts in macromolecular crystallography and works in close co-ordination with the high throughput protein expression and synchrotron instrumentation groups at the outstation. One of our major areas of development is data management. We are collaborating with the EBI, EMBL Hamburg and other laboratories in Europe towards the development of a common Laboratory Information Management System (LIMS) for macromolecular crystallography. We are also working in collaboration with the outstation's Instrumentation Group in order to develop strategies to close the gap between crystallisation and data collection by facilitating operations like crystal mounting or freezing that are required before data collection.

We are currently applying high throughput methods to the study of signalling molecules and transcriptional regulators. We have recently solved the structure of the extracellular domain of the human inhibitory receptor IREM-1 expressed in myeloid cells and we are now investigating other members of this receptor family. We are also investigating several members of the transcriptional mediator complex, which is required for activated transcription in Eukaryotes.

Selected references

Ravelli, R.B., Haselmann-Weiss, U., McGeehan, J.E., McCarthy, A.A., Marquez, J.A., Antony, C., Frangakis, A.S. & Stranzl, G. (2007). Plastic-embedded protein crystals. *J. Synchrotron Radiat.*, 14, 128-132

Marquez, J., Reinelt, S., Koch, B., Engelmann, R., Hengstenberg, W. & Scheffzek, K. (2006). Structure of the full-length enzyme I of the phosphoenolpyruvate-dependent sugar phosphotransferase system. *J. Biol. Chem.*, 281, 32508-32515

Hothorn, M., D'Angelo, I., Marquez, J.A., Greiner, S. & Scheffzek, K. (2004). The invertase inhibitor Nt-CIF from tobacco: a highly thermostable four-helix bundle with an unusual N-terminal extension. *J. Mol. Biol.*, 335, 987-995

Marquez, 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

Ramesh Pillai

PhD 2002, University of Bern, Switzerland.
Postdoctoral research at the Friedrich Miescher Institute, Basel, Switzerland.
Group Leader at EMBL Grenoble since 2006.

Regulation of gene expression by non-coding RNAs

Previous and current research

The aim of our research is to understand the molecular mechanisms by which non-coding RNAs, specifically microRNAs (miRNAs), regulate gene expression. MicroRNAs are an abundant class of small non-protein-coding RNAs that function as negative gene regulators at the post-transcriptional level. They are involved in a wide variety of biological processes and it is becoming clear that these tiny RNAs perform critical functions during development and cell differentiation. Recently, mis-expression of miRNAs has been implicated in human cancers, underscoring the relevance of these RNAs in human health. Our recent research has been aimed at determining how miRNAs repress translation of target mRNAs. MicroRNAs act as guides for their associated proteins to bring them to their target mRNAs. They inhibit a step very early in translation and lead to the accumulation of these targets in cytoplasmic structures called P-bodies or processing bodies. MiRNAs can also lead to target mRNA degradation and deadenylation, which could be a consequence of accumulation in P-bodies. Currently, we are trying to understand the role of some of the components of these structures in miRNA-mediated repression.

Future projects and goals

The detailed mechanism by which miRNAs repress translation will be investigated using a combination of genetic, biochemical and cell biology approaches. To complement the bioinformatic predictions of miRNA targets, attempts will be made to biochemically identify endogenous targets of miRNAs with relevance to disease states. Our goal is to understand the mechanism and biology of regulation of gene expression by non-coding RNAs.

Selected references

Pillai, R.S. *et al.* (2007) Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol.*, 17, 118-26.

Schmitter, D. *et al.* (2006). Effects of Dicer and Argonaute down-regulation on mRNA levels in human HEK293 cells. *Nucleic Acids Res.*, 34, 4801-15

Pillai, R.S. *et al.* (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science,* 309, 1573-1576

Pillai, R.S. (2005). MicroRNA function: multiple mechanisms for a tiny RNA? *RNA*, 11, 1753-1761

Pillai, R.S. *et al.* (2004). Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA*, 10, 1518-1525

EMBL Hamburg, Germany

The Hamburg outstation is situated in one of the most beautiful areas of northern Germany's Hanseatic city. Hamburg is centred around its harbour, which has a long tradition of overseas trade, and offers a wide variety of cultural activities. In the close vicinity of EMBL there are diverse neighbourhoods ranging from residential, wealthy suburbs like Blankenese, to areas with a lot of students, entertainment and a multicultural ambience like St. Pauli, Altona and Ottensen.

EMBL Hamburg is located on the site of DESY (the German Synchrotron Research Centre), with synchrotron radiation (DORIS-IIII) and laser (FLASH) facilities available. In addition, the PETRA ring will be converted into a world-leading, dedicated synchrotron radiation facility with operation expected to start in 2009/10 (PETRA-III), and a powerful X-ray Free Electron Laser will be built during the following years. To date, EMBL is operating seven synchrotron radiation beamlines at DORIS-III with applications in life sciences, ranging from biocrystallography to small angle X-ray scattering and X-ray absorption spectroscopy (EXAFS) of biological samples. Since 2006, EMBL has also operated one of the largest high-throughput crystallisation facilities. In addition, infrastructures for expression of protein targets in *E. coli*, *M. smegmatis* and insect cells are available, with an option for high-density cell fermentation. Protein samples can be characterised by isothermal microcalorimetry, mass spectrometry, static and dynamic light scattering and life-time fluorescence spectroscopy. For future applications, EMBL is building a new integrated facility, EMBL@PETRA-III, which will comprise three state-of-the-art beamlines for applications in protein crystallography and small angle X-ray scattering. These beamlines will be complemented with facilities for sample preparation and characterisation and data evaluation.

EMBL Hamburg has a well-established record for the development of novel, innovative technologies in structural biology. World-leading software packages for the automation of data interpretation have been developed and are used in a large number of projects across the world's research community. One is the ARP/wARP package which, in its current version, allows automatic X-ray structure determination with X-ray data higher than 2.5 Å resolution. It is integrated into an automated software pipeline, Auto-Rickshaw, which has also been compiled at EMBL Hamburg. Another package, ATSAS, allows the automatic interpretation of small angle X-ray scattering data for structural shape determination.

Finally, EMBL has set up an ambitious research programme for structures of multifunctional proteins and protein complexes. Most of these projects are driven by questions about biological function and are carried out frequently with collaborators with specific abilities for *in vivo* functional assays and cellular imaging techniques, either from other EMBL Units or external research groups. Present research interests include sarcomeric protein assemblies, protein kinases and phosphatases, viral replication and translation, two-component systems and protein translocation into peroxisomes. In addition, several research groups and teams have joined into a common effort to determine 3D structures from potential drug targets of *Mycobacterium tuberculosis*, with the aim to contribute to the discovery of new drugs against this deadly disease.

Matthias Wilmanns
Head of EMBL Hamburg

Matthias Wilmanns

PhD 1990, University of Basel, Switzerland.
Postdoctoral research at the University of California, Los Angeles, USA.
Group Leader at EMBL Heidelberg 1993-1997.
Head of EMBL Hamburg since 1997.

Structural biology of macromolecular complexes

Our central focus is on the structural characterisation of interactions in networks of biological molecules. We are particularly interested how proteins are regulated either by post-translational modification or by interactions with cellular ligands. We use synchrotron radiation to determine molecular structures at high resolution by X-ray crystallography, complemented by other structural biology methods such as NMR spectroscopy, small angle X-ray scattering, *in vitro* FRET and *in vivo* imaging methods. We are interested in exploiting the opportunities provided by synchrotron and laser facilities at DESY (DORIS-III, PETRA-III, FLASH, X-FEL), developing new methods and applying them to biology-driven projects.

Protein–protein complexes of the muscle sarcomere, including titin, myomesin and binding partners. Titin is the largest gene product of the human genome, comprising up to 38,000 residues in its largest isoform. It is known as the third filament of the muscle sarcomere and is involved in multiple functions, such as acting as a 'molecular ruler' keeping major components of the sarcomere in place, muscle development, passive elasticity of the muscle sarcomere and muscle signalling. We have determined the structure of the N-terminal assembly complex of titin in the Z-disk (Zou *et al.*, 2006), in which two titin filaments are found in an unprecedented antiparallel arrangement, which is mediated by the Z-disk protein telethonin (Fig. 1). We have determined about ten other structures of components or complexes of sarcomeric proteins, and our future work will concentrate on titin-ligand complexes from the sarcomeric Z-disk, the I-band and the M-band.

The architecture of the translocon of peroxisomes. Persoxisome are cell organelles that allow sequestered metabolic processes that would interfere with other processes that, for instance, take place in the cytosol. The proteins that are involved in these processes are generally translocated as active and folded targets. We have been able for the first time to unravel the mechanism of the recognition of peroxisome protein targets by the peroxisome import receptor Pex5p, by determining the structure of the cargo binding domain of the receptor in the absence and presence of the cargo protein sterol carrier protein 2 (Stanley *et al.*, 2006). Our present focus is on structural/functional studies of several other protein components of the peroxisomal translocation machinery.

Structural Proteomics on *Mycobacterium tuberculosis* targets. During the last three years we have determined the X-ray structures of about ten protein targets, some of them with an already known function and others of unknown function. We were able to identity Rv2217 as a novel cysteine/lysine dyad acyltransferase, which allows activation of several important protein complexes by lipoylation (Ma *et al.*, 2006). With the help of the high resolution structures of a total of six protein-ligand complexes, we have been able to unravel the catalytic mechanism of Rv1603 (PriA) as a bifunctional isomerase (Kuper *et al.*, in press) and, by screening a large compound library, we have been able to find an inhibitor that specifically blocks the bifunctional enzyme *in vitro* and *in vivo* (Due *et al.*, in press). Our future interest will focus on the human host/*M. tuberculosis* interactome, with the aim to quantitatively analyse protein-ligand interactions, where the ligands may range from protein, lipids, metabolites and other compounds.

See www.embl-hamburg.de/~wilmanns/home.html for more information.

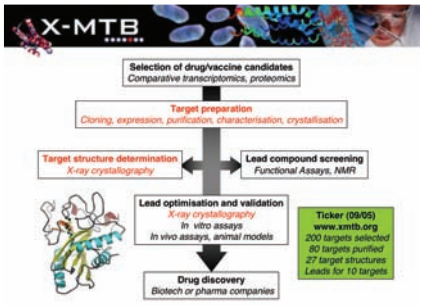


Figure 1: Work flow of the X-MTB structural proteomics project (www.xmtb.org).

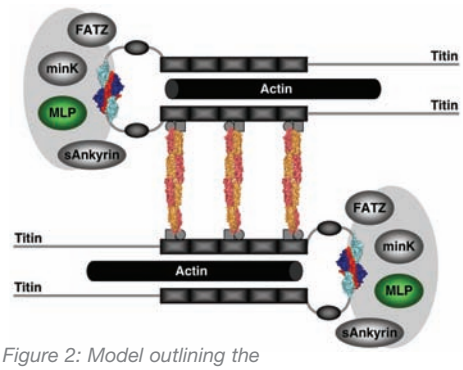


Figure 2: Model outlining the architecture of the sarcomeric Z-disk. Titin filaments are assembled by a dual Z-disk bridging system, by a-actinin rods on a variable number of titin Z-repeats (three bridges are shown), and by telethonin via the N-terminal IG domains Z1 and Z2 (our data).

Selected references

Stanley, W.A., Filipp, F.V., Kursula, P., Schuller, N., Erdmann, R., Schliebs, W., Sattler, M. & Wilmanns, M. (2006). Recognition of a functional peroxisome type 1 target by the dynamic import receptor pex5p. *Mol. Cell*, 24, 653-663

Ma, Q., Zhao, X., Eddine, A.N., Geerlof, A., Li, X., Cronan, J.E., Kaufmann, S.H. & Wilmanns, M. (2006). The *Mycobacterium tuberculosis* LipB enzyme functions as a cysteine/lysine dyad acyltransferase. *Proc. Natl. Acad. Sci. USA*, 103, 8662-8667

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

Christoph Hermes

PhD 1981, Technical University, Munich, Germany.
At EMBL Hamburg since 1981.
Group Leader since 1988.

Instrumentation for structural research at EMBL Hamburg

Previous and current research

Research carried out at the Hamburg outstation since 1974 covers a broad spectrum of X-ray methods used for structural investigations in biology. X-ray absorption spectroscopy (EXAFS) permits high-precision investigation of the environment around specific metal atoms in proteins. Small angle solution scattering (SAXS) allows to study biological macromolecules and their complexes in their native environment while the complete 3D picture can be obtained by protein X-ray crystallography (PX), a method which has become the dominant structural research tool in molecular biology not only at synchrotron radiation sources.

Each of the above-mentioned methods has specific instrumentation needs and our group designs, constructs and builds the appropriate equipment. Our activities include mechanical engineering, vacuum technology, X-ray optics, and data acquisition and control electronics. Recently we rebuilt three beamlines on bending magnet fan K of the DORIS storage ring. Two of the stations for data collection at fixed wavelength from protein crystals have been operating successfully for a couple of years, whereas the third one, which will be used mainly for multi-wavelength anomalous diffraction (MAD) data collection, was opened for external user operation in 2005. Its key component is a fixed exit focusing double monochromator system, which allows an easy and rapid change of the X-ray energy over a wide range. The SAXS station has been up-graded over the last two years by improving the X-ray optics (monochromator and mirror) and the beamline control system (electronics and software). As a pilot project we have installed a sample mounting robot on one of our PX-beamlines following the general trend to increase the level of automation of experiments leading to high-throughput facilities. A system is being developed and tested which should eventually allow remote monitoring and control of the robot and relevant beamline features.

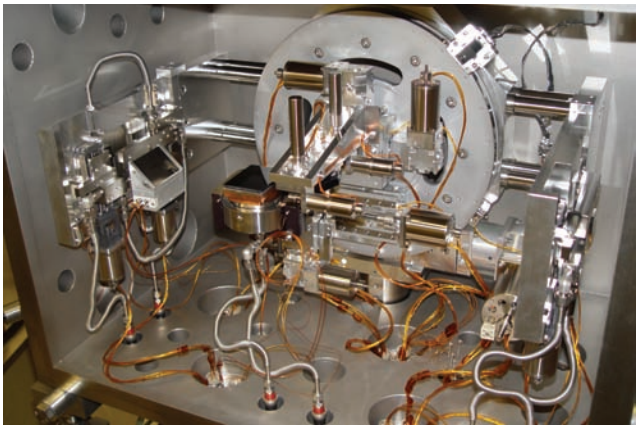
Future projects and goals

We will continue to improve the spectral quality of our existing beamlines, while at the same time we plan to make our lines more 'user-friendly'. This also means that we will continue to increase the level of automation of our experimental stations, which is a necessary condition to perform highthroughput data collection.

In 2007 the transformation of one octant of the PETRA accelerator into a dedicated synchrotron radiation source will start, and EMBL will build and operate beamlines on this unique radiation source. In this context major challenges and opportunities in the field of beamline instrumentation are coming up.

The opening of the European X-Ray Laser (XFEL) in Hamburg is scheduled for 2013, offering unprecedented research opportunities. To design experiments which will exploit the potential of this unique facility requires a large number of problems in various areas to be addressed and solved.

Our goal is to create optimal conditions for state-of-the-art experiments in structural biology at a modern synchrotron radiation source.



The double monochromator set-up of the MAD beamline X12.

Selected references

Hermes, C. Gehrmann, T., Jahn, D., Ristau U., Robrahn, B. & Siambanis, T. (2004). New beamlines for protein crystallography at the EMBL Outstation Hamburg. *AIP Conference Proceedings*, 705, 384-388

Pohl, E., Ristau, U., Gehrmann, T., Jahn, D., Robrahn, B., Malthan, D., Dobler, H. & Hermes, C. (2004). Automation of the EMBL Hamburg protein crystallography beamline BW7B. *J. Synchrotron Radiat.,* 11, 372-377

Victor S. Lamzin

PhD 1987, Moscow State University, Russia.
Scientist, Inst. Biochemistry, Russian Academy of Sciences, Moscow, until 1991.
Postdoctoral research at EMBL Hamburg until 1995; Staff Scientist until 1997.
Group Leader and Deputy Head of EMBL Hamburg since 1997.

Protein crystallography

Macromolecular X-ray crystallography provides the majority of atomic models in the Protein Data Bank and is nowadays the most commonly used technique for structure determination, in the context of structural proteomics and structural biology in general. The major activity of the group is the development of novel approaches and software tools for high throughput 3D structure determination and detailed interpretation of biological macromolecules and their complexes.

Previous and current research

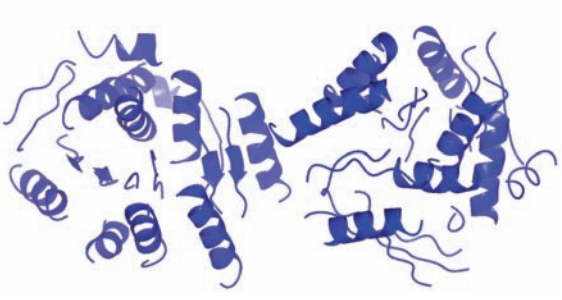
An ongoing development is the ARP/wARP software suite for refinement and modelling of protein structures (in collaboration with the Perrakis group at the NKI, Amsterdam). Owing to its continuous advancement, ARP/wARP has become the method of choice in macromolecular crystallography when the data extend to sufficient resolution. Highlights of 2006 include automated modelling of secondary structural elements at resolution below 3.0 Å (see figure), provision of web-based computational services for remote execution of crystallographic structure determination, and enhanced modelling of bound ligands with the new functionality to screen a cocktail for a likely ligand candidate at a known or even unknown location. The first encouraging results were obtained for the recognition of DNA/RNA planar bases in electron density maps, which is becoming increasingly important, e.g. in the context of studies of mechanisms of gene expression and transcription regulation.

The automation and the enhancement of the crystallographic experiment continue to be of special interest. Considerable progress has been achieved with software XREC for automated recognition and centring of crystalline samples on a beamline goniostat (in cooperation with colleagues at EMBL Grenoble). XREC has a success rate of about 75%, and is now used at several synchrotron sites worldwide. The software BEST (in collaboration with Popov and Bourenkov) proved to be very efficient in collecting the X-ray data from a crystal. Recently implemented modelling of the radiation-induced intensity variation will be particularly valuable for weakly diffracting crystals and large unit cells. A related development is the radiation damage protection of the crystals in the X-ray beam by quick-soaking with radical scavengers (in collaboration with the Weiss team, page 79). Furthermore, we have succeeded in the exploitation of atomic resolution protein structures for the extraction of the protonation state in active sites, unusual substrate binding and functionally relevant directional atomic motion.

In collaboration with S. Panjikar from the Tucker group (page 78) and the Weiss team we are developing the automated structure determination pipeline AutoRickshaw, which utilises a complex underlying decision-making system. With the help of ARP/wARP and other software, AutoRickshaw can produce an interpretable electron density map and a partial structure shortly after data acquisition; consequently, a synchrotron beamline user can receive immediate feedback about whether the measured data are of good enough quality for successful structure determination.

Future projects and goals

In the future we will witness the further integration of macromolecular crystallography, and the underlying methodology will face a number of important challenges. Crystalline samples will be obtained from ever more complex biological systems and will be smaller and more difficult to handle, with data harder to extract. Along with this, more state-of-the-art synchrotron beamlines will become available, including those at the PETRA ring at DESY. Already on the horizon are the first coherent X-ray sources, which will allow experiments to be carried out that are simply impossible today. The activity of the group will be cast within these anticipated trends.



Automatically built helices and strands in an electron density map for a 662-residue long protein with ARP/wARP at 3.2 Å resolution. The task took 1 minute on a modern workstation.

Selected references

Kauffmann, B., Weiss, M.S., Lamzin, V.S. & Schmidt, A. (2006) How to avoid premature decay of your macromolecular crystal – a quick soak for long life. *Structure,* 14, 1099-1105

Pothineni, S.B., Strutz, T. & Lamzin, V.S. (2006) Automated detection and centring of cryo-cooled protein crystals. *Acta Crystallogr. D Biol. Crystallogr.,* 62, 1358-1368

Schmidt, A. & Lamzin, V.S. (2005). Extraction of functional motion in trypsin crystal structures. *Acta Crystallogr. D Biol. Crystallogr.,* 61, 1132-1139

Zwart, P.H., Langer, G.G. & Lamzin, V.S. (2004). Modelling bound ligands in protein crystal structures. *Acta Crystallogr. D Biol. Crystallogr.,* 60, 2230-2239

Wolfram Meyer-Klaucke

PhD 1996, Hamburg University, Germany.
Postdoctoral research at the Medical University of Lübeck, Germany.
At EMBL Hamburg since 1998. Team Leader since 2000.

Trace elements in biological systems

Previous and current research

Trace elements such as metals play a key role in the structure and function of about 30% of all proteins. Many biocatalytic processes depend on the presence of metal ions. Our research deals with metal functionality, binding and selectivity in biological systems. The group's projects combine structural techniques with molecular biology, biochemistry and further methods aiming at a complete understanding of metal-related biological processes. Apart from methods development (Korbas, 2006), current research includes:

Proteins of the metallo-β-lactamase superfamily. This superfamily, with an active site capable of binding up to two metal ions, catalyses a variety of enzymatic processes. Beside the global metal binding motif the overall fold of α-sheets and β-helices is conserved within the superfamily:

- zinc dependent enzymes (class B lactamases, zinc phosphodiesterase (ZiPD));
- enzymes flexible in metal used (Glyoxalase II);
- iron dependent redox enzymes (rubredoxin oxygen oxidoreductase (ROO));
- yet uncharacterised members.

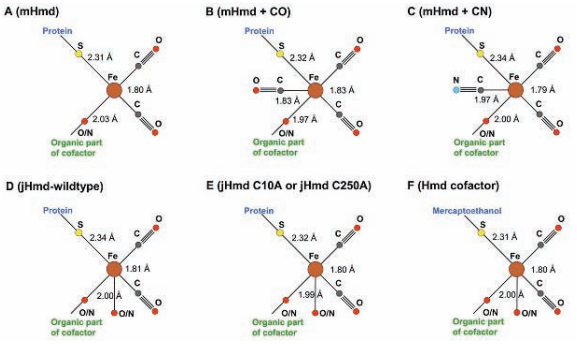
Their physiological importance varies, from putative association with cancer and antibiotic resistance to different roles in cellular detoxification.

An exosite unique to the metallo-b-lactamase superfamily is required for substrate binding in ZiPD: To investigate the function of an exosite unique in the metallo-β-lactamase superfamily we constructed a ZiPD variant that lacks this module, ZiPDΔ. This has almost wild-type-like catalytic properties for hydrolysis of a chromogenic substrate. Hence, the exosite is not required for the intrinsic phosphodiesterase activity of ZiPD. In contrast to the wild-type enzyme, ZiPDΔ does not process pre-tRNA, and gel shift assays demonstrate that only the wild-type enzyme, but not ZiPDΔ, binds mature tRNA. These findings show that the exosite is essential for pre-tRNA recognition. In conclusion, we identify a ZiPD exosite that guides physiological substrate recognition in the ZiPD/ElaC protein family. This was further substantiated by the proteins crystal structure (Kostelecky, 2006).

New metal binding motifs: Hydrogenases are enzymes that catalyze the reversible oxidation of molecular hydrogen. Their structure and catalytic mechanism are of considerable applied interest as models for the development of efficient catalysts for hydrogen-fueled processes. Despite intensive efforts, however, the understanding of how hydrogenases react with H_2 is only in its infancy. The only mononuclear hydrogenase, Hmd, harbours an iron containing cofactor of yet unknown structure. X-ray absorption spectroscopy determined two CO, one sulfur, and two nitrogen/oxygen atoms coordinated to the iron, the sulfur ligand being most probably provided by the protein. In active Hmd holoenzyme, the number of iron ligands increased by one when one of the Hmd inhibitors (CO or KCN) were present, indicating that in active Hmd, the iron contains an open coordination site, which is proposed to be the site of H_2 interaction (Korbas, 2006).

Future projects and goals

In addition to the metal specificity of proteins we will focus on the regulation of metal concentrations in cells. At present we combine structural analysis (e.g. XAS, SAXS, protein crystallography) and biochemical methods (element analysis, isothermal calorimetry, enzyme kinetics). In future spatial resolved spectroscopy and advanced spectroscopic techniques will play an increasing role.



Structural models for the iron site in mononuclear hydrogenase under different biochemical conditions.

Selected references

Korbas, M., Vogt, S., Meyer-Klaucke, W., Bill, E., Lyon, E.J., Thauer, R.K. & Shima, S. (2006). The iron-sulfur cluster-free hydrogenase (Hmd) is a metalloenzyme with a novel iron binding motif. *J. Biol Chem.*, 281, 30804-13

Boyault, C., Gilquin, B., Zhang, Y., Rybin, V., Garman, E., Meyer-Klaucke, W., Matthias, P., Muller, C.W. & Khochbin, S. (2006). HDAC6-p97/VCP controlled polyubiquitin chain turnover. *EMBO J.*, 25, 3357-3366

Kostelecky, B., Pohl, E., Vogel, A., Schilling, O. & Meyer-Klaucke, W. (2006). The crystal structure of the zinc phosphodiesterase from *Escherichia coli* provides Insight into function and cooperativity of tRNase Z-family proteins. *J. Bacteriol.*, 188, 1607-1614

Korbas, M., Marsa, D.F. & Meyer-Klaucke, W. (2006). A program script for automated biological X-ray absorption spectroscopy data reduction. *Review of Scientific Instruments*, 77

Jochen Müller-Dieckmann

PhD 1994, Albert-Ludwigs-Universität, Freiburg, Germany.
Postdoctoral research at the University of California, Berkeley, USA.
Associate Director, SGX, San Diego, USA until 2004.
Team Leader at EMBL Hamburg since 2004.

High-throughput crystallisation

A high-throughput crystallisation facility capable of performing all steps of a crystallisation experiment, including the preparation of crystallisation cocktails from stock solutions, the set up of experiments with nano drops, the imaging of experiments and remote access to data, has been established. This facility is the largest in Europe which is open to the general user community and it has the explicit goal to offer users as many options as possible in the design of their experiments, rather than enforce compliance with pre-formulated protocols. The following options are currently available to users during on-line registration of their experiments on our electronic booking system (www.embl-hamburg.de/htpx/2007/):

- experimental set up as vapour diffusion – sitting drop or as free interface diffusion in micro-fluidic chips (Topaz from Fluidigm Inc.). The latter needs just 1µL of sample for 96 experiments;
- initial screening (~1,000 different conditions) or set up from customised screens;
- drop volume of 200 to 1,500nL (in 100nL steps) for sitting drops;
- free ratio of reservoir to drop volume (in sitting drop);
- two different crystallisation plate types.

Development of a new crystallisation device. Based on micro-fluidic technologies, the system consists of a CD-like crystallisation chip and a console to load the chip and initiate the experiment. We were able to demonstrate the ability to grow crystals in the chip from as little as 5nL of sample for all four test proteins (see figure below). Crystals grown on the chip can be analysed *in situ* in the X-ray beam. This allows the determination of the intrinsic quality of crystals prior to potentially detrimental handling. The console itself, being very compact (~20x20x20cm) and working without the use of pumps or pressure, is very easy to operate.

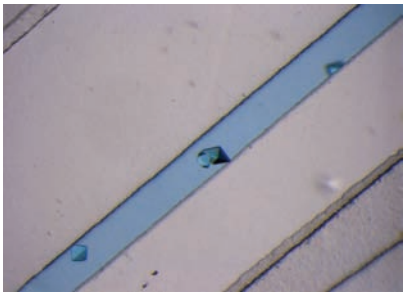
Proteins from Mycobacterium tuberculosis. Tuberculosis remains the single most infectious disease, killing over 2 million individuals every year. The disease is curable; treatment, however, is prolonged and requires a combination of several drugs. At the same time, multidrug-resistant strains of *Mycobacterium tuberculosis* (Mtb), the bacilli that cause TB, have been detected in almost all 109 countries surveyed in 2006.

The German X-MTB consortium focuses on a subset of validated gene products which were identified by comparing the expression and transcription profiles of Mtb during different life cycles or under different growth conditions. These differences often indicate that the corresponding proteins are involved in and important for the persistence or pathogenicity of Mtb. We originally started with twelve targets from the X-MTB list of proteins. Targets have been deleted and added over time to accommodate the scientific development.

We have solved and refined the structure of IF-1 to 1.5 Å resolution. The structure of IF-1 (Rv3462c) is the first X-ray structure of an un-complexed initiation factor 1 (there is a NMR structure of free IF-1 from *E. coli*). The only X-ray structure of an IF-1 is in complex with the 30S ribosome from *Thermus thermophilus*.

Future projects and goals

The importance of construct design on protein expression, solubility and crystallisability has been widely accepted. We have established a collaboration with Fluidigm and Darren Hart's group (page 67) in order to demonstrate that it is possible to rapidly screen a large number (~40) of protein constructs and to identify those constructs which are more amenable to crystallisation using the Fluidigm technology. The transfer of crystallisation conditions from free interface diffusion experiments with pico-litre volumes to vapour diffusion experiments with nano-litre volumes is a known obstacle which we will try to characterise and hopefully overcome within this collaboration.



Proteinase K crystal in reaction chamber of newly-developed crystallisation device. The crystal shown is about 100x30x30µm3.

Selected references

Müller-Dieckmann, J. (2006). The open-access high-throughput crystallization facility at EMBL Hamburg. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 1446-1452

Kernstock, S., Koch-Nolte, F., Müller-Dieckmann, J., Weiss, M.S. & Müller-Dieckmann, C. (2006). Cloning, expression, purification, crystallization and preliminary X-ray diffraction analysis of human ARH3, the first eukaryotic protein-ADP-ribosylhydrolase. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.*, 62, 224-227

Thomas Schneider

PhD 1996, Technical University of Munich and EMBL, Germany
Postdoctoral research at the Max-Planck-Institute for Molecular Physiology, Dortmund, Germany.
Researcher at the University of Göttingen, Germany.
Group Leader at the FIRC Institute of Molecular Oncology, Milan, Italy.
Group Leader at EMBL since 2007. Coordinator of the EMBL@PETRA-III project.

Tools for structure determination and analysis

Previous and current research

The group pursues two major activities: (1) the construction of three beamlines for structural biology at the new PETRA III synchrotron in Hamburg; and (2) the development of computational methods to extract the information from large amounts of structural data.

The three beamlines we are designing will harness the extremely brilliant beam of the PETRA III synchrotron for small angle X-ray scattering on solutions and X-ray crystallography on crystals of biological macromolecules. The beamlines will be embedded in an integrated facility for structural biology (www.embl-hamburg.de/services/petra). This facility will support non-specialists not only in performing the actual experiments with synchrotron radiation but also in sample preparation and the evaluation of the measured data. The design of the beamlines is done in close collaboration with the team of Stefan Fiedler.

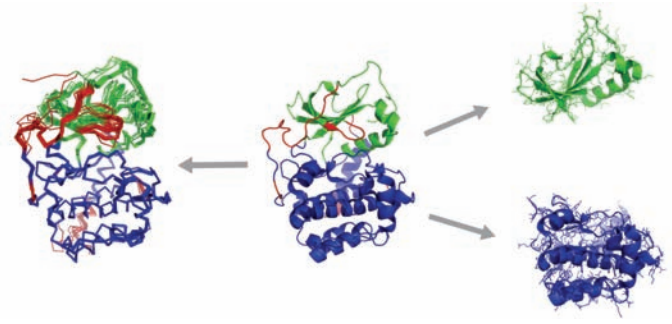
Partly due to the enormous progress in synchrotron radiation-based structural biology, structural data on biological macromolecules are produced at an ever-increasing rate, creating the need to develop tools for efficient mining of structural data. We are creating tools for which the central concept is to use coordinate errors throughout all calculations. The necessity of this approach becomes clear when one considers that in the contrast to sequence data where a nucleotide entry can only be right or wrong, the precision in the location of an atom in a crystal structure can vary over several orders of magnitude; while the position of an atom in a rigid region of a protein giving diffraction data to high resolution may be known to within 0.01 Å, for an atom in a flexible region of a poorly diffracting protein, the coordinate error may reach more than 1.0 Å.

From a technical point of view, extracting information from large amounts of raw structural data (up to hundreds of structures containing thousands of atoms each) is a very complex task and requires sophisticated algorithms both for the analysis and for the presentation and 3D visualisation of the results. During the last few years, we have been implementing various algorithms in a framework for the analysis of different conformations of the same molecule. Presently, we are expanding the scope of the methods to the investigation of homologous structures.

Future projects and goals

For the integrated facility for structural biology, our goal is to provide beamlines that are ready for user experiments by early 2010. In small-angle X-ray scattering, the new beamlines will enable us to work with more complex and more dilute samples than presently possible. In macromolecular crystallography, the beamlines will provide features such as micro-focusing and energy tunability, allowing imaging of the content of small crystals containing large objects such as multi-component complexes.

On the computational side, we will work on improving the error models underlying our methods and on expanding our computational framework using genetic and graph-based algorithms. We also plan to use recurrent structural fragments extracted from ensembles of structures as search models in molecular replacement and for the interpretation of low resolution electron density maps. In fact, this aspect of our computational work will be very helpful in the interpretation of diffraction experiments on weakly diffracting large systems on the future PETRA III beamlines.



Rigid regions (centre; blue and green) identified from ensembles of structures can be used for superposition and subsequent analysis (left) or as fragments to interpret experimental data from methods with lower resolution than X-ray crystallography (right).

Selected references

Schneider, T.R. (2004). Domain identification by iterative analysis of error-scaled difference distance matrices. *Acta Cryst. D.*, 60, 2269-2275

Penengo, L., Mapelli, M., Murachelli, A., Confalonieri, S., Magri, L., Mustacchio A, Di Fiore, PP, Polo, S. & Schneider. T.R. (2006). Crystal structure of the ubiquitin binding domains of Rabex-5 reveals new modalities of interaction with ubiquitin. *Cell*, 124, 1183-1195

Dall'Antonia, F. & Schneider, T.R. (2006). SITCOM: A program for comparing sites in macromolecular structures. *J. Appl. Cryst.*, 39, 618-619

Dmitri Svergun

PhD 1982, Institute of Crystallography, Moscow, Russia.
Dr. of Science 1997, Institute of Crystallography, Moscow, Russia.
At EMBL since 1991.

Development and application of new methods for X-ray and neutron solution scattering data analysis

Previous and current research

Fundamental biological processes, such as cell-cycle control, signalling, DNA duplication, gene expression and regulation and some metabolic pathways, depend on supra-molecular assemblies and their changes over time. There are objective difficulties in studying such complex systems, especially their dynamic changes, with high resolution structural techniques like X-ray crystallography or NMR.

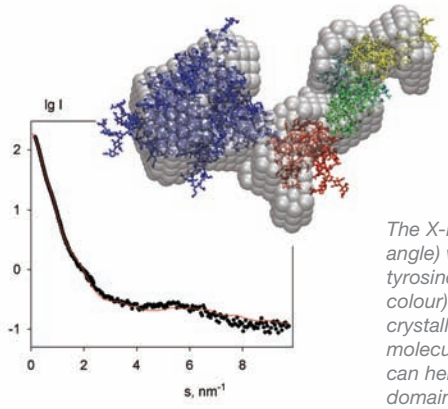
Small-angle X-ray scattering (SAXS) allows us to study native biological macromolecules, from individual proteins to large complexes, in solution under nearly physiological conditions. SAXS not only provides low resolution 3D models of particle shapes but yields answers to important functional questions. In particular, SAXS elucidates structural changes in response to variations in external conditions, protein-protein and protein-ligand interactions, and time-resolved studies elucidate structural kinetics of assembly/dissociation or folding/unfolding.

Our group runs the EMBL SAXS beamline X33 at the storage ring DORIS-III of synchrotron DESY, which has recently been upgraded in response to the growing demand for SAXS in biological community. The group develops novel methods to construct 3D models of individual macromolecules and their complexes from the X-ray and neutron scattering data. Advanced mathematical approaches (regularisation, non-linear optimisation, heuristic algorithms, neural networks, bioinformatics etc.) are employed. Most of the ongoing projects of the external user groups at the SAXS beamline are collaborative projects, whereby the members of the group are applying the new methods to solve biological problems using SAXS. Special emphasis is put to the joint use of the results of X-ray crystallography, NMR and EM with SAXS data (see figure).

Future projects and goals

The present and future work of the group includes:

- development of algorithms for *ab initio* analysis of the tertiary and quaternary structure of macromolecules, in particular of flexible and intrinsically disordered proteins, from X-ray and neutron scattering data;
- methods for rigid-body modelling (docking) of macromolecular complexes using high-resolution structures of individual domains from crystallography or NMR;
- automation of SAXS experiments and development of an integrated data analysis pipeline;
- maintenance and upgrade of the existing X33 beamline and collaborative user projects;
- collaboration with the Petra-3 group at EMBL-Hamburg in designing of a new high-brilliance biological SAXS beamline at the third-generation PETRA storage ring at DESY.



The X-ray scattering curve on the left (intensity versus scattering angle) was used to create a low-resolution model of receptor tyrosine kinase MET, shown in grey. Superimposed on this (in colour) are atomic models of separate domains obtained from crystallography and positioned to fit the SAXS data. The protein molecule is about 10 nm across. This example illustrates how SAXS can help to assemble together high resolution models of individual domains into the model of the entire macromolecular complex.

Selected references

Gherardi, E., Sandin, S., Petoukhov, M.V., Finch, J., Youles, M.E., Ofverstedt, L.G., Miguel, R.N., Blundell, T.L., Vande Woude, G.F., Skoglund, U. & Svergun, D.I. (2006). Structural basis of hepatocyte growth factor/scatter factor and MET signalling. *Proc. Natl. Acad. Sci. USA*, 103, 4046-4051

Konarev PV, Petoukhov MV, Volkov VV, Svergun DI (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. *Journal of Applied Crystallography*, 39, 277-286

Petoukhov, M.V., Monie, T.P., Allain, F.H., Matthews, S., Curry, S. & Svergun, D.I. (2006). Conformation of polypyrimidine tract binding protein in solution. *Structure*, 14, 1021-1027

Petoukhov, M.V. & Svergun, D.I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys. J.*, 89, 1237-1250

Paul Tucker

PhD 1972, University of Essex, Colchester, UK.
Postdoctoral research at the University of Groningen, the Netherlands, University of Leicester, UK, Australian National University, Canberra and University of Exeter, UK.
Lecturer in Inorganic Chemistry, University of the West Indies.
At EMBL Heidelberg 1985-1998; at EMBL Hamburg since 1998.

Structural studies of proteins from pathogens

Previous and Current Research

A major focus has been trying to understand, at a structural level, the mechanisms by which bacteria respond to external conditions. We are therefore interested in classical two-component systems and the way they regulate gene expression by responding to the environment. *Mycobacterium tuberculosis* (MtB) is our organism of choice by virtue of its threat to human health. We have determined the structures of several response regulators (Figure 1) and have also identified a new type of two-component system in MtB that acts at the level of transcriptional anti-termination rather than transcriptional activation/repression. We have also been studying an alternative sigma factor in MtB and they way that its availability is regulated by the anti-sigma factor.

A second area of study has been on proteins that are involved in replication of the genome of a variety of RNA viruses. The current work is to obtain structural information on the RNA dependent RNA polymerases of the caliciviridae (Figure 2), the flaviviridae and, more recently a vesiculovirus. We have also continued our work on the structural and functional aspects of single-stranded DNA binding proteins found in the dsDNA viruses, and are interested in how these proteins interact with other components of the replication machinery.

A third focus, largely with Staff Scientist Santosh Panjikar and the Weiss team (opposite), remains on improving the structure determination process by, for example, developing improved phasing methods and automating structure determination. Santosh Panjikar is also actively engaged in elucidated the enzymes of the biosynthetic pathway of various indole alkaloids in Indian medicinal plants.

Future Projects and Goals

As part of a major EU-funded Integrated Project (VIZIER) and we are also expanding our work on viral replication proteins to structural work on proteins of unknown function but which seem essential for viral replication. This includes, for example work on domains of the coronaviridae NSP3 protein.

We have also begun a project to obtain structural information on the lipid binding proteins of parasitic nematodes, because these proteins are essential for the viability of the nematode in the host.

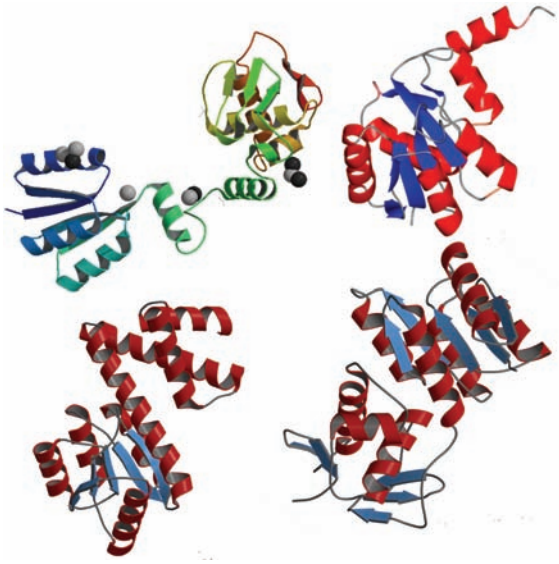


Figure 1: The structures of the two-component system response regulators of Mycobacterium tuberculosis *that were solved recently in the group. Shown from top left going clockwise are proteins with genomic location tags Rv0491, Rv0844, Rv0903c and Rv1626.*

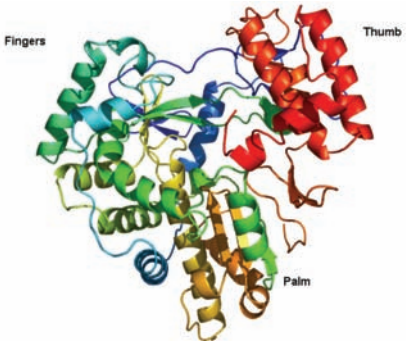


Figure 2: The structure of the RNA dependent RNA polymerase of saporovirus.

Selected references

Fullerton, S.W., Blaschke, M., Coutard, B., Gebhardt, J., Gorbalenya, A., Canard, B., Tucker, P.A. & Rohayem, J. (2007). Structural and functional characterization of sapovirus RNA-dependent RNA polymerase. *J Virol.*, 81, 1858-71

Nowak, E., Panjikar, S., Morth, J.P., Jordanova, R., Svergun, D.I. & Tucker, P.A. (2006). Structural and functional aspects of the sensor histidine kinase PrrB from *Mycobacterium tuberculosis*. *Structure*, 14, 275-285

Mapelli, M., Panjikar, S. & Tucker, P.A. (2005). The crystal structure of the herpes simplex virus 1 ssDNA-binding protein suggests the structural basis for flexible, cooperative single-stranded DNA binding. *J. Biol. Chem.*, 280, 2990-2997

Morth, J.P., Gosmann, S., Nowak, E. & Tucker, P.A. (2005). A novel two-component system found in *Mycobacterium tuberculosis*. *FEBS Lett.*, 579, 4145-4148

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

Manfred Weiss

PhD 1992, Albert-Ludwigs-Universität, Freiburg, Germany.
Postdoctoral research at the University of California, Los Angeles, USA.
Senior Research Assistant at the IMB Jena.
Team Leader at EMBL Hamburg since 2001.

X-ray crystallography of biological macromolecules

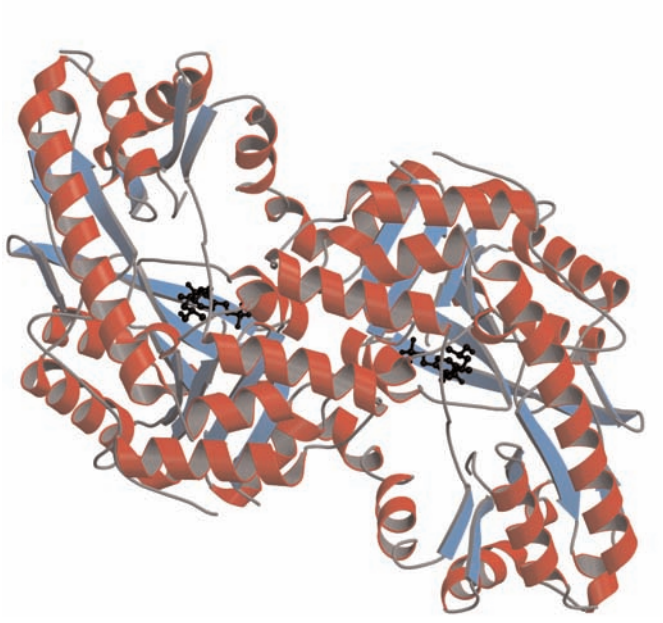
Previous and current research

Structural analysis of proteins from *Mycobacterium tuberculosis*. Within the Hamburg group of the X-Mtb consortium (www.xmtb.org) a total of about 240 proteins from *M. tuberculosis* are currently under investigation. In our team the focus has been placed on proteins from the leucine and lysine biosynthesis pathways of this organism. Recently, we have been able to determine the structures of five enzymes from *M. tuberculosis*: one enzyme from the leucine biosynthesis and four from the lysine biosynthesis pathway. As an example the dimeric structure of N-succinyldiaminopimelate aminotransferase (Rv0858c) is depicted below. We are currently in the process of co-crystallising these enzymes with substrates, cofactors, inhibitors, etc. in order to better understand the metabolic pathways in *M. tuberculosis*. These structures will also provide the basis for the design of new drugs. In addition, we have determined the structure of a hypothetical protein from *M. tuberculosis*, the function of which is currently under investigation.

Development of new methods for macromolecular structure determination. In collaboration with the group of Paul Tucker (opposite) and within the frame of a project funded by the DFG we are investigating the use of longer X-ray wavelengths in macromolecular crystallography. More specifically, we are interested in determining phases based on the very small anomalous signal provided by sulphur and phosphorus atoms, which are naturally present in proteins and nucleic acids. Recently, we have been able to establish that almost independent of the protein and the anomalous substructure the wavelength at which the largest anomalous signal-to-ratio can be obtained is about 2.0 Å. We are now in the process of extending this work towards a general strategy of phase determination from weak anomalous signals.

Future projects and goals

The structure analysis of proteins from *M. tuberculosis* will be continued. We hope to be able to completely structurally characterise the two pathways. Also, we will lay further emphasis on complex formation between enzymes catalysing adjacent reactions in the pathways. In cases where protein-protein complexes exist, we will try to isolate and crystallise them. In cases where only transient complex formations occurs, we will try to characterise them by different means. Another important aspect in the *M. tuberculosis* project is the search for small molecules which bind to the proteins and which, in the case of enzymes, have the capacity to inhibit them. This work will be carried out in collaboration with the screening facility led by Jens-Peter von Kries at the Forschungsinstitut für Molekulare Pharmakologie (FMP) in Berlin.



Schematic representation of the dimer of Rv0858c from M. tuberculosis*. The view is down the twofold axis. α-helices are represented in red and β-strands in light blue. The PLP-moiety bound to Lys232 is shown as a ball-and-stick model in black.*

Selected references

Müller-Dieckmann, C., Kernstock, S., Lisurek, M., von Kries, J.P., Haag, F., Weiss, M.S. & Koch-Nolte, F. (2006). The structure of human ADP-ribosylhydrolase 3 (ARH3) provides insights into the reversibility of protein ADP-ribosylation. *Proc. Natl. Acad. Sci. USA*, 103, 15026-15031

Kauffmann, B., Weiss, M.S., Lamzin, V.S. & Schmidt, A. (2006). How to avoid premature decay of your macromolecular crystal: a quick soak for long life. *Structure*, 14, 1099-1105

Müller-Dieckmann, C., Panjikar, S., Tucker, P.A. & Weiss, M.S. (2005). On the routine use of soft X-rays in macromolecular crystallography. Part III. The optimal data-collection wavelength. *Acta Crystallogr. D Biol. Crystallogr.*, 61, 1263-1272

Singh, R.K., Kefala, G., Janowski, R., Mueller-Dieckmann, C., von Kries, J.P. & Weiss, M.S. (2005). The high-resolution structure of LeuB (Rv2995c) from *Mycobacterium tuberculosis*. *J. Mol. Biol.*, 346, 1-11

EMBL-EBI, Hinxton, UK

The European Bioinformatics Institute (EMBL-EBI) lies in the 55 acres of landscaped parkland in rural Cambridgeshire that make up the Wellcome Trust Genome Campus, which also houses the Wellcome Trust Sanger Institute. Together, these institutes provide one of the world's largest concentrations of expertise in genomics and bioinformatics.

EMBL-EBI has a fourfold mission:

- to provide freely available data and bioinformatics services to all facets of the scientific community in ways that promote scientific progress;
- to contribute to the advancement of biology through basic investigator-driven research in bioinformatics;
- to provide advanced bioinformatics training to scientists at all levels, from PhD students to independent investigators;
- to help disseminate cutting-edge technologies to industry.

As a hub of bioinformatics in Europe, EMBL-EBI provides data resources in all the major molecular domains. The EBI grew out of EMBL's pioneering work in providing public biological databases to the research community. Its comprehensive range of data resources now includes EMBL-Bank (DNA and RNA sequences), Ensembl (animal genomes), ArrayExpress (microarray data), UniProt (protein sequences and functional information), the Macromolecular Structure Database (protein structures), InterPro (protein families, motifs and domains) IntAct (molecular interactions) and Reactome (pathways). All of these resources are the products of international collaborations with other data providers.

We have a broad palette of research interests that complement our data resources, and these two strands of activity are mutually supportive, with many collaborations between research groups and service teams. Seven research groups aim to understand biology through the development of new approaches to interpreting biological data. These approaches are diverse, and include: classifying and understanding proteins and their interactions; mathematical analyses of evolutionary models; computational modelling of neuronal signalling; text mining; statistical approaches to functional genomics; and large scale analysis of regulatory systems and differentiation. In addition, our services teams perform extensive research to enhance existing data resources and develop new ones.

Almost all of our groups offer PhD places through EMBL's international PhD Programme, and EMBL-EBI's PhD places give students registration at and, if successful, degrees from the University of Cambridge. For a list of those groups with PhD places available for 2008/2009, please see the PhD studies section of our training web pages: www.ebi.ac.uk/training/Studentships/. Other positions are advertised through the EMBL jobs pages (www.embl.org/jobs).

Janet Thornton
Head of EMBL-EBI

Janet Thornton

PhD 1973, King's College & National Inst. for Medical Research, London. Postdoctoral research at Oxford University, NIMR & Birkbeck College, London. Lecturer, Birkbeck College 1983-1989. Professor of Biomolecular Structure, University College London since 1990. Consultant at EMBL-EBI since 1994. Bernal Professor at Birkbeck College, 1996–2002. Director of the Centre for Structural Biology at Birkbeck and University College London, 1998–2001. Director of EMBL-EBI since 2001.

Computational biology of proteins – structure, function and evolution

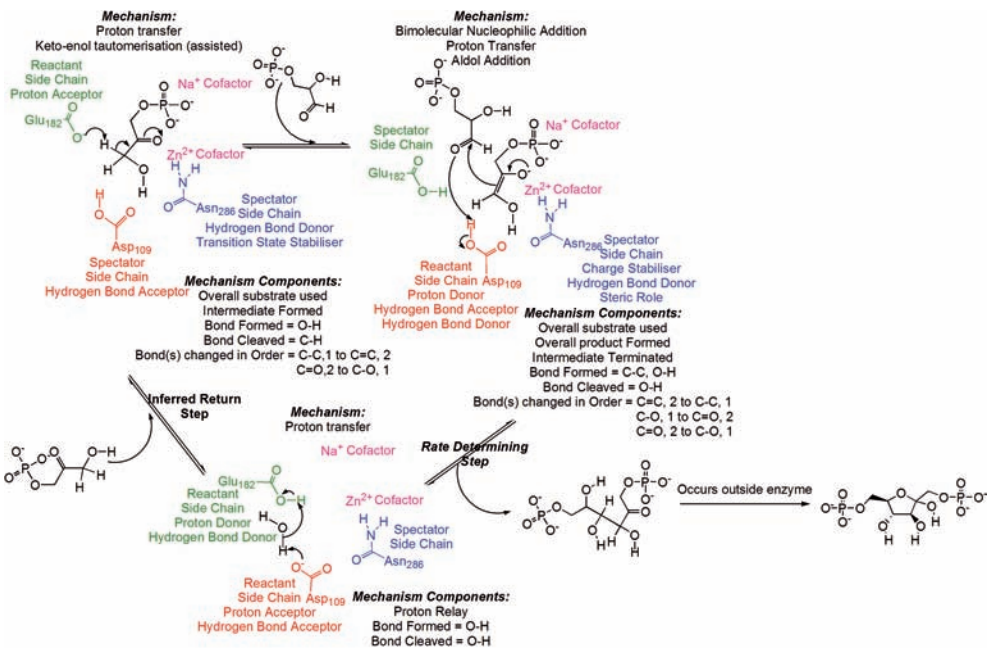
Previous and current research

The goal of our research is to understand more about how biology works at the molecular level. We analyse how enzymes perform catalysis, how these molecules recognise their cognate ligands, and how proteins and organisms have evolved to create life. We develop and use novel computational methods, gathering data either from the literature or by mining the data resources, to answer specific questions. Much of our research is collaborative, involving either experimentalists or other computational biologists. During 2006 our major contributions have been in the following five areas:

- Enzyme structure and function
- Using structural data to predict protein function and to annotate genomes
- Evolutionary studies of genes and their expression
- Functional genomics analysis of ageing
- Development of tools and web resources

Future projects and goals

We will continue our work on understanding more about enzymes and their mechanisms, including a study of how the enzymes, their families and their pathways have evolved. This will include a study of the structural regulation of enzymes by proteins and small molecules. We would like to combine this detailed understanding of molecular evolution with a study of gene expression in different cell types and organs. In our work on ageing, we will study ageing-related diseases and how these are affected in long-lived model organisms.



An example of the annotation found in a MACiE entry. The reaction shown here corresponds to fructose-bisphosphate aldolase (entry 52). MACiE (www.ebi.ac.uk/thornton-srv/databases/MACiE/) is a database of enzyme reaction mechanisms, and is publicly available as a web-based data resource. (Holliday et al.*, 2006).*

Selected references

Bashton M., Nobeli I. & Thornton, J.M. (2006). Cognate ligand domain mapping for enzymes. *J. Mol. Bio.*, 365, 836-852

Freilich S. *et al.* (2006). Relating tissue specialisation to the differentiation of expression of singleton and duplicate mouse proteins. *Genome Res.*, 9, R89

Holliday, G. M. *et al.* (2006) MACiE (Mechanism, Annotation and Classification in Enzymes): novel tools for searching catalytic mechanisms. *Nucl. Acids Res.*, 35, D515-D520, 2007.

Stockwell, G.R. & Thornton, J.M. (2006). Conformational diversity of ligands bound to proteins. *J. Mol. Biol.*, 356, 928-944

Paul Bertone

PhD 2005, Yale University, USA.
At EMBL-EBI since 2005.
Group Leader since 2006.
Joint appointment with the Gene Expression and Developmental Biology Units.

Differentiation and development

Previous and current research

Much of my previous work has entailed the development of novel microarray platforms for genomic and proteomic analyses, extending these technologies beyond conventional array-based gene expression profiling and toward the global analysis of complex biological systems. Related platforms include chromosome- and genome-scale DNA tiling arrays for the discovery of novel functional elements in eukaryotic genomes (Bertone *et al.* 2005), and protein-based microarrays for the large-scale characterisation of biochemical activities (Bertone & Snyder, 2005).

My research group investigates the cellular and molecular processes underlying mammalian stem cell differentiation, using a combination of experimental and computational approaches. Embryonic stem (ES) cells are similar to the transient population of self-renewing cells within the inner cell mass of the preimplantation blastocyst, capable of pluripotential differentiation to all specialised cell types comprising the adult organism. Stem cells undergo continuous self-renewal and produce identical daughter cells, or can develop into specialised progenitors and terminally differentiated cells. Each regenerative or differentiative division involves a decision whereby an individual cell remains in self-renewal or commits to a particular lineage. Pluripotent ES cells can give rise to lineage-specific precursors, with an accompanying restriction in commitment potential; these exist *in vivo* as self-renewing multipotent progenitors localised in reservoirs within developed organs and tissues. The properties of proliferation, differentiation and lineage specialisation are fundamental to cellular diversification and growth patterning during organismal development, as well as the initiation of cellular repair processes throughout life.

A variety of molecular pathways involved in embryonic development have been elucidated, including those influencing stem cell differentiation. As a result, we know of a number of key transcriptional regulators and signalling molecules that play essential roles in nuclear potency and self-renewal capacity of embryonic and tissue-specific stem cells. Despite these efforts however, a small number of components have been identified and large-scale characterisation of cellular specialisation and terminal differentiation remains incomplete. In studying the components involved in these processes, we hope for a better understanding of which factors confer the capacity for self-renewal, and the events that trigger the onset of lineage commitment. The generation of large-scale data from functional genomic and proteomic experiments will define the global regulatory influence of key transcription factors, signalling genes and non-coding RNAs involved in early developmental pathways, leading to a more detailed understanding of the molecular mechanisms controlling vertebrate embryogenesis.

Future projects and goals

A long-term goal of this work is to elucidate an accurate model of stem cell differentiation and lineage commitment at various biological levels. Despite the importance of transcription factors and the interaction of co-factor proteins on the repression and activation of genes, eukaryotic cells utilise many layers of regulatory control. These range from histone acetylation and methylation events affecting chromatin accessibility, variations in transcript splicing producing alternate isoforms in certain cell types or conditions, the attenuation of message levels and/or inhibition of translation by antisense RNAs, and myriad posttranslational modifications affecting protein function and subcellular localisation. Computational approaches will be vital for the analysis and integration of these data in context with existing knowledge.

We eventually wish to characterise the complex interaction of signalling pathways, gene regulation by key transcription factors and non-coding RNAs, and chromatin modifications that function in concert to induce distinct morphological and physiological outcomes. A first step in the process of system-level modelling is the construction of regulatory networks from time-resolved gene expression profiles. Such an approach can be applied to data generated from the projects described above to build regulatory networks from experimental results, augmented by existing information from external resources. Using this approach, we can examine changes in network topology and gene expression patterns in response to permutations of the system. Linked to biological data from many sources, this will become a powerful framework for exploring the biological activities and system-wide impact of transcriptional and translational regulators during cell differentiation.

Selected references

Bertone, P. *et al.* (2006). Design optimization methods for genomic DNA tiling arrays. *Genome Res.*, 15, 674-680

Bertone, P., Gerstein, M. & Snyder, M. (2005). Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery. *Chromosome Res.*, 15, 259-274

Bertone, P. & Snyder, M. (2005). Advances in functional protein microarray technology. *FEBS J.*, 272, 5400-5411

Bertone, P. *et al.* (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science*, 306, 2242-2246

Nick Goldman

PhD 1992, Cambridge University, UK.
Postdoctoral research at National Institute for Medical Research, London, and Cambridge University, UK. Wellcome Trust Senior Fellow 1995-2006.
Group Leader at EMBL-EBI since 2002.

Evolutionary analysis of biological sequence

The group's focus is in the development of improved methods for analysing the patterns of nucleic acid and protein sequence evolution, and in the application of those methods to gain insights into the functions encoded within genomes and the trajectories plotted by past evolution.

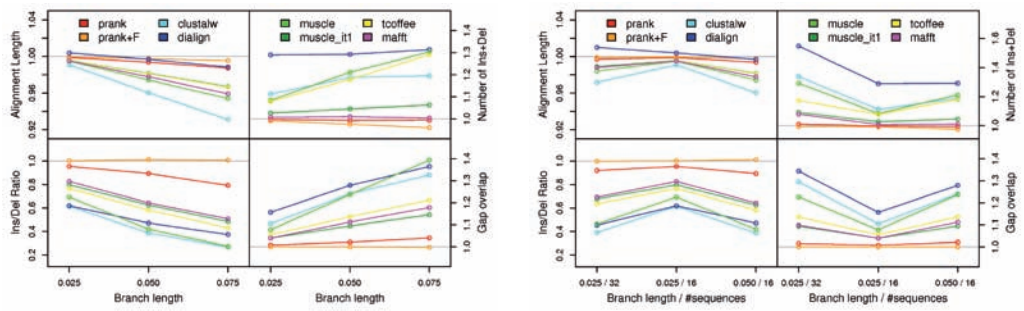
Our work includes the study of the theoretical foundations of biological sequence-based phylogenetic analyses, aiming to understand and improve inferential techniques and develop statistical methods that test the accuracy of our mathematical models. We also concentrate on devising better mathematical models to describe sequence evolution. Typically this involves incorporating greater mathematical complexity with the aim of adding biological reality. By determining which biological features add most to the accuracy of the models, we hope to find out which proposed evolutionary forces are responsible for the patterns of sequence variation within and among species. Finally, methods and models are combined in software applications designed to learn about living organisms' evolutionary history and about the processes of evolution that have shaped the diversity of life on Earth.

Most sequence-based phylogenetic analyses rely on sequence alignments, the accuracy of which is crucial to results and conclusions drawn. We have fed back insights from phylogenetic analysis to develop, implement and validate improved alignment methodologies, in turn leading to further insights into the mechanisms of evolution. Results from our phylogenetic analyses have generated hypotheses of evolutionary constraint and biological function, which have been fed-forward to experimental biologists, for example within the international ENCODE consortium.

Future projects and goals

The accelerating accumulation of large-scale genomic sequence datasets such as whole genome sequencing projects and targeted multi-locus studies is generating datasets with a depth and diversity that were previously only available for a handful of loci. We will capitalise on our expertise in powerful and realistic evolutionary models, adapting and developing them as necessary for their application to this genomic-scale data. This work has already started to bear fruit in the context of the ENCODE project, but there are major issues still to tackle such as sequence accuracy and alignment quality, and many more biological questions that can be asked. We will use comparative genomic data to identify functional elements and probe the validity of previously annotated elements.

Continuing the theme of genome-scale analyses, we have established a collaboration with the group of Nick Luscombe (page 87) to investigate the evolutionary dynamics of genome wide transcriptional regulatory networks. This work will involve the generation of reliable multiple sequence alignments, orthology assignment and the application of sequence evolution models to identify constraint and diversification in gene control regions.



Left: The performance of alternative methods in the alignment of data at different evolutionary distances. Most methods overestimate the total number of insertion-deletion events and the overlapping gaps, and underestimate the total length of the alignment and the insertion/deletion rate ratio. These errors get greater with longer evolutionary distances. The values are normalised and unbiased estimates should be close to 1. Right: As the error-generating bias in the alignments is caused by the alignment algorithms' inability to handle insertions, it cannot be corrected by adding more sequences and making the branch lengths shorter. In comparison to the error in a 16-taxon tree with branch length 0.025 (middle), the errors in a 32-taxon tree with branch length 0.025 (left) and in a 16-taxon tree with branch length 0.050 (right), covering roughly the same evolutionary time, are of similar magnitude.

Selected references

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W. & Steinmetz, L.M. (2006). A high-resolution map of transcription in the yeast genome. *Proc. Natl Acad. Sci. USA*, 103, 5320–5325

Klosterman, P.S., Uzilov, A.V., Bendana, Y.R., Bradley, R.K., Chao, S., Kosiol, C., Goldman, N. & Holmes, I. (2006). XRate: a fast prototyping, training and annotation tool for phylo-grammars. *BMC Bioinformatics*, 7, 428

Löytynoja, A. & Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl Acad. Sci. USA*, 102, 10557–10562

Massingham, T. & Goldman, N. (2005). Detecting amino acid sites under positive and purifying selection. *Genetics*, 169, 1753–1762

Wolfgang Huber

PhD 1998, Albert-Ludwigs-Universität, Freiburg, Germany.
Postdoctoral Research at IBM Almaden Research Center, San José, California and German Cancer Research Center (DKFZ), Heidelberg, Germany.
Group Leader at EMBL-EBI since 2004. Joint appointment with the Gene Expression Unit.

Functional genomics

Previous and current research

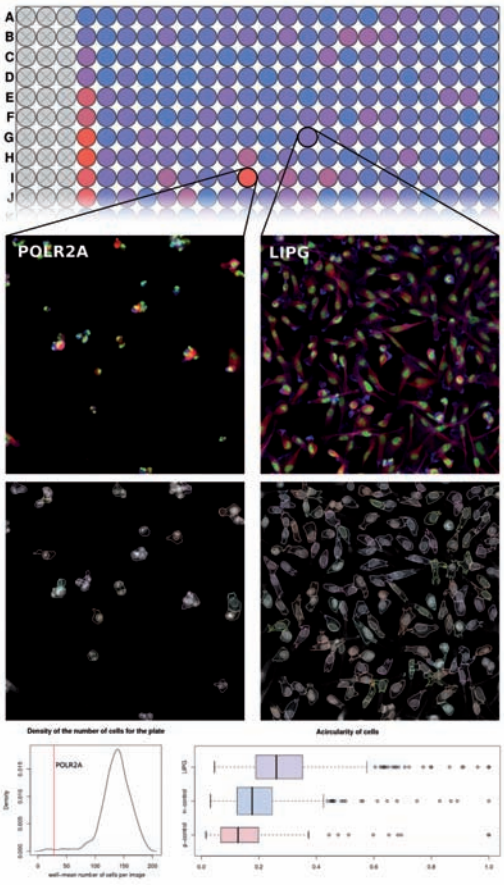
The Huber group develops advanced mathematical and statistical methods for the understanding of functional genomics data. This includes:

- analysis methods for emerging microarray and other functional genomic technologies;
- methods for high-content phenotype data from systematic functional screens and modelling of genetic interaction networks;
- statistical and probabilistic methods for graph-like data and graphical models.

In addition, we regard the publication of high-quality scientific software as complementary to the publication of new methodical approaches and contribute to the Bioconductor project (www.bioconductor.org).

Future projects and goals

Biology and its applications to human health will continue to be driven by advances in experimental technologies. Of particular interest to us are array-based techniques (DNA, mRNA, protein), genome-scale functional cellular assays, and high-content phenotyping using automated microscopy. To make these fruitful for systematic models of biological processes, we aim to stay at the forefront of developments in experimental design, data analysis, statistical software, and mathematical modelling. An emphasis is on project-oriented collaborations with experimenters.



Analysis of a whole genome RNAi screen on HeLa cells. Knock-down phenotypes were recorded using automated microscropy (F. Fuchs, DKFZ Heidelberg). The top panel shows, schematically, how the experiment was conducted in 384-well plate format. Each well corresponds to an siRNA pool specific for one human gene. For each well, replicate images were taken at three different wavelengths (shown in false colour representation in the second row): DAPI staining for the nuclei, tubulin for microtubules and phalloidin for F-actin. Images were automatically segmented (third row) and phenotypic descriptors were extracted (fourth row).

Selected references

Boutros, M., Bras, L. & Huber, W. (2006). Analysis of cell-based RNAi screens. *Genome Biology*, 25, 7, R66

David, L. *et al.* (2006). A high-resolution map of transcription in the yeast genome. *Proc. Natl Acad. Sci. USA.*, 103, 5320-5325

Hahne, F., Arlt, D., Sauermann, M., Majety, M., Poustka, A., Wiemann, S. & Huber, W. (2006). Statistical methods and software for the analysis of high throughput reverse genetic assays using flow cytometry readouts. *Genome Biology*, 17, 7, R77

Huber, W. *et al.* (2006). Transcript mapping with high-density oligonucleotide tiling arrays. *Bioinformatics*, 22, 1963-1970

Nicolas Le Novère

PhD 1998, Pasteur Institute, Paris, France.
Postdoctoral research at Cambridge University, UK and CNRS, Paris, France.
Group Leader at EMBL-EBI since 2003.

Computational systems biology

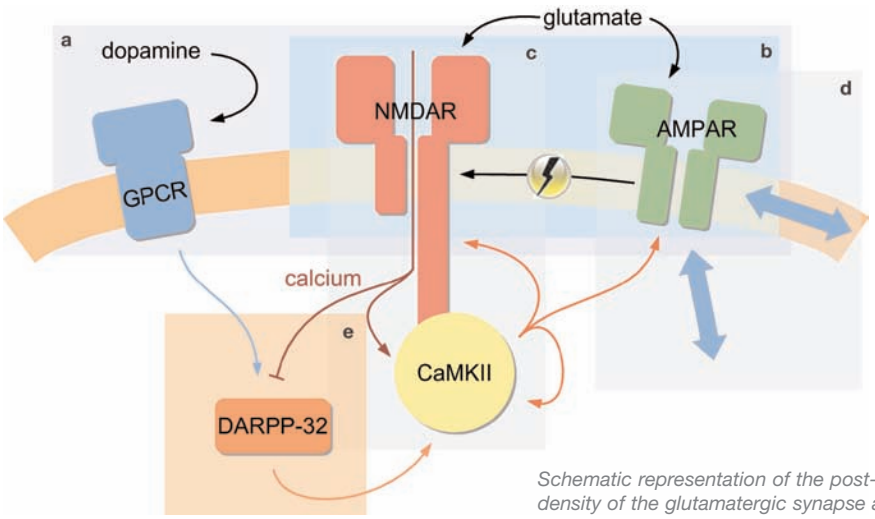
Previous and current research

The Le Novère group's research interests revolve around signal transduction in neurons, ranging from the molecular structure of membrane proteins involved in neurotransmission to modelling signalling pathways. In particular, we focus our work on the molecular and cellular basis of neuroadaptation in neurons of the basal ganglia. The supra-macromolecular structure of the postsynaptic membrane strongly influences signal transduction. Moreover, the whole structure is dynamic and evolves, for example under the control of neuronal activity. By building detailed and realistic computational models, we try to understand how neurotransmitter-receptor movement and clustering, interactions between membrane and cytoplasmic proteins, and spatial location influence synaptic signalling. Downstream from the transduction machinery, we build quantitative models of the integration of signalling pathways known to mediate the effects of neurotransmitters, neuromodulators and drugs of abuse.

We also provide services that facilitate our research, including the creation of standards, database production and software development. The Systems Biology Markup Language (SBML) is designed to facilitate the exchange of biological models between different types of software. Nicolas Le Novère is one of the editors of SBML, and the group also works on extensions to SBML as well as developing software to support SBML usage. The Systems Biology Graphical Notation is an effort to develop a common graphical notation that biochemists and modellers can use to represent pathways and systems. Moving from the form to the content, we are also developing standards for model curation (MIRIAM) and controlled vocabularies to improve the models (the Systems Biology Ontology). Finally, a model is only as useful as it can be accessed. BioModels Database (www.ebi.ac.uk/biomodels/) is a data resource that allows biologists to store, search and retrieve published mathematical models of biological interest.

Future projects and goals

In forthcoming years, the activity of the group will continue in two orthogonal directions. Our research work on modelling neuronal signalling at the level of the dendritic spine will expand to include other signalling pathways (MAPK, PI3K) and tackle problems such as the role of scaffolding proteins or the synchronisation of Ca2+ waves and phosphorylation gradients. Building on the growth of BioModels Database, we will also carry out research on model composition, with the aim of improving component identification and reaction matching to build large-scale models of cellular compartments such as dendritic spines. Our involvement in developing standards and resources for Systems Biology will continue, with the goal of completing the puzzle of representations and ontologies so as to integrate efficiently the different levels of description of biochemical and cellular processes, qualitative, quantitative and experimental.



Schematic representation of the post-synaptic density of the glutamatergic synapse and the relation between the research projects of the group.

Selected references

Fernandez, E., Schiappa, R., Girault, J.A. & Le Novère, N. (2006). DARPP-32 is a robust integrator of dopamine and glutamate signals. *PLoS Comput. Biol.*, 2, e176

Le Novère, N., Bornstein, B., Broicher, A., Courtot, M., Donizelli, M., Dharuri, H., Li L., Sauro, H., Schilstra, M., Shapiro, B., Snoep, J.L. & Hucka, M. (2006). BioModels Database: A Free, Centralized Database of Curated, Published, Quantitative Kinetic Models of Biochemical and Cellular Systems. *Nucleic Acids Res.*, 34, D689-D691

Tolle, D. & Le Novère, N. (2006). Particle-based Stochastic Simulation in Systems Biology. *Current Bioinformatics*, 1, 315-320

Le Novère, N. *et al.* (2005). Minimum Information Requested In the Annotation of biochemical Models (MIRIAM). *Nat Biotechnol.*, 23, 1509-1515

Nicholas Luscombe

PhD 2000, University College London, UK.
Postdoctoral research at the Department of Molecular Biophysics & Biochemistry, Yale University, USA.
Group Leader at EMBL-EBI since 2005.
Joint appointment with the Gene Expression Unit.

Genome-scale analysis of regulatory systems

Previous and current research

Cellular life must recognise and respond appropriately to diverse internal and external stimuli. By ensuring the correct expression of specific genes at the appropriate times, the transcriptional regulatory system plays a central role in controlling many biological processes: these range from cell cycle progression and maintenance of intracellular metabolic and physiological balance, to cellular differentiation and developmental time-courses. Numerous diseases result from a breakdown in the regulatory system and a third of human developmental disorders have been attributed to dysfunctional transcription factors. Furthermore, alterations in the activity and regulatory specificity of transcription factors are now established as major sources for species diversity and evolutionary adaptation. Indeed, increased sophistication in the regulatory system appears to have been a principal requirement for the emergence of metazoan life.

Much of our basic knowledge of transcription regulation has derived from molecular biological and genetic investigations. In the past decade, the availability of genome sequences and development of new laboratory techniques have generated (and continue to generate) information describing the function and organisation of regulatory systems on an unprecedented scale. Genome-scale studies now allow us to examine the regulatory system from a whole-organism perspective; on the other hand, however, observations made with these data are often unexpected and appear to complicate our view of gene expression control.

This continued flood of biological data means that many interesting questions require the application of computational methods to answer them. The strength of bioinformatics is its ability to uncover general principles providing global descriptions of entire systems. Armed with these biological data we are now in a great position to do this.

Much of our work so far has focused on the regulatory system in the yeast Saccharomyces cerevisiae: by integrating diverse data sources – from genome sequence to the results of functional genomics experiments – we can study the regulatory system at a whole-organism level. More recently, we also expanded our interests to understanding regulation in enterobacteria and humans. Below we describe some of our new findings in these new areas as well as our continued work in yeast.

Our current projects include:

- Developing graphical models to integrate disparate data sources and provide a probabilistic framework for understanding the relationship of regulatory network structure to its function.
- Studying the structure and robustness of the regulatory network to identify regions of vulnerability that have potential implications for diseases.
- Examining how the transcriptional regulatory network interacts with other cellular components such as the metabolic system.
- Analysing the repertoire, usage and cross-species conservation of transcription factors in the human genome.
- Wet/dry lab collaborations to uncover the regulation governing complex bacterial behaviour.

In 2007 we will continue advancing analysis techniques and our understanding of regulatory systems in relatively simple organisms such as yeast. We will also consolidate our work with the regulatory apparatus in bacteria and mammalian organisms. A major focus continues to be our close interactions with research groups performing genome-scale experiments.

Future projects and goals

We will continue to develop new techniques to advance our understanding of regulatory systems, and expand our approaches towards alternative regulatory processes. Finally, we will continue to interact closely with research groups performing functional genomics experiments.

Selected references

Balaji, S., Babu, M.M., Iyer, L.M., Luscombe, N.M. & Aravind, L. (2006). Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. *J. Mol. Biol.*, 360, 213-227

Royce, T.E. *et al.* (2006). Extrapolating traditional DNA microarray statistics to the tiling and protein microarray technologies. *Methods Enzymol.*, 411, 282-311

Luscombe, N.M. *et al.* (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, 431, 308-312

Seshasayee, A.S.N., Bertone, P., Fraser, G.M. & Luscombe, N.M. (2006). Transcriptional regulatory networks in bacteria: from input signals to output responses. *Curr. Opin. Microbiol.*, 9, 511-519

Dietrich Rebholz-Schuhmann

Master in Medicine 1988, and PhD 1989, Immunology, University of Düsseldorf, Germany.
Master in Computer Science 1993, Passau, Germany.
Senior Scientist at gsf, Munich, and LION bioscience AG, Heidelberg, Germany.
Group Leader at EMBL-EBI since 2003.

Information extraction from scientific literature

Previous and current research

The number of scientific publications published every year is high and constantly increasing (currently 600,000 Medline abstracts per year). This requires new techniques to automatically integrate facts from the scientific literature into public data resources. In theory, text mining is the perfect solution to transform factual knowledge from publications into database entries, but no IT solution exists that solves this task and satisfies the diversity of demands. One reason is that scientific language is so complex that corpus annotation must still be done with human intervention and thus insufficient data is available to train machine learning techniques to do this task. Another reason is that only a flexible solution can fulfil the growing demands from curation teams for more diverse data. Finally, a solution has to tie biological data from public data sources into the annotation process to support curators in their interpretation of findings in the text.

In this context, research in the Rebholz group is focused on information extraction from the scientific literature, thus supporting curators and the scientific community. Our goal is to make solutions available that automatically integrate literature with other biomedical data resources (e.g. bioinformatics databases). Such solutions analyse a document, recognise biomedical terms and link them to relevant biomedical resources, all in an automated way. In addition we extract events such as protein–protein interactions, and recognise nomenclatures (e.g. karyotypes). As a result, authors and curators can submit a document and instantly receive an analysis of its content.

During 2006 we further developed two services, EBIMed and Protein Corral, that we established in 2005. EBIMed (www.ebi.ac.uk/Rebholz-srv/ebimed/index.jsp) is the first public text mining service that analyses Medline abstracts in real time to identify and link large sets of terms to their corresponding public data resources (including UniProtKB, GO and Medline Plus). Protein Corral (www.ebi.ac.uk/Rebholz-srv/Pcorral) extracts protein–protein interactions from Medline abstracts using language patterns. Other tools and services that we are developing are concerned with providing local context for query terms (MedEvi) and integrating annotations into full-text documents (Text-to-GO).

As part of a new European research project, BOOTStrep, we are developing terminological and ontological resources from the scientific literature. This project will focus on information relevant to gene regulation. Currently we are building a large-scale terminological resource (BioLexicon) that integrates lexical and contextual information.

In the last two years we have tackled the problem of harmonising annotations in the scientific literature by proposing a universal schema ('IeXML') for annotations. This is an important step to enable semantic enrichment of scientific literature using text mining components from different research groups.

Future projects and goals

Several goals are high on the agenda for the next two years. First we will continue ongoing research work in term recognition and term mapping to biomedical data resources. This requires sophisticated handling of terminological resources, disambiguation of terms in the scientific literature and assessment of the final result.

Second we will invest more effort into the extraction of content from the scientific literature. The BOOTStrep project is a good example of this goal. In this project we will shape an ontology representing gene regulation concepts to improve information extraction techniques and to fill a fact database with content from the scientific literature.

Third, we will investigate workflow systems that use text mining as part of bioinformatics information retrieval solutions, where public biomedical data resources are integrated into the data from the biomedical scientific literature.

Finally, we will increase the availability of information extraction solutions based on SOAP Web services for the benefit of the bioinformatics community. This requires standards in the annotation of scientific literature and will automatically lead to semantic enrichment of the scientific literature.

Selected references

Kirsch, H. *et al.* (2006). Distributed modules for text annotation and IE applied to the biomedical domain. *Int. J. Med. Inform.*, 75, 496-500

Rebholz-Schuhmann, D. *et al.* (2006). Protein annotation by EBIMed. *Nat. Biotechnol.*, 24, 902-903

Jensen, L.J., Saric, J. & Bork, P. (2006). Literature mining for the biologist: from information retrieval to biological discovery. *Nat. Rev. Genet.*, 7, 119-29

Liu, H., Hu, Z.Z., Zhang, J. & Wu, C. (2006). BioThesaurus: a web-based thesaurus of protein and gene names. *Bioinformatics*, 22, 103-5

Rolf Apweiler

PhD 1994, University of Heidelberg, Germany.
At EMBL since 1994.
Group Leader since 2000.

The activities of the Sequence Database Group

Previous and current research

The activities of the Sequence Database group are focused on the production of protein sequence, protein family and nucleotide sequence databases at the EMBL-EBI. We maintain and host EMBL-Bank (the European nucleotide sequence database), the UniProt protein resource, and a range of databases around these two core databases. Our main research activities are on automatic annotation of proteins, genes and genomes, as well as mining high-throughput genomics and proteomics datasets.

The main achievements of the Sequence Database group in 2006 have been:

- Handling of an ever-growing amount of data, including nucleotide and protein sequences and complete proteomes (Integr8 for example incorporated 150 new genomes, and Genome Reviews have been extended to include three eukaryotes).
- The creation of the UniProtKB Sequence/Annotation Version database (UniSave), a comprehensive archive of UniProtKB entry versions.
- Programmatic access to Integr8 has been made available via a web services application programming interface, and Genome Annotation Scores measure completeness of annotation for each genome.
- Analysis of the HUPO Human Brain Proteome Project Data.

Future projects and goals

In addition to major improvements of our current systems, we intend to expand our family of UniProt databases by creating the UniProt Metagenomic and Environmental Sequences database (UniMes) in order to respond to the increasing number of available environmental sequences. The source of this data will be the 'environmental and other taxonomically unassigned sequences' in DDBJ/EMBL/GenBank and will use the same production pipeline as that of UniProtKB/TrEMBL. Cross-references will only be provided to DDBJ/EMBL/GenBank and InterPro and its member databases. We will annotate the UniMes sequences completely automatically without any manual annotation. Due to the nature of the sequence data in this database, it would not be a good use of valuable curator time to manually annotate UniMes records. This extreme example well illustrates the fact that with rapid growth of sequence databases there is an increasing need for reliable, automatic, functional annotation of newly predicted proteins.

Also, we will add mining of high-throughput genomics and proteomics datasets to our automatic annotation toolset. Despite the abundance of data from large-scale experimentation on a genome-wide level, such as expression profiling, protein–protein interaction screens or protein localisation, the systematic and integrated use of this type of information for high-throughput annotation of proteins remains largely unexplored. We therefore intend to build on ongoing research activities at the EMBL-EBI to develop and assess new protocols to integrate and analyse functional genomics datasets for the purpose of high-throughput annotation of uncharacterised proteins. This will include the analysis of different data types regarding their suitability for the approach, development of data structures that allow the efficient integration and mining of data of different types and quality as well as benchmarking of the obtained results and the application of the new methodologies to annotation of UniProtKB/TrEMBL records.

Selected references

Kulikova T. *et al.* (2007). EMBL Nucleotide Sequence Database in 2006. *Nucleic Acids Res.*, 35, D16-20

Mueller, M. *et al.* (2006). Functional annotation of proteins identified in human brain during the HUPO Brain Proteome Project pilot study. *Proteomics*, 6, 5059-5075

Sterk, P. *et al.* (2006). Genome Reviews: Standardizing content and representation of information about complete genomes. *Omics*, 10, 114-118

The UniProt Consortium (2007). The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, 35, D193-197

Ewan Birney

PhD 2000, Sanger Institute,
Hinxton, Cambridge, UK.
Team Leader at EMBL-EBI since 2000.

Ensembl, Reactome and the Birney Research Team

Previous and current research

Our joint service and research group runs both the Ensembl system for metozoan genome annotation and Reactome, a human pathway database.

The Ensembl team is one of the leading groups for genome annotation, which is the process of finding functional elements in large genomes using computational approaches. Ensembl provides a framework for working with the genomes of higher animals, focusing on vertebrates. It presents, via an interactive website (www.ensembl.org), the human genome together with other genomes that are important for addressing questions in medical research and molecular biology. Progress in the prediction of non-coding RNA genes has also contributed significantly to the richness of annotation in Ensembl.

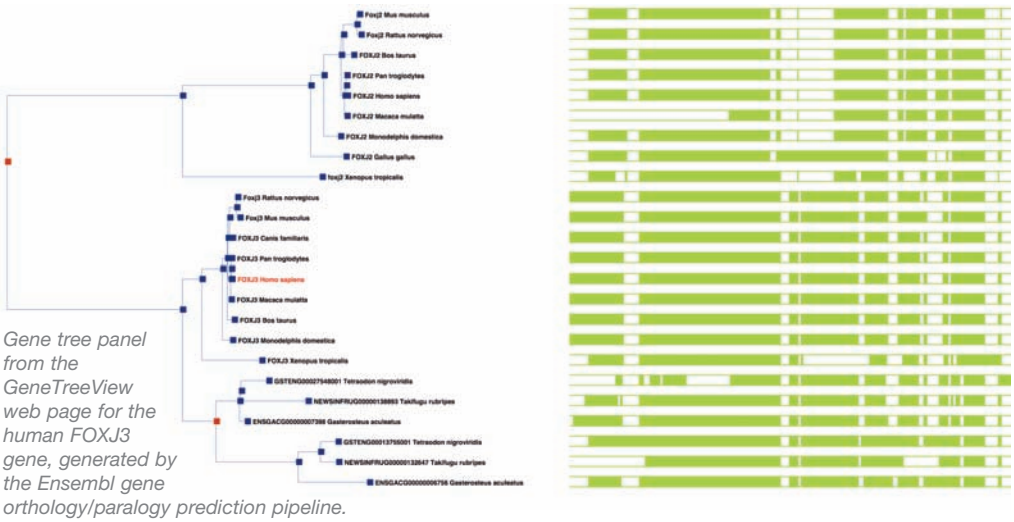
Reactome (www.reactome.org), a curated database of biological processes in humans, is a dual-purpose project that can be used by general biologists as an online textbook of biology, or by bioinformaticians to make discoveries about biological pathways. Reactome is produced in collaboration with Lincoln Stein's group at Cold Spring Harbor Laboratory (CSHL), NY, USA. The basic information in Reactome is provided by bench biologists who are experts on a particular pathway. The information is then managed by groups of curators at CSHL and the EMBL-EBI, peer-reviewed by other researchers and published on the web. Its coverage ranges from the basic processes of metabolism to complex regulatory pathways such as hormonal signalling. New pathways include the complete insulin receptor pathway and the electron transport chain.

Research in the Birney group is directed in these broad areas, with a focus on developing novel algorithms and data-mining techniques to discover new biological information, such as finding new cis-regulatory elements or discovering how a pathway has evolved.

Future projects and goals

Future projects include developing new methods in a variety of areas, including:

- Sequence manipulation methods in graph structures to represent assembly and efficient search strategies;
- Integration of expression and protein–protein interaction data into Reactome;
- Examination of the evolution of pathways in metozoans;
- Development of novel hidden Markov models for gene prediction.
- Investigating new genome-wide functional data such as ChIP–chip.



Gene tree panel from the GeneTreeView web page for the human FOXJ3 gene, generated by the Ensembl gene orthology/paralogy prediction pipeline.

Selected references

Birney, E. *et al.* (2006). Ensembl 2006. *Nucleic Acids Res.*, 34, D556-D561

Cunningham, F. *et al.* (2006). TranscriptSNPView: a genome-wide catalog of mouse coding variation. *Nat. Genet.*, 38, 853-853

Hoffman, M. & Birney, E. (2006). Estimating the neutral rate of nucleotide substitution using introns. *Molecular Biology Evolution*, 24, 522-31

Kahlem, P. & Birney, E. (2006). Dry work in a wet world: computation in systems biology. *Mol. Syst. Biol.*, 2, 40

Alvis Brazma

PhD 1987, Computer Science, Moscow State University, 1987.
Postdoctoral research at New Mexico
State University, Las Cruses, USA.
At EMBL-EBI since 1997. Team Leader since 2001.

The Microarray Informatics Team

Previous and current research

The Microarray Informatics Team is working in four main directions:

- Development and running of the ArrayExpress database system and tools to provide services for functional genomics data.
- Medical informatics-related development projects.
- Development of data standards for systems biology and medical informatics.
- Research in high-throughput data analysis and development of algorithm for systems biology.

Our group was among the first to use microarray data to study transcription regulation mechanisms on a genomic scale (Brazma *et al.*, 1998). In 1999 we realised the importance of standards in microarray data reporting (Brazma *et al.*, 2001) and began work to establish the ArrayExpress database, which is now roughly doubling in size every year; as of November 2006 it holds data from over 50,000 microarrays. More recently, we created a separate database of gene expression profiles. While participating in several large European projects, we have extended our work to other domains of high-throughput data, such as proteomics and metabonomics, which we are integrating with medical informatics data. Our PhD students and postdocs focus mostly on analysing these data to build models for systems biology (e.g., Rustici *et al.*, 2004, Schlitt & Brazma, 2006).

Future projects and goals

The ongoing development of ArrayExpress from a prototype to a robust 24/7 service database will remain the main focus of the group. We will continue to improve the ease with which users can submit, access and mine data in ArrayExpress, and to populate ArrayExpress with high-quality data describing basic biological processes, disease and toxicity. We will continue applying cutting-edge software engineering methodologies to achieve these goals. New data-analysis algorithms will be developed and implemented in Expression Profiler – our online data-analysis tool. Integration and meta-analysis of different datasets, with the goal of understanding and modelling basic biological processes, will continue to be the major focus of our research. For instance, the study and modelling of the cell cycle will be an important collaborative project with the Wellcome Trust Sanger Institute and other partners. We will use microarray data to reconstruct combinatorial relationships in transcription regulation. A different approach will be dynamic modelling using the recently developed finite state linear model. As part of our involvement in the EMBL Centre for Computational Biology we will continue to develop the four-dimensional gene expression atlas integrating expression data from different technologies, organisms, tissue types and developmental stages. Large collaborative projects, such as integration of transcriptomics, proteomics, and metabonomics data to understand the molecular mechanisms of disease, will be an important activity related to medical bioinformatics.

The new ArrayExpress Interface showing several ways of searching and visualising the data.

Selected references

Schlitt, T. & Brazma, A. (2006). Modelling in molecular biology: describing transcription regulatory networks at different scales. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 361, 483-494

Brazma, A. *et al.* (2001). Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.,* 29, 365-371

Brazma, A. *et al.* (1998). Predicting gene regulatory elements *in silico* on a genomic scale. *Genome Res.*, 8, 1202-1215

Rustici, G. *et al.* (2004). Periodic gene expression program of the fission yeast cell cycle. *Nat. Genet.*, 36, 809-817

Midori Harris

PhD 1997, Cornell University, Ithaca, NY, USA.
Scientific Curator, Saccharomyces Genome Database, Stanford University, Stanford, CA, USA.
At EMBL-EBI since 2001.
GO Editor.

The GO Editorial Office

Previous and current research

The Gene Ontology (GO) project (www.geneontology.org/) is a collaborative effort to construct and use ontologies to facilitate the biologically meaningful annotation of genes and their products in a wide variety of organisms. At the EBI, the GO Editorial Office plays a key role in managing the distributed task of developing and maintaining the GO vocabularies, and contributes to a number of other GO project efforts, including documentation, web presence, software testing and user support.

The Gene Ontology Consortium (GOC) provides the scientific community with a consistent and robust infrastructure, in the form of biological ontologies, for describing, integrating, and comparing the structures of genetic elements and the functional roles of gene products within and between organisms. Participating groups include major model organism databases and other bioinformatics resource centres. The GO ontologies cover three key biological domains that are shared by all organisms (The GO Consortium, 2000, 2001):

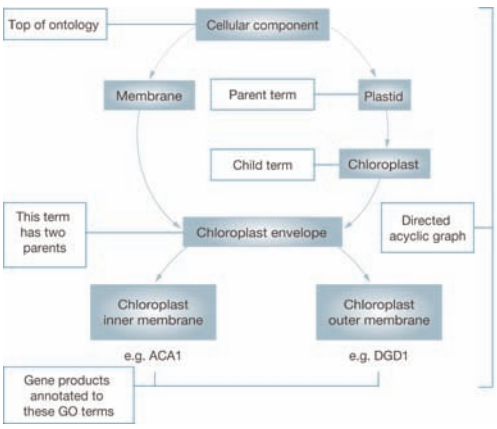
- **Molecular function.** Molecular function defines the tasks performed by individual gene products; examples include aminoacyl-tRNA ligase activity and translation elongation factor activity.
- **Biological process.** Biological process defines broad biological goals, such as protein biosynthesis or ribosome assembly, that are accomplished by ordered assemblies of molecular functions.
- **Cellular component.** Cellular component describes subcellular structures, locations and macromolecular complexes; examples include cytoplasm, ribosome and translation release factor complex.

In addition, sequence features are covered by the Sequence Ontology, which is maintained separately from the three GO ontologies (Eilbeck *et al.*, 2005).

The ontologies in GO are structured as directed acyclic graphs (DAGs), wherein any term may have one or more parents and zero, one, or more children. Within each vocabulary, terms are defined, and parent-child relationships between terms are specified. A child term is a subset of its parent(s). At present, the GO vocabularies define two semantic relationships between parent and child terms: is_a and part_of. The is_a relationship means that a term is a subclass of its parent; part of may mean 'physically part of' (as in the cellular component ontology) or 'subprocess of' (as in the biological process ontology). The figure shows a portion of the GO Cellular Component DAG.

Future projects and goals

The GO Editorial Office will continue to work closely with the rest of the GO Consortium and with biological experts to ensure that the ontologies are comprehensive, logically rigorous and biologically accurate. Among the targeted biological topics are interactions between organisms, heart development, and peripheral nervous system development. Specific logical improvement goals include completing the Biological Process ontology subsumption hierarchy, adding a new relationship type to represent regulation, and recasting many complex process terms as explicit cross-products with orthogonal ontologies such as the ChEBI Ontology and the Cell Ontology.



GO terms are organised in directed acyclic graphs (DAGs) – hierarchical structures in which any 'child' (more specialised term) can have many 'parents' (less specialised terms). For example, the cellular component term chloroplast envelope has two parents, reflecting the fact that it is a part of the chloroplast and a type of membrane. Any gene that is annotated to this term is automatically annotated to both chloroplast and membrane. Some terms and relationships have been omitted for clarity.

Selected references

Harris, M.A. *et al.* (2006). The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.*, 34, D322–D326

The Gene Ontology Consortium (2000). Gene Ontology: tool for the unification of biology. *Nat. Genet.*, 25, 25–29

The Gene Ontology Consortium (2001). Creating the Gene Ontology resource: design and implementation. *Genome Res.*, 11, 1425–1433

Eilbeck, K. *et al.* (2005). The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.*, 6, R44

Kim Henrick

PhD 1974, University of Western Australia, Perth.
Postdoctoral research at the Polytechnic of North London, Imperial College of Science and Technology, SERC Darebury Laboratory, Cambridge Centre for Protein Engineering and National Institute for Medical Research. At EMBL-EBI since 1996. Team Leader since 2001.

Macromolecular Structure Database Team activities

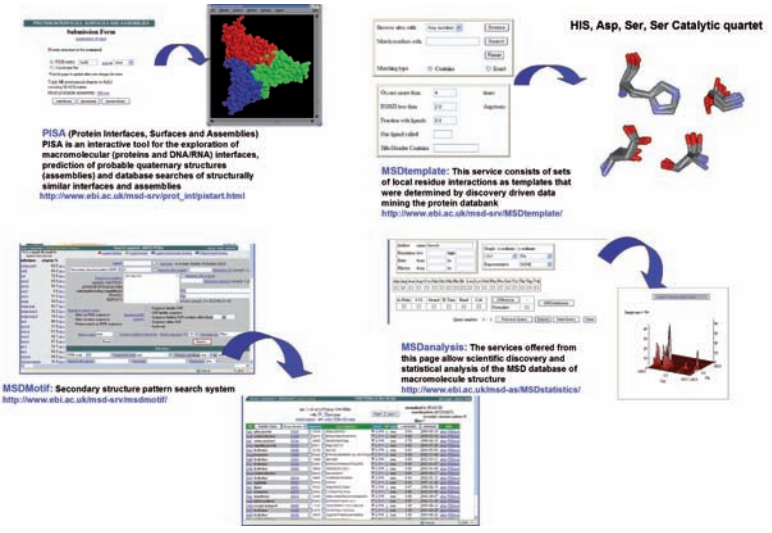
Previous and current research

The Macromolecular Structure Database (MSD; www.ebi.ac.uk/msd/) group is one of the three partners in the worldwide Protein DataBank (wwPDB), the consortium entrusted with the collation, maintenance and distribution of the global repository of macromolecular structure data. Since its inception, the MSD group has worked with partners around the world to improve the quality of PDB data, through a clean-up programme that addresses inconsistencies and inaccuracies in the legacy archive. The improvements in data quality in the legacy archive have been achieved largely through the creation of a unified data archive, in the form of a relational database that stores all of the data in the wwPDB. The three partners are working towards improving the tools and methods for the deposition of new data by the community at large. The implementation of the MSD database, together with the parallel development of improved tools and methodologies for data harvesting, validation and archival, has led to significant improvements in the quality of data that enters the archive. Through this and related projects in the realms of nuclear magnetic resonance spectroscopy and electron microscopy, the MSD continues to improve the quality of publicly available structural data.

Future projects and goals

Biological science will continue to be one of the major areas of focus of scientific research worldwide. Many scientists believe that molecular biology is the primary driver of medical advances in the 21st century. The rapidly increasing volume of molecular data and the need to decipher its patterns has demanded continuous requirements for databases and analysis tools, special curatorial expertise, and unique physical facilities. The new NIH Roadmap recognises that one of the most powerful and unifying concepts of 21st century biology is that of bioinformatics (http://nihroadmap.nih.gov/), as does the European FP7 programme of infrastructure investment, which specifically lists bioinformatics as a key area for development.

The MSD will continue to provide free access to 3D molecular structure data. We face a moving target. There will be a continued increase in the amount of information, and the type of data to be collated is changing. The PDB started as a collection of atomic coordinates that described a protein's structure. Now, our users need 'meta-data' on how the protein was produced and purified, and how its structure was solved. The high-throughput projects continue to make yet greater demands. The success of the structural databases in keeping pace with these changes is due to the creation of the wwPDB and having the partners marshal their resources. Over the next 2–3 years the wwPDB partners intend to merge their development efforts in deposition interfaces and annotation tools into a single toolkit collection. We face not only an ever-increasing flow of data but further developments in methods, making it essential that the wwPDB sites pool their resources.



New MSD services, MSDpisa, MSDtemplate, MSDmotif and MSDanalysis.

Selected references

Albeck, S. *et al.* (2006). SPINE bioinformatics and data-management aspects of high-throughput structural biology. *Acta Crystallogr D Biol Crystallogr.*, 62, 1184-95

Berman, H.M. *et al.* (2006). Outcome of a workshop on archiving structural models of biological macromolecules. *Structure*, 14, 1211-1217

Berman, H.M. *et al.* (2006). Is one solution good enough? – Response. *Nat. Struct. Mol. Biol.*, 13, 185-185

Tagari, M. *et al.* (2006). E-MSD: improving data deposition and structure quality. *Nucleic Acids Res.*, 34, D287-290

Henning Hermjakob

Dipl. Inf. (MSc) 1996, Bioinformatics, University of Bielefeld, Germany.
Research assistant at the National Research Centre for Biotechnology (GBF), Braunschweig, Germany, in the Transfac Database team.
At EMBL-EBI since 1997. Team Leader since 2004.

The activities of the Proteomics Services Team

Previous and current research

The team provides databases and tools for the deposition, distribution and analysis of proteomics and proteomics-related data. We contribute to the development of community standards for proteomics data in the context of the HUPO Proteomics Standards Initiative (PSI) (http://psidev.sf.net), and develop reference implementations for these standards. Extending the 'standards' theme, we provide the IntEnz and ChEBI reference resources for enzyme and chemical nomenclature.

The HUPO PSI Molecular Interaction standard (PSI MI) has been jointly developed by major interaction-data providers from both the academic and commercial sector, among them BIND, Cellzome, DIP, Hybrigenics, IntAct, MINT and MIPS, and is now widely accepted as the standard format for exchange of protein-interaction data. The IntAct (www.ebi.ac.uk/intact) project provides an open source, production-quality, fully portable molecular interaction database and toolset implementing the PSI MI standard. Beyond IntAct, we are now collaborating with other major providers to establish a regular exchange of molecular interaction data between large public data providers. In the framework of the EU ProteomeBinders project, we are also extending the scope of the PSI MI schema to antibodies and other protein binders.

The PSI mass spectrometry work group (PSI MS) has developed the mzData format, a vendor-independent representation of mass spectra, providing a unified format for data archiving, exchange and search engine input. It has been jointly developed by academic users, commercial users and instrument vendors. The PRIDE (www.ebi.ac.uk/pride) database provides a PSI MS-compatible database for protein identifications.

In addition to the MI and MS work groups, HUPO PSI develops the 'Minimum Information About a Proteomics Experiment (MIAPE)' document analog to the MIAME requirements for a microarray experiment, as well as data models to fully represent a proteomics experiment.

In the context of the EU-funded BioSapiens (www.biosapiens.info), Transfog, and Enfin (www.enfin.org) projects, the team contributes to the development of tools for proteomics data integration, in particular the DAS UniProt reference server (www.ebi.ac.uk/uniprot-das/) and Dasty client (www.ebi.ac.uk/dasty/).

In the framework of the IntEnz integrated Enzyme database (www.ebi.ac.uk/IntEnz/) and the ChEBI small molecules database (www.ebi.ac.uk/chebi/), the Proteomics Services Team contributes to the standardisation of proteomics-related nomenclature and data representation.

Future projects and goals

In 2006, the infrastructure of the core resources IntAct and PRIDE has been significantly improved, enabling us to cope with significantly higher throughput in 2007. Through the IMEx and ProDac consortia we expect the coverage of newly published proteomics data to increase significantly, providing ever more valuable proteomics resources.

A major focus for the coming year will be the integration of projects within the group, in terms of software, strategies and user interfaces, to provide integrated resources for systems biology, and to contribute efficiently to large European networks and projects such as BioSapiens, Transfog and Enfin. Extending efforts begun in 2006, we will contribute to a series of international workshops to raise awareness for the Proteomics Services Team resources, both for data deposition and for data analysis. While the primary focus of the team is and will remain the service aspect, we aim to increase research activities based on the resources we develop, for example the large-scale comparative analysis of experiments currently under development for PRIDE. We will also continue to seek partnerships for specific curation projects in collaboration with experimental groups, such as the recently published analysis of protein interactions in schizophrenia (Camargo, 2006).

Major activities in the biochemical area will be the automated analysis of large-scale patent data in collaboration with the European Patent Office as well as the definition and implementation of a reaction ontology in the IntEnz database. Specific projects expected to produce published results are the Enfin data integration platform, standards and infrastructure developed for the representation of protein binders developed in the ProteomeBinders project, and the release of major modules of the standards developed by the PSI.

Selected references

Camargo, L.M. *et al.* (2006). Disrupted in Schizophrenia 1 Interactome: evidence for the close connectivity of risk genes and a potential synaptic basis for schizophrenia. *Mol. Psychiatry*, 12, 74-86

Cote, R.G. *et al.* (2006). The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics*, 7, 97

Hermjakob, H. & Apweiler, R. (2006). The Proteomics Identifications Database (PRIDE) and the ProteomExchange Consortium: making proteomics data accessible. *Expert Rev. Proteomics*, 3, 1-3

Taylor, C.F. *et al.* (2006). The work of the Human Proteome Organisation's Proteomics Standards Initiative (HUPO PSI). *Omics*, 10, 145-151

Rodrigo Lopez

Vet. Med Degree 1984, Oslo Vet. Hoyskole and NASAS Cand. Scient.
Molecular Toxicology and Informatics 1987, University of Oslo.
At EMBL-EBI since 1995.
Team Leader since 2001.

The EMBL-EBI External Services Team

Previous and current research

The EMBL-EBI's External Services team is in charge of the web infrastructure through which services become visible to both external and internal users. Providing access to a wide range of services that include sequence similarity searches, various sequence alignment and phylogeny analysis applications, protein function prediction tools, data search and retrieval systems and general support for the development of web applications and project-specific web sites are amongst the team's main tasks. All members of the group are involved in providing training and support within the institute as well as to the EBI's web users. They also actively participate in various conferences and workshops around the world. In December 2006, the team launched the EBI's new web site and search engine. Emphasis has been placed on making it easy for the user to find and navigate the vast amounts of information stored in the EBI databases, and making this navigation consistent across the entire web site. The new search engine, EB-eye, provides a fast and powerful way of exploring all our databases.

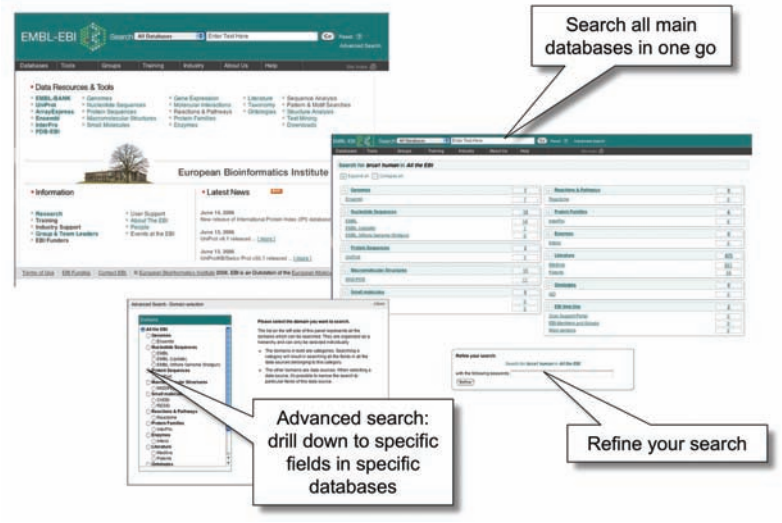
Future projects and goals

The new web site and search engine will pose interesting challenges in the near future. If our expectations are correct, we will have peaks of more than 500 queries per second on the search engine and will require additional resources to appropriately tackle these. Furthermore, if successful, this service will produce an ever-growing demand for the resources available for storage and indexing as well as in the data production pipelines.

Our dispatcher middleware continues to provide access to synchronous and asynchronous job requests. A new version of this middleware component that mediates access to the computational resources is in the making. This new version is being developed in Java and will provide a state-of-the-art API for internal as well as external application developers.

On the training and education fronts, the team will continue to improve on the 2can educational resource. Interestingly, this resource has seen some healthy growth in usage recently, possibly as a result of *Science*'s NetWatch having made a mention of it on 6 October 2006.

User support is an important daily task for the team. We are currently investigating novel means of enabling our users to communicate with us and have better tracking in order to provide faster response to requests.



The EBI's new search engine, EB-eye, allows you to perform an integrated keyword search of all the EBI's data resources. You can refine your search to reduce the number of hits to a manageable level, or drill right down to specific fields in individual databases.

Selected references

Albeck, S. *et al.* (2006). SPINE bioinformatics and data-management aspects of high-throughput structural biology. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 1184-95

Cochrane, G. *et al.* (2006). EMBL Nucleotide Sequence Database: developments in 2005. *Nucleic Acids Res.*, 34, D10-15

Labarga, A. & Brooksbank, C. (2006). A Quick Guide to Web Services at the EMBL-EBI. OBBeC Life Sci. *Bioinformatics*, June, 24-26

Whelan, S., de Bakker, P.I.W., Quevillon, E., Rodriguez, N. & Goldman, N. (2006). PANDIT: an evolution-centric database of protein and associated nucleotide domains with inferred trees. *Nucleic Acids Research*, 34, D327-D331

Wu, C.H. *et al.* (2006). The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.*, 34, D187-D191

Nicola Mulder

PhD 1998, University of Cape Town, South Africa.
At EMBL-EBI since 1999. Team Leader since 2004.
Joint appointment between EMBL-EBI and the
University of Cape Town since 2004.

The InterPro Team

Previous and current research

The InterPro team coordinates the InterPro, CluSTr and Gene Ontology Annotation (GOA) projects at EMBL-EBI.

InterPro is an integrated documentation resource for protein families, domains and functional sites. The project integrates signatures from the major protein signature databases into a single resource, and currently includes data from Pfam, PRINTS, PROSITE, ProDom, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, Gene3D and PANTHER.

During the integration process, InterPro rationalises where more than one protein signature describes the same protein family/domain, and unites these into single InterPro entries, with relationships between them, where applicable. Additional biological annotation is included, together with links to external databases such as GO, PDB, SCOP and CATH. InterPro precomputes all matches of its signatures to UniProtKB proteins using the InterProScan software, and displays the matches in various formats, including table and graphical views and the InterPro Domain Architectures view.

InterPro has a number of important applications, including the automatic annotation of proteins for UniProtKB/TrEMBL and genome annotation projects. InterPro is used in the Integr8 genome and proteome reviews database and in the GOA project to provide large-scale mapping of proteins to GO terms.

GOA provides GO annotations for UniProtKB proteins through manual and electronic annotation, and integrates GO annotations from all model organism databases that are part of the GO Consortium. All annotations are provided regularly as gene association files available for download. The CluSTr project aims to cluster all UniProtKB proteins and protein sets from complete genomes. The resulting clusters and similarity scores are accessible via a web interface.

Future projects and goals

Our future goals in InterPro are to extend InterPro's existing functionality and integrate new methods to increase coverage. For new features, we will group InterPro entries by functional categories and are redesigning the look of InterPro entries to make each feature more intuitive to use. We also aim to provide access to UniParc matches via the web interface. In addition, InterPro aims to provide links to gene expression data and metabolic pathway information. The addition of links such as these will increase the biological value of InterPro and its interoperability with other databases.

Future plans for CluSTr include:

- Work on better visualisation tools for CluSTr
- Continue working on refining CluSTr-based homology predictions (e.g. inclusion of in-paralogues)
- Incorporate CluSTr data into the EMBL-EBI's automatic annotation pipeline to predict annotation of UniProtkB/TrEMBL entries.
- Achieve full coverage of UniParc (excluding future environmental data).

Future plans of GOA are to increase the breadth and depth of GO annotations. The focus will continue to be on human proteins, and grant funding to maintain the curators will be pursued. To increase our coverage of UniProtKB, new ontologies developing within the UniProt Consortium, e.g. UniPathway and subcellular locations, will be manually mapped to GO. Our outreach efforts will increase to ensure that users' needs are served and that we receive feedback and data from the community. The QuickGO browser will be developed to allow batch querying with multiple ids, interpro2GO mappings will be ranked according to the quality of their prediction, and a new text mining aid developed for curator training will be released to the public. Finally, external well-maintained GO Tools will be reviewed by our curators and listed on the GOA Homepage.

NB: Nicola Mulder will be leaving EMBL-EBI in September 2007 and will be replaced as Team Leader by current team member Sarah Hunter.

Selected references

Mulder, N.J. *et al.* (2007). New developments in the InterPro database. *Nucleic Acids Res.*, 35, D224-228

Quevillon, E., Silventoinen, V., Pillai, S., Harte, N., Mulder, N., Apweiler, R., & Lopez, R. (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.,* 33, W116-120

Petryszak, R., Kretschmann, E., Wieser, D., & Apweiler, R. (2005). The predictive power of the CluSTr database. *Bioinformatics* 21, 3604-3609

Lee, V., Camon, E., Dimmer, E., Barrell, D., & Apweiler, R. (2005). Who tangos with GOA? – Use of Gene Ontology Annotation (GOA) for biological interpretation of '-omics' data and for validation of automatic annotation tools. *In Silico Biol.*, 5, 5-8

Peter Rice

BSc 1976, University of Liverpool, UK.
Previously at EMBL Heidelberg (1987-1994), The Sanger Centre (1994-2000) and LION bioscience AG (2000-2002).
Group Leader at EMBL-EBI since 2003.

Grid and e-Science R&D at EMBL-EBI

Previous and current research

The team's focus is on the integration of bioinformatics tools and data resources. We have the remit to investigate and advise on the e-Science and Grid technology requirements of the EMBL-EBI, through application development, training exercises and participation in international projects and standards development. Our group is responsible for the EMBOSS open source sequence analysis package, the Taverna bioinformatics workflow system (originally developed as part of the myGrid UK eScience project) and for various projects (including EMBRACE and ComparaGrid) that integrate access to bioinformatics tools and data content.

Future projects and goals

The services provided by the group remain largely SOAP-based web services. These have proved themselves to be highly useful to prototype and develop service and metadata standards. We are looking, especially through the EMBRACE project, to migrate to true Grid services, but like many other groups we are waiting for the long-anticipated merging of web and grid service standards.

The EMBOSS project plans to expand in the coming few years to cover bioinformatics more generally, including genomics, protein structure, gene expression, proteomics, phylogenetics, genetics and biostatistics. This will require the participation of external groups to expand the project beyond its current EBI base, and we are actively seeking potential partners in each area. We will expect to build a service-based eScience architecture around the applications and data resources through the EMBRACE project, with support and guidance from the community of users in academia and industry.

The EMBRACE project will move beyond sequence data and analysis services to cover the remaining areas of the EBI's core databases and to integrate services from our partners using the same standards and interfaces.

Selected references

Oinn, T. *et al.* (2006). Lessons in creating a workflow environment for the life sciences. *Concurr. Comput.-Pract. Exp.*, 18, 1067-1100

Rice, P. *et al.* (2006). EMBRACE: Bioinformatics data and analysis tool services for e-Science. *Proc. 2nd IEEE Conf. on e-Science and Grid Computing*, in press

Ugis Sarkans

PhD 1998, Computer Science, University of Latvia.
Postdoctoral research at the University of Wales, Aberystwyth.
At EMBL-EBI since 2000.
Technical Team Leader since 2006.

The Microarray Software Development Team

Previous and current research

Our team has been developing software for ArrayExpress since 2001 (Sarkans *et al.*, 2005). As of November 2006 ArrayExpress holds around 50,000 microarray hybridisations and has become one of the major data resources at EMBL-EBI.

The current ArrayExpress infrastructure consists of:

- repository – the archival MIAME-compliant database for the data that supports publications;
- data warehouse – a query oriented database of gene expression profiles;
- MIAMExpress – a web-based data annotation and submission system;
- Expression Profiler – a web-based data analysis toolset;
- various components used internally by the curation team.

A more recent activity of the team is moving towards medical informatics. We are developing a data management and integration platform for high-throughput experimental data beyond transcriptomics. The project that drives this work is MolPAGE (Molecular Phenotyping to Accelerate Genomic Epidemiology), an Integrated EU project that aims to find biomarkers for genetic diseases, type II diabetes in particular, by employing a range of high-throughput functional genomics technologies together with advanced statistical analysis.

We have also participated in data standardisation efforts, in particular MAGE (Spellman *et al.*, 2002), and see this work as crucial for the successful development and support of a high-throughput data management infrastructure.

Future projects and goals

The main focus of the team is improving the robustness and usability of ArrayExpress software infrastructure. In 2007 we are planning to redesign the database tier of ArrayExpress, with the aim of tightening integration among its various components. This will minimise the resources needed to move data around inside the system.

We will continue to improve user experience when working with ArrayExpress, both for data access and for data acquisition. We have improved and simplified the data access interfaces in 2006; further work is needed for data acquisition, where the MIAMExpress batch uploader system will be aligned with the emerging MAGE-TAB standard. A challenging task here is to be able to easily customise the submission tool for various user communities.

The number of different data processing tools has been continuously increasing. We are planning to design and implement a workflow system that will streamline various data management tasks in the group.

We will continue to develop new data-analysis algorithms for Expression Profiler. The data warehouse will be more tightly integrated with Expression Profiler so that various data analysis options are easily available for the warehouse users.

For medical informatics we will continue to further develop the set of tools developed for the MolPAGE project. We believe the same architecture can be used for other similar projects, so we are building generalised software components that can be customised for other contexts.

Selected references

Brazma, A. *et al.* (2006). Standards for systems biology. *Nat. Rev. Genet.*, 7, 593-605

Rayner, T.F. *et al.* (2006). A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. BMC *Bioinformatics*, 7, 489

Sarkans, U. *et al.* (2005). The ArrayExpress gene expression database: a software engineering and implementation perspective. *Bioinformatics*, 21, 1495-501

Spellman, P.T. *et al.* (2002). Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol.*, 3

Peter Stoehr

MSc 1978 (Applied Genetics), Birmingham University, UK.
Statistician in Agriculture Faculty, Newcastle University.
Analyst/Programmer with AFRC, Cambridge and Harpenden, UK.
At EMBL since 1988; at EMBL-EBI since 1994.

Literature resource development

Previous and current research

A key goal is to integrate the scientific literature with the biological databases and provide public services to exploit this. Where feasible we achieve this by implementing state-of-the-art information retrieval methods and novel biomedical text-mining methods.

We maintain a local copy of PubMed under lease from the US National Library of Medicine (NLM). Biological patent abstracts are extracted monthly from data products of the European Patent Office.

CiteXplore is available at www.ebi.ac.uk/citexplore as a tool for querying these data, showcasing text-mining methods, and linking to biological databases and full-text. UMLS, GO, the NCBI taxonomy and gene synonyms from UniProt are used as thesauri. 'Whatizit' text-mining methods from the Rebholz research group at EMBL-EBI provide filters for enrichment of text with annotation. Gene and protein names from UniProt and GO terms are examples of entities identified in text and hyperlinked to the underlying data resources.

The technical implementation features Oracle database technology (XML-DB) and the open source Lucene full-text retrieval engine. Lucene provides good performance at both indexing and retrieval, and is rich in lexical functionality including word stemming, wild-carding, proximity searches, ranking of search results and highlighting of search terms.

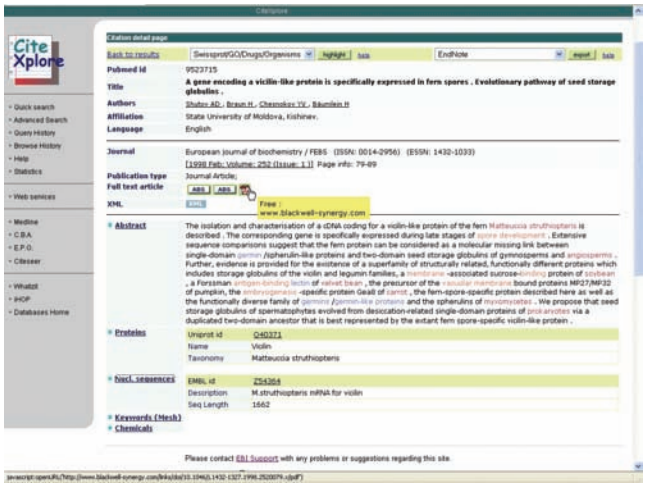
Future projects and goals

In addition to PubMed, other abstract and indexing resources are being explored for research areas that are not fully covered by PubMed. Examples of these are AGRICOLA (USDA) for plant sciences, agriculture and food, and AMED (British Library) for allied and complementary medicine. Grey literature and theses are also possible sources to harvest, albeit with much more human effort.

We aim to incorporate the growing PubMed Central archive and full-text patents into our services. Most of the tools and techniques to manage full-text will be the same as for the abstracts, but some adaptations will be needed to cater for the larger size and the sectioned nature of a complete article.

Further Web Services interfaces to much of the CiteXplore functionality will be developed to make the bibliographic data that we have aggregated at the EBI available conveniently to third-party information systems, workflows and research tools.

The feasibility of generating citation networks from references cited in scientific articles (eg in PDFs) has been explored recently. We hope that a future CrossRef 'forward-linking' service will provide us with another way of populating this network cleanly. This would naturally fit the browsing function of CiteXplore, to enable users to conveniently navigate cited articles, but would also open opportunities for bibliometric analyses.



Screenshot of a PubMed record in CiteXplore, showing mark-up of proteins, organisms and GO terms in the text, links to the complete article (a free PDF at Blackwell Synergy), and cross-references to UniProtKB and EMBL sequence databases.

Selected references

Rebholz-Schuhmann, D. *et al.* (2007). EBIMed – text crunching to gather facts for proteins from Medline. *Bioinformatics*, 23, e237-e244

EMBL Monterotondo, Italy

The EMBL Monterotondo campus is situated in a green park on the banks of the Tiber river, 20km north of central Rome. The Eternal City is an easy train ride away, and the nearby Lazio countryside features medieval hill towns, spectacular mountain terrain and lakes for hiking, biking, skiing and swimming. EMBL shares the Monterotondo campus with Italian national research groups (IBC-CNR) and the headquarters of the European Mouse Mutant Archive (EMMA).

The EMBL Mouse Biology Unit, which was established to capture new opportunities in mouse functional genomics and to exploit applications of mouse genetic manipulation to biomedical problems, is centred in Monterotondo. Alliances with other European academic research and clinical centres have established EMBL Monterotondo as a hub for the international mouse research community. Participation of the outstation in several EU-wide mouse research and information initiatives serves to link genetics/genomics, phenotyping, physiology and translational features.

Recent expansion of the outstation has retained the original emphasis on developmental mechanisms and extended capabilities to cover adult mammalian physiology and disease. Naturally converging interests between the groups include new collaborative explorations in developmental neurobiology, genetic and pharmacologic manipulations of adult behaviour, inflammation and regenerative processes, stem cell biology and biomedical applications.

A state-of-the-art animal facility provides a full range of mouse transgenic and gene knockout production, rederivation and cryopreservation services, and a fully-equipped behaviour phenotyping suite. The continued refinement of gene regulation *in vivo* is focused on generating more accurate models of human pathologies and multigenic disorders through the development of conditional and inducible mouse mutations. Other centralised core facilities provide a wide array of technologies to the scientific community. Interaction with clinical groups through the EMBL Centre for Disease Mechanisms provides exciting opportunities for application of basic research to advanced disease diagnosis and treatment. A new partnership established with Imperial College London and its associated hospitals adds additional opportunities for translational collaborations.

Development of an EMBL Monterotondo course curriculum, focusing on mouse genetic manipulation in collaboration with local faculty at CNR, EMMA and The Jackson Laboratory, reflects EMBL's long-standing tradition of transmitting new information and expertise. A dynamic seminar series and a visiting researcher programme, together with active collaborations with research groups throughout the world, integrates the science at EMBL Monterotondo with the international mouse biological community.

Nadia Rosenthal
Head of EMBL Monterotondo

Nadia Rosenthal

PhD 1981, Harvard Medical School, USA.
Postdoctoral research at the National Cancer Institute, USA.
Assistant Professor, Boston University Medical Center, USA.
Associate Professor, Massachusetts General Hospital, USA.
Group Leader and Head of EMBL Monterotondo Outstation since 2001.

Regenerative mechanisms in heart and skeletal muscle

Previous and current research

Our laboratory focuses on the biology and enhancement of muscle and cardiovascular regeneration, using mouse as a model. Regeneration is a homeostatic mechanism evolved to maintain or restore the original architecture of a damaged tissue by recapitulating part of its original embryonic development. Despite inevitable physiological changes, most mammalian organs are relatively resilient to age because they can rebuild themselves according to their original body plans, although regenerative capacity is compromised by ageing and chronic disease.

Our approach has been to identify and intervene in the pathways controlling the mammalian response to damage, disease and ageing, by reducing the impediments to effective regeneration. These studies support the feasibility of recapturing regenerative capacity by modulating key signalling pathways to restore injured or degenerating mammalian tissues. In one project, mice expressing locally acting Insulin-like Growth Factor 1 isoforms as muscle-specific transgenes maintain skeletal muscle integrity and ageing, counter muscle decline in degenerative muscle disease, and enhance stem cell homing to sites of damage. Different IGF-1 isoforms invoke distinct responses, suggesting specific mechanisms through which combinations of supplemental growth factors can improve regeneration, and providing new targets for clinical intervention.

Another project explores calcium-activated phosphatase Calcineurin (Cn) that transduces physiological signals through intracellular pathways. A occurring splicing variant CnAb1 acts as a constitutive enzyme, activating its targets in a cyclosporin-insensitive manner. In myoblasts, CnAb1 knockdown activates FoxO-regulated genes, reduces proliferation and induces myoblast differentiation, increasing regenerative activity after muscle injury. This unique mode of action distinguishes the CnAb1 isoform as a candidate for interventional strategies in muscle regeneration.

A third project focuses on nuclear factor-kappa B (NF-κB), a pleiotropic transcription factor that regulates the expression of many genes involved in inflammation, immune response and cell survival. In mice, skeletal muscle-specific ablation of NF-κB action contributes to improved regeneration with minimal scar formation. However mice with cardiac-specific ablation of NF-κB action develop dilated cardiomyopathy, underscoring the deployment of distinct cell type-specific components, and providing new targets for clinical intervention.

Selected references

Mourkioti, F., Kratsios, P., Luedde, T., Song, Y.-H., Delafontaine, P., Adami, R., Parente, V., Bottinelli, R., Pasparakis, M. & Rosenthal, N. (2006). Targeted ablation of IKK2 improves skeletal muscle strength, maintains mass and promotes regeneration. *J. Clin. Invest.,* 116, 2954-2945

Schulze, P.C., Fang, J., Kassik, K., Gannon, J., Cupesi, M., MacGillivray, C., Lee, T., & Rosenthal, N. (2005). Transgenic overexpression of IGF-1 Inhibits skeletal muscle activation of the ubiquitin-proteasome pathway in chronic ventricular dysfunction. *Circ. Res.*, 97, 418-2

Song, Y.H., Li, Y., Du, J., Mitch, W.B., Rosenthal, N. & Delafontaine, P. (2005). Muscle-specific expression of insulin-like growth factor-1 in angiotensin II-infused mice blocks caspase-3 activation, actin cleavage and skeletal muscle wasting. *J Clin Invest.*, 115, 451-8

Mourkioti, F. & Rosenthal, N. (2005). IGF-1 and the immune system: interactions during muscle regeneration. *Trends in Immunology*, 26, 535-42

Rosenthal, N. (2005). Youthful prospects for human stem cell therapy. *EMBO Reports*, 6, S30-4

Cornelius Gross

PhD 1995, Yale University, USA.
Postdoctoral research at Columbia University, USA.
Group Leader at EMBL Monterotondo since 2003.

Developmental programming of anxiety

Previous and current research

Anxiety disorders are debilitating mental illnesses characterised by excessive worry or inappropriate thoughts about responses to or threatening stimuli. Epidemiological studies suggest that both genetic and environmental factors contribute to the prevalence of these disorders. Exposure to adverse events such as trauma, maltreatment or negligence during childhood is known to result in an increased risk for anxiety disorders during adulthood. However, not all persons subjected to such events develop anxiety, and genetic factors are thought to influence the long-term outcome of such experiences. Recently a number of specific genetic polymorphisms have been identified that moderate susceptibility to mental illness following exposure to childhood adversity. However, we know little about the neural circuits and molecular substrates that underlie such gene-by-environment risk factors. A better understanding of the molecular mechanisms involved could lead to novel diagnostic and therapeutic approaches for mental illness in humans.

We are using pharmacological, histochemical, electrophysiological, and behavioural genetic approaches in mice to study how exposure to early adverse rearing can program anxiety-related behaviour in adulthood. We have shown that exposure to low levels of maternal care is associated with increased anxiety and depression-related behaviour in adulthood and that this effect is moderated by specific mutations in genes that are known to play a role in brain development and plasticity. We are using tissue-specific and temporally controlled gene expression technology in transgenic mice to identify the neural circuits and critical time periods for these effects. Until now much of the work in the lab has focused on the role of the serotonergic system in the developing brain as a major determinant of neuronal wiring and excitability in brain circuits controlling anxiety-related behaviour.

Future projects and goals

The major focus of future research will be aimed at understanding how genetic and environmental influences act during development to modulate anxiety behaviour in the mouse. This research will involve a number of approaches, including:

- determination of the molecular mechanisms underlying the serotonin-dependent programming of anxiety and depression-related behaviour by maternal environment;
- identification of novel genes that moderate the maternal programming of anxiety-related behaviour using unbiased genetic methods;
- creation of mouse models of specific human genetic variations that have been associated with anxiety by genetic linkage studies;
- development of pharmaco-genetic transgenic technologies for the tissue and cell-type specific suppression of neural activity in behaving animals.

Together these approaches are aimed at discovering the long-term plastic mechanisms that underlie susceptibility to anxiety. A better understanding of the molecular signals that trigger these plastic changes will allow us to form specific hypotheses about how human anxiety is determined and may lead to improved diagnostic and therapeutic tools in the clinic.

Selected references

Depino, A.M. & Gross, C. (2007). Simultaneous assessment of autonomic function and anxiety-related behavior in BALB/c and C57BL/6 mice. *Behav. Brain Res.*, 177, 254-60

Reuveni, E., Ramensky, V. & Gross, C. (2007). Mouse SNP Miner: An annotated database of mouse functional single nucleotide polymorphisms. *BMC Genomics*, 8, 24

Carola, V., Frazzetto, G. & Gross, C. (2006). Identifying interactions between genes and early environment in the mouse. *Genes Brain Behav.*, 5, 189-199

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*, 416, 396-400

Liliana Minichiello

PhD, University of Naples, Italy, and The National Cancer Institute, NIH, Bethesda,(MD), USA.
Postdoctoral work at EMBL Heidelberg, 1993-1999.
Group Leader at EMBL Monterotondo since 2000.
Visiting Professor, Stem Cell Center, University of Lund, Sweden since 2005.

Signalling mechanisms and gene regulation in the nervous system

Previous and current research

Nerve cells are the elementary signalling units of the nervous system. The structural and functional integrity of the nervous system is controlled by various signals that are in part transduced by transmembrane receptor tyrosine kinases. Extensive studies have established that neurotrophin receptor tyrosine kinases (called Trk) specifically and potently regulate diverse neuronal responses during development.

The main research focus of my laboratory is studying mechanisms of action of Trk receptors in the mouse nervous system. Our primary aim is to correlate the biological functions of neurotrophins with signal transduction events. For example, to investigate the mechanism of TrkB action in hippocampal synaptic plasticity, we have chosen an approach that combines sophisticated genetic tools allowing interference with single phosphorylation sites on a large receptor protein (by generating mice with a targeted mutation in the PLCγ docking site of TrkB or in the Shc site of TrkB receptor) with biochemical and electrophysiological approaches. This provided us with a relatively clean way to tease out specific roles of the various signalling players involved in TrkB-dependent synaptic plasticity.

Moreover, previous *in vitro* studies have characterised the electrophysiological properties and molecular events associated with long-term potentiation (LTP), but as yet there are no *in vivo* data from molecular-level dissection that directly identify LTP as the biological substrate for learning and memory. It is clearly important to understand whether the molecular pathways required for learning are also those generating LTP when measured directly on the relevant circuit of a learning animal, although so far it has proved technically difficult. Using a novel combination of *in vivo* methods we have been able to show that signalling through the TrkB receptor and its PLCγ binding site is important for associative learning and parallel long term potentiation (LTP). These results indicate that the same molecular mechanism forms the basis for learning a task and for changes in synaptic plasticity seen during LTP in awake animals.

Future projects and goals

For further comprehensive understanding of how signalling molecules interlink with each other in the formation of a transductosome downstream TrkB receptor we are employing the co-precipitation/mass spectrometry approach to identify signalling complexes (in particular we are using the TAP-tag strategy, in which a tandem affinity purification tag is inserted into the mouse gene of interest by homologous recombination in ES cells). This allows protein complexes to be directly purified from mouse tissues and subjected to mass spectrometric analysis for the identification of *in vivo* associated proteins.

Current areas of interest include also determining functional and/or morphological feature of GABAergic interneurons. Selective knockouts of key molecules expressed in GABAergic interneurons are fundamental to demonstrate their functional importance. As the neurotrophins and their cognate receptors are expressed also in these cell type, in particular BDNF/TrkB, our aim is to specifically ablate trkB from GABAergic interneuros or from a particular subset of these neurons by the use of the cre-lox system. We have started generating transgenic mice expressing the cre recombinase under specific promoters in bacterial artificial chromosomes (BACs).

We are using a similar approach to understand the *in vivo* relevance of neurotrophins and their cognate receptors in neurodegenerative disorders like Alzheimer's and Huntington's diseases.

Our long term goals are:

- to define the molecular mechanisms by which neurotrophin receptors perform their various functions (such us survival, migration, differentiation, and in particular synaptic plasticity) in the developing and adult vertebrate nervous system;
- to understand the mechanisms underling neuronal diversification;
- investigate the post-transcriptional mechanisms governing the development of the nervous system.

Selected references

Calella, A.M., Nerlov, C., Lopez, R., Sciarretta, C., von Bohlen und Halbach, O., Bereshchenko, O. & Minichiello, L. (2007). Neurotrophin/Trk receptor signalling mediates C/EBPα, -β and NeuroD recruitment to immediate-early (IE) gene promoters in neuronal cells and requires C/EBPs to induce IE gene transcription. *Neural Development*, 2, 4

Gruart, A., Sciarretta, C., Valenzuela-Harrington, M., Delgado-García, J.M. & Minichiello, L. (2007). Mutation at the TrkB-PLCγ docking site affects hippocampal LTP and associative learning in conscious mice. *Learning & Memory*, 14, 54-62

Medina, D.L., Sciarretta, C., Calella, A.M., Von Bohlen und Halbach, O., Unsicker, K. & Minichiello, L. (2004). TrkB regulates neocortex formation through the Shc/PLCγ-mediated control of neuronal migration. *EMBO J.*, 23, 3803-14

Minichiello, L., Calella, A.M., Medina, L.D., Bonhoeffer, T., Klein, R., & Korte, M. (2002). Mechanism of TrkB-mediated hippocampal long-term potentiation. *Neuron*, 36, 121-137

Claus Nerlov

PhD 1995, the University of Copenhagen, Denmark.
Postdoctoral research at EMBL Heidelberg.
Head of Laboratory of Gene Therapy Research, Copenhagen University Hospital, Denmark.
Group Leader at EMBL since 2001. Visiting Professor, Stem Cell Center, University of Lund, Sweden from 2005.

Transcription factor function in development, physiology and disease

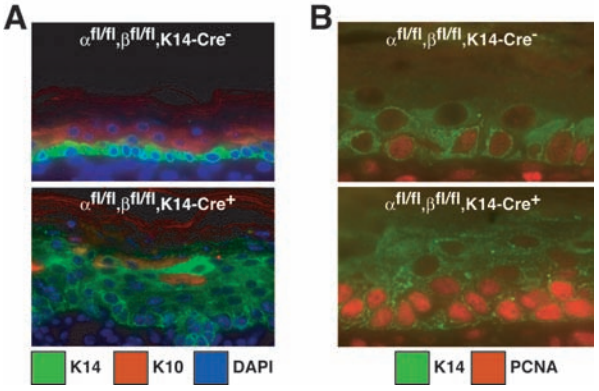
Transcription factors play important roles in regulation of cellular proliferation, differentiation and in the function of fully differentiated cells. We look at how transcription factors link the processes of cellular differentiation and self-renewal/proliferation, both at the stem cell level and during terminal cellular differentiation. We use mouse genetics to study *in vivo* the C/EBP family of basic regionleucine zipper transcription factors, proteins which play essential roles in the development of the hematopoietic system, adipose tissues, epithelia and granulosa cells. We use conditional mutagenesis to delete one or more C/EBPs from specific cell types, and point mutagenesis to specifically alter C/EBP protein-protein interactions or post-translational modifications.

Previous and current research

We previously defined E2F repression and interaction with the SWI/SNF complex as C/EBPα functions essential for adipose and myeloid differentiation (Pedersen *et al.*, 2001: *Genes & Dev.* 15; Porse *et al.*, 2001: *Cell,* 107). We have now found that E2F repression is required also for myeloid tumour suppression, as mice homozygous for mutations that disable C/EBPα-E2F interaction have increased myeloid progenitor proliferation and develop an acute myeloid leukemia (AML)-like disease (Porse *et al.*, 2005). In contrast, we did not observe any effect on progenitor proliferation upon deletion of the Cdk2/Cdk4 interaction domain of C/EBPα (Porse *et al.*, 2006). Mutations in the gene encoding C/EBPα are found in AML patients, and the most common type results in specific loss of the 42kDa C/EBPα isoform (p42), while preserving expression of the 30kDa isoform (p30). Only p42 has E2F repression activity, and when we generated p42 knockout mice we found that also these developed AML. We are now generating knockin mice with other AML-derived mutations, and investigating the role of C/EBPα mutations in the formation of self-renewing leukemic stem cells. We are currently investigating the role of C/EBPs in keratinocytes, and have observed a similar role for C/EBPs in the transition from proliferation to differentiation; however, in this case C/EBPα and C/EBPβ function redundantly, and removal of both is necessary to cause keratinocyte hyperproliferation and impair their differentiation (see figure). Other main projects involve studying the role of post-translational modifications of C/EBPs in metabolism and macrophage activation, and the transcriptional regulation of C/EBPβ.

Future projects and goals

The future focus of the laboratory will be to elucidate the molecular mechanisms by which C/EBPs control differentiation in non-hematopoietic tissues (neurons, skin, liver), and to determine the signalling pathways that regulate C/EBPs through post-translational modification. A major effort will deal with the effects of leukemogenic mutations on hematopoietic stem cell function, in order to determine how malignant, self-renewing tumour stem cells arise.



Defective keratinocyte differentiation and epidermal hyperplasia upon conditional ablation of C/EBPα and C/EBPβ expression in the skin. A) Skin sections from control mice (carrying floxed alleles of both C/EBPα and β, but not the Keratin 14-Cre transgene; upper panel) and C/EBPα/β double knockout mice (lower panel. As above, but with the K14-Cre transgene). The sections were stained with antibodies against the basal cell marker (Keratin 14, green), a marker for differentiating keratinocytes (Keratin 10, red) and the DNA stain DAPI (blue). The lower boundary of the epidermis is defined by the Keratin 14 positive basal cells; the outer surface of the skin is at the top of the panels. Note the expansion of the Keratin 14 expression domain and diminished Keratin 10 expression in double knockout mice. B) Skin sections as in A), stained with antibodies against Keratin 14 (green) and Proliferating Cell Nuclear Antigen (PCNA, a marker of proliferating cells; red). Note the expanded domain of cell proliferation in double knockout mice.

Selected references

Kirstetter, P., Anderson, K., Porse, B.T., Jacobsen, S.E. & Nerlov, C. (2006). Activation of the canonical Wnt pathway leads to loss of hematopoietic stem cell repopulation and multilineage differentiation block. *Nat Immunol.*, 7, 1048-1056

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/EBPα is dispensable for C/EBPα-mediated growth regulation *in vivo*. *Mol. Cell Biol.*, 26, 1028-1037

Porse, B.T., Bryder, D., Theilgaard-Monch, K., Hasemann, M.S., Anderson, K., Damgaard, I., Jacobsen, S.E. & Nerlov, C. (2005). Loss of C/EBPα cell cycle control increases myeloid progenitor proliferation and transforms the neutrophil granulocyte lineage. *J. Exp. Med.*, 202, 85-96

Nerlov, C. (2004). C/EBPα mutations in acute myeloid leukaemias. *Nat. Rev. Cancer*, 4, 394-400

Walter Witke

PhD 1991, University of Munich, Germany.
Postdoctoral research at the Brigham and Women's Hospital, Harvard Medical School, USA.
Group Leader at EMBL since 1996.

Role of the actin cytoskeleton in cell migration and mouse physiology

Previous and current research

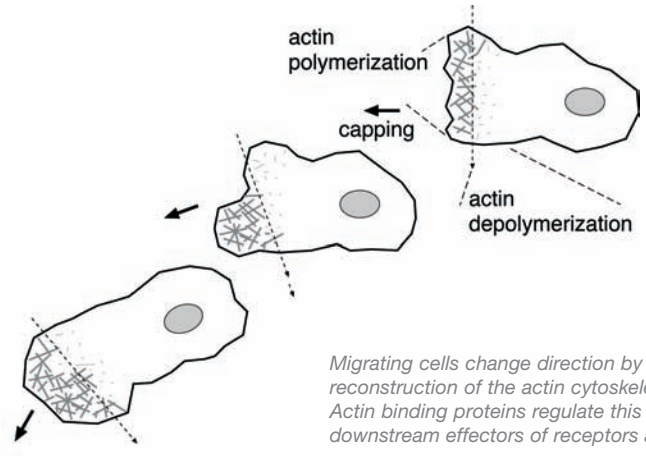
The actin cytoskeleton is the key structure for regulating cell motility, membrane trafficking, cell polarity and cytokinesis. How the actin filament system participates in all these processes is a fundamental question in cell biology. In mammals the situation is even more complex, because different cell types have different requirements in terms of migration and motility. Furthermore, cell migration and cytokinesis during embryonic development have a different quality than, for example, cell migration in adult animals. While the first aims to build structured tissues from cells, the latter type of motility rather relates to aspects of the animal's physiology.

To serve these different requirements, cells need mechanisms to regulate the dynamics of actin filaments, and to link actin polymerisation to a specific cellular processes. A large number of actin binding proteins have been characterised by their activity to modulate actin polymerisation and filament alignment. Our group is focusing on two protein families which are thought to be essential for remodelling of the actin cytoskeleton *in vivo*. The first group of proteins – the profilins – sequester monomeric actin, catalyse the nucleotide exchange on actin monomers, and finally deliver the monomers to the growing ends of actin filaments. In the mouse, three profilin genes were identified. The second group of proteins – the actin depolymerising factors – display the opposite activity. These proteins recognise old actin filaments and depolymerise the filaments from the slow growing end. In mouse these activities are provided by the actin depolymerising factors Gelsolin, Cofilin and ADF. The main interest of our group is to understand the role of these proteins for regulating cell crawling during embryonic developments as well as later in the adult animal in the brain and the immune system.

Future projects and goals

Or results on cofilin/ADF strongly suggest a role of actin dynamics in cell fate determination. Symmetric versus asymmetric cell division is critical for cell homeostasis during organ development and data from the various tissue specific knockouts of cofilin/ADF imply an essential function of these actin binding proteins for the balance of symmetric/asymmetric division. Future work will focus on the mechanisms by which cofilin/ADF utilises actin for controlling the cell cycle and whether cofilin/ADF plays a role in tumour progression.

In the brain the actin cytoskeleton plays a critical role in synapse formation and function. Our data show that mutations in cofilin/ADF can lead to lissencephaly-like symptoms in the mouse due to defects in neuronal migration in the cortex, while profilin 2 is having a different role in controlling neurotransmitter release and behaviour. Future projects will address the role of cofilin/ADF and profilins in dendritic spine plasticity, memory and learning.



Migrating cells change direction by a coordinated breakdown and reconstruction of the actin cytoskeleton at the leading edge. Actin binding proteins regulate this highly dynamic process as downstream effectors of receptors and signalling cascades.

Selected references

Aidinis, V., Carninci, P., Armaka, M., Witke, W., Harokopos, V., Pavelka, N., Koczan, D., Argyropoulos, C., Thwin, M.M., Moller, S., Kazunori, W., Gopalakrishnakone, P., Ricciardi-Castagnoli, P., Thiesen, H.J., Hayashizaki, Y. & Kollias, G. (2005). Cytoskeletal rearrangements in synovial fibroblasts as a novel pathophysiological determinant of modeled rheumatoid arthritis. *PLoS Genet.* 1, e48

Gareus, R., Di Nardo, A., Rybin, V. & Witke, W. (2006). Mouse profilin 2 regulates endocytosis and competes with SH3 ligand binding to dynamin 1. *J. Biol. Chem.* 281, 2803-2811

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

Witke, W. (2004). The role of profilin complexes in cell motility and other cellular processes. *Trends Cell Biol.* 14, 461-469

Index

Notes

Notes

Notes