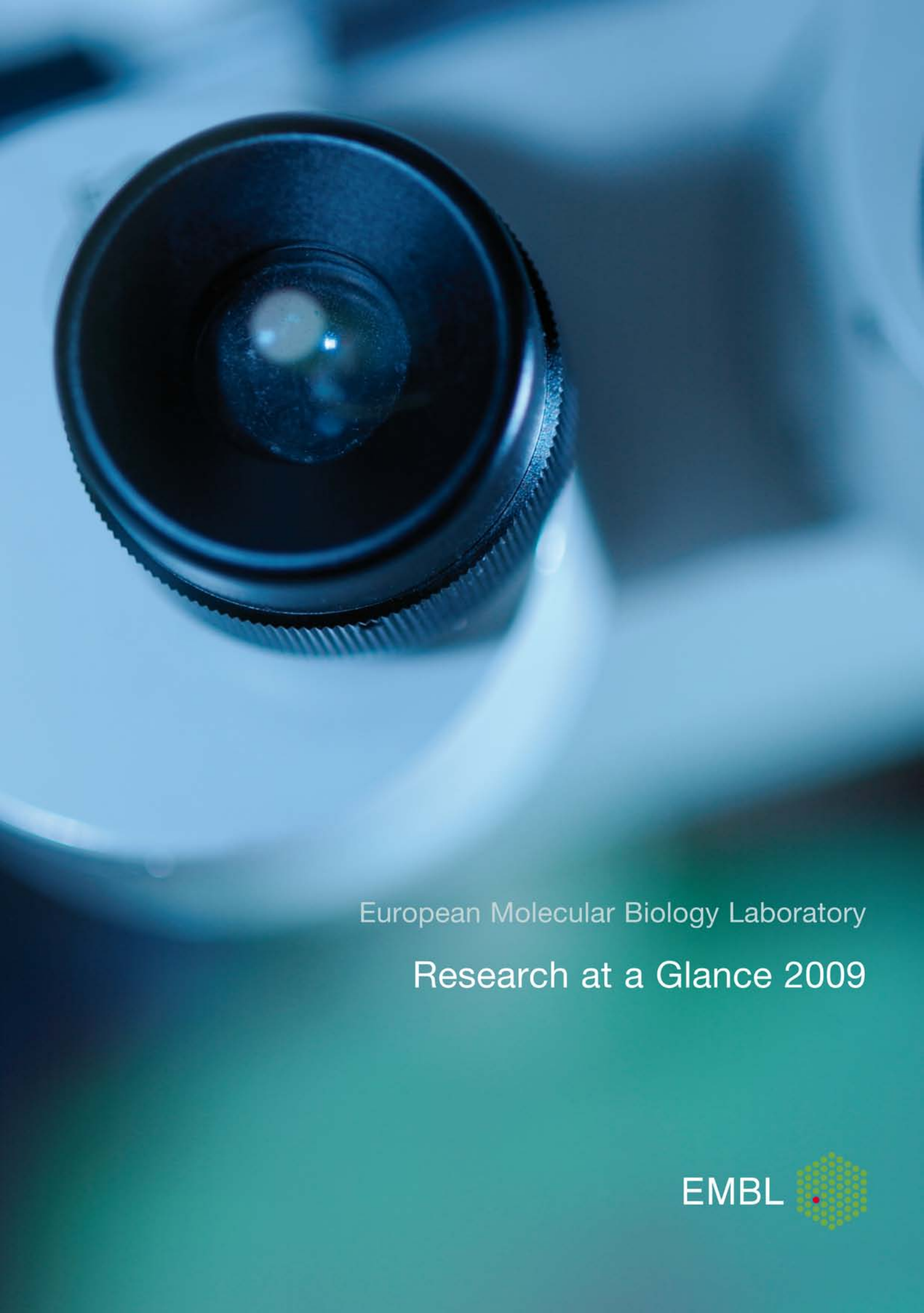


European Molecular Biology Laboratory

Research at a Glance 2009

EMBL

European Molecular Biology Laboratory

Research at a Glance 2009

EMBL

Contents

Foreword by EMBL's Director General

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

The vision of the nations which 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, and its member states now number twenty-one, including the first associate member state, Australia. 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 for 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 informs the public about the impact modern biology has 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.

But what really distinguishes EMBL is the large number of inter-unit collaborations, bringing people with common interests but distinct expertise together to tackle ambitious projects. Cross-unit networking and training 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 140,000 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: Gene Expression, Cell Biology and Biophysics, Developmental Biology, Structural and Computational Biology and Directors' Research, as well as the Core Facilities and the central administration, from which service functions are provided for the use of staff at all five EMBL sites. Heidelberg is also home to EMBLEM, the laboratory's technology transfer company.

Today more than 900 personnel 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. Integrated in the EMBL campus the newly-built Advanced Training Centre (ATC) will host state-of-the-art training facilities for practical courses and computer labs together with a 450-seat auditorium, setting the scene for a new area of scientific conferences at EMBL promoting advanced scientific training and education in Europe.

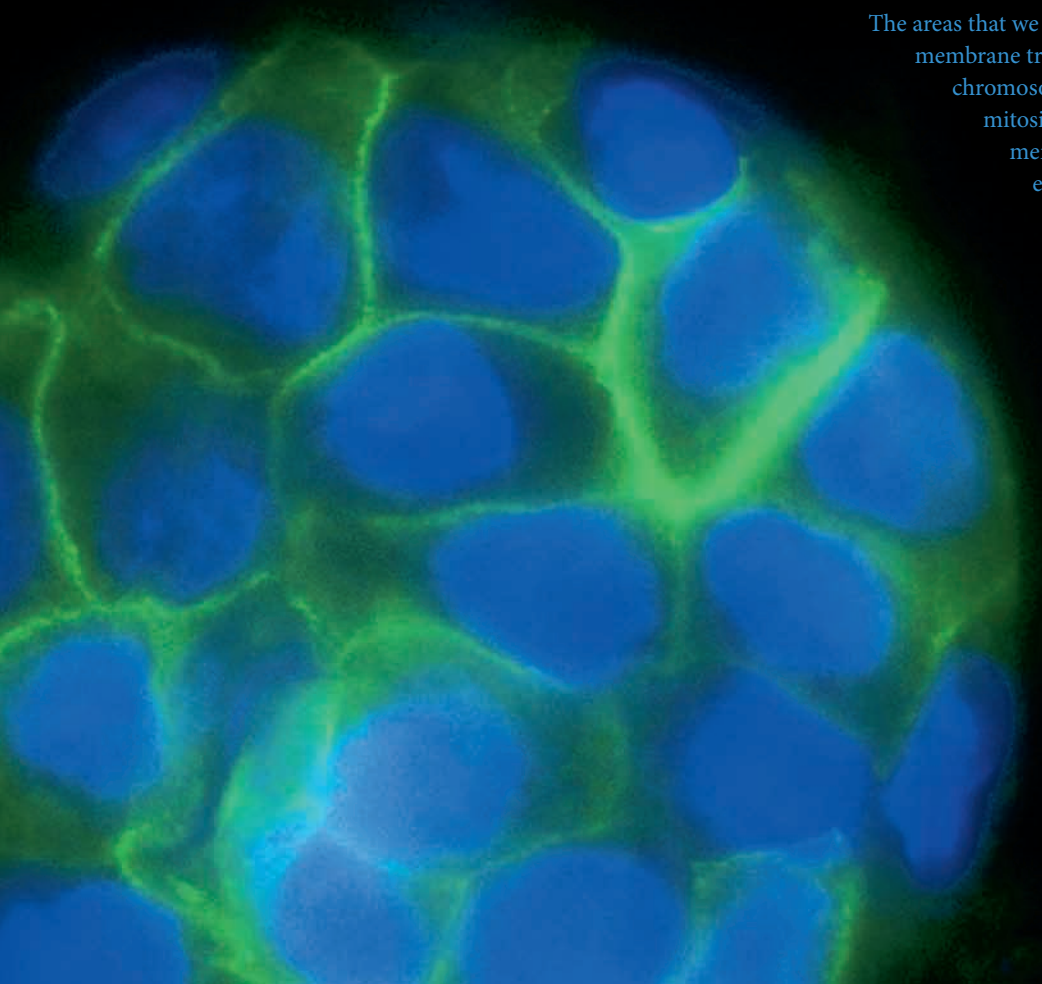
Cell Biology and Biophysics Unit

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 and function.

All cells (including prokaryotes) are divided into functional domains, each with different molecular compositions. In addition, eukaryotes have compartments such as the nucleus, the cytoskeleton and the endomembrane system. These compartments are permanently renewed by mechanisms that are still poorly understood.

Research in the Cell Biology and Biophysics Unit focuses on the mechanisms and principles that underlie the organisation and function of these different compartments and the distribution of specific molecules to each cellular sub-system. Cell biologists and physicists at EMBL are therefore trying to define the role of targeting events, as well as that of more complex self-organisation processes in organising cellular space. These principles are best understood at transitions when the organisation of the cell undergoes dramatic changes to carry out new functions. This is the case when cells divide, or when they change their fate during the development of the organism to form specific tissues and organs. Both opportunities are exploited in the unit.

As a cell prepares to divide, all the microtubules suddenly depolymerise to reassemble into the mitotic spindle. At the same time, the nucleus is disassembled, mitotic chromosomes are formed, the Golgi complex fragments and membrane traffic ceases. After segregation of the genome is achieved, cellular organisation is re-established. Thus every cell cycle provides the opportunity to study the principles of the biogenesis of cellular compartments. Similarly, during development, when progenitor cells differentiate into new cell types, not only do the daughter cells receive a complement of chromosomes and organelles from the parent cell, but the genetic program is changed. A reorganisation of cellular architecture takes place, guided by rules that we begin to unravel. The elucidation of such rules and principles is a major challenge to contemporary biology.



The areas that we are presently concentrating on are membrane trafficking, cytoskeletal networks and chromosomes and the nucleus and their role in mitosis and meiosis as well as in development. New directions are therefore being explored at the interface between cell and developmental biology to understand how the cell organisation and collective cell behaviour leads to organ formation. Physicists and chemists working together with biologists are trying to elucidate the fundamental rules that govern dynamic cell organisation and function while developing new instruments and tools. Novel developments in microscopy and computer simulations are a particular strength of the unit.

Jan Ellenberg and Eric Karsenti
Joint Coordinators, Cell Biology and Biophysics Unit

Jan Ellenberg

PhD 1998, Freie Universität Berlin.

Postdoctoral research at the Cell Biology and Metabolism Branch, NICHD, NIH, Bethesda.

Group leader at EMBL since 1999.

Head of Gene Expression Unit since 2006. Joint Unit Coordinator of Cell Biology and Biophysics Unit since 2009.

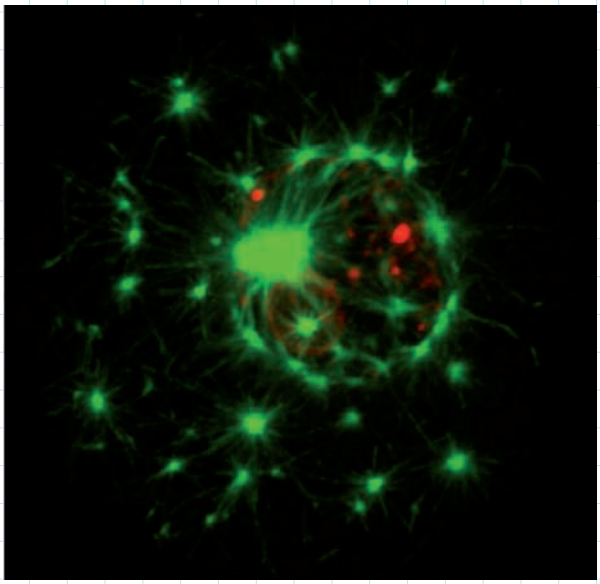
Functional dynamics of nuclear structure during the cell cycle

Previous and current research

The genome of eukaryotic cells is compartmentalised inside the nucleus, delimited by the nuclear envelope (NE) whose double membranes are continuous with the endoplasmatic reticulum (ER) and stabilised by the nuclear lamina filament meshwork. The NE is perforated by nuclear pore complexes (NPCs), which allow selective traffic between nucleus and cytoplasm. In M-phase, most metazoan cells reversibly 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 structural proteins and their regulators using advanced fluorescence microscopy methods coupled with computerised image processing and simulations to extract biophysical parameters and build mechanistic models.

In the past, we could define the ER as the reservoir and means of partitioning for nuclear membrane proteins in mitosis and found that nuclear breakdown is triggered by disassembly of the NPC and then further facilitated by microtubule mediated tearing of the nuclear lamina. During the meiotic division of starfish oocytes, we could show that long-range chromosome motion after nuclear breakdown is driven by actin filaments. In mouse oocytes this occurs only after formation of an acentrosomal spindle, which we could show assembles by self-organisation of cytoplasmic microtubule asters. In mitotic cells, we have analysed chromosome dynamics during their segregation and could show that their overall spatial arrangement is transmitted through mitosis and that their maximal compaction is reached only at the end of anaphase, just before nuclear reformation.



Acentriolar microtubule organising centres (green) form a 3D network around the chromosomes (red) during spindle assembly in a mouse oocyte

Future projects and goals

The objective of our future work is to gain further mechanistic insight into nuclear remodelling in live cells. In particular, we are focussing 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. As a first step, we have developed high-throughput live cell imaging in combination with RNAi screening to identify novel genes that function in the above cell division processes.

Selected references

Dultz, E., Zanin, E., Wurzenberger, C., Braun, M., Rabut, G., Sironi, L. & Ellenberg, J. (2008). Systematic kinetic analysis of mitotic dis- and reassembly of the nuclear pore in living cells. *J. Cell Biol.,* 180, 857-65

Schuh, M. & Ellenberg, J. (2008). A new model for asymmetric spindle positioning in mouse oocytes. *Curr. Biol.,* 18, 1986-92

Schuh, M. & Ellenberg, J. (2007). Self-organization of MTOCs replaces centrosome function during acentrosomal spindle assembly in live mouse oocytes. *Cell,* 13, 484-98

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, 385-90

Self-organisation principles in cell morphogenesis

Eric Karsenti

PhD 1979, University of Paris.

Postdoctoral research at University of California, San Francisco.

Group leader at EMBL since 1985.

Unit Coordinator since 1998. Joint Unit Coordinator since 2009.

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 and forming reaction networks with specific topologies; 2) reaction-diffusion processes that define steady-state spatial distributions of regulatory molecules; and 3) the collective behaviour of cytoskeletal systems that self-organise in large structures. These principles may feed back to each other to generate precise steady state or evolving dissipative structures with specific patterns.

Last year we managed to demonstrate how complex cascades of reaction-diffusion processes actually control microtubule dynamics around chromosomes. Using chromatin spots micro-patterned on coverslips, an engineering procedure developed in collaboration with François Nédélec's group (page 17), and centrosomes added to frog egg extracts, we directly measured the shape and extent of the gradient of aster asymmetry around chromosomes. Using reaction-diffusion equations and the program Cytosim, also developed by François Nédélec, we were able to find which type of reaction network topology could generate a gradient of microtubule dynamic instability that fits the shape and extent of the observed gradient of aster asymmetry (Chaitanya Athale, Ana Dinarina).

We have also identified one of the enzymes probably involved in this network. This is a Cyclin-dependent kinase named CDK11, which is essential for the asymmetric growth of microtubules towards chromosomes in mitosis (Hideki Yokoyama).

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.

We have also initiated a new line of research with the group of Darren Gilmour (page 12), which has to do with the role of reaction-diffusion coupled to structural events in organogenesis. During fish development, a group of cells that derive from the primordial ear starts to migrate on each side of the fish to finally form the lateral line composed of a fixed number of sensory organs that feel water pressure. Each organ contains sensory cells coupled to a nerve cell. Sensory cells contain cilia that, like in the ear, feel water flows. The question is, how do these organs form while the slug migrates during embryogenesis, how cells interchange from mesenchymal-like to epithelial like in the migrating slug, how is the number of organs determined in each line and how is the distance between organs fixed?

This is a typical problem of pattern formation. This can be taken from the genetic angle or from a physico-chemical point of view. Genetics will tell us which genes are involved in the process, physics and chemistry how they work together to generate a pattern. A pattern is a physics problem, not a genetics problem. I therefore hired a physics student, Sebastian Streichan, to work on this in collaboration with the groups of Darren Gilmour and Lars Hufnagel (page 14). Sebastian is developing new methods to acquire dynamic and physical data about the experimental process, while developing a reaction-diffusion model of regulators of cell organisation coupled to a physical model of cell behaviour, with promising results so far.

Selected references

Athale, C.A., Dinarina, A., Mora-Coral, M., Pugieux, C., Nédélec, F. & Karsenti, E. (2008). Regulation of microtubule dynamics by reaction cascades around chromosomes. *Science*, 322, 1243-1247

Karsenti, E. (2008). Self-organization in cell biology: a brief history. *Nat. Rev. Mol. Cell Biol.*, 9, 255-262

Yokoyama, H., Gruss, O.J., Rybina, S., Caudron, M., Schelder, M., Wilm, M., Mattaj, I.W. & Karsenti, E. (2008). Cdk11 is a RanGTP-dependent microtubule stabilization factor that regulates spindle assembly rate. *J. Cell Biol.*, 180, 867-875

Caudron, M., Bunt, G., Bastiaens, P. & Karsenti, E. (2005). Spatial coordination of spindle assembly by chromosome-mediated signaling gradients. *Science*, 309, 1373-1376

Claude Antony

PhD 1984, Université Paris VI.

Postdoctoral research at EMBL 1987-1989.

Group Leader at CNRS 1994-2003.

Facility head and team leader at EMBL since 2003.

Cell cytoskeleton: an electron tomography analysis

Previous and current research

In January 2008 we established a new electron tomography set-up, which will be invaluable for future innovative projects at EMBL. Our focus of interest is the organisation of the fission yeast and budding yeast microtubular cytoskeleton using electron tomography, which not only allows the reconstruction, modelling and quantifying of subcellular elements, but also enables the visualisation of a number of structural features in very fine detail.

Having performed a detailed analysis of the wild type fission yeast cytoskeleton array (Höög *et al.*, 2007), we have carried out investigations on EB1/mal3 and CLIP170/tip1 MAPs (microtubule associated proteins) mutants which regulate microtubule dynamics, and also on microtubule bundling factors.

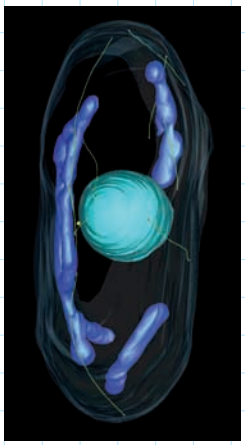
Plus end tracking protein mutants: Mal3 deletion mutant displays disordered and shorter microtubule bundles in the cytoplasm, although the number of microtubules within these bundles is similar to that in wt cells. There is no apparent antiparallel overlapping bundle and this deletion also causes alteration to the spindle pole body (SPB). The deletion of Tip1 causes severe alterations as the total microtubule polymer length is considerably reduced and a loss of SPB-microtubule connection is observed. Our data suggest that Tip1p may play a role in microtubule nucleation which is supported by abnormally thin filaments, numerous in the deletion strain.

We also studied microtubule regrowth after depolymerisation of the whole microtubular array using the carbendazim microtubule depolymerising drug (MBC). Upon drug wash-out we captured the early stages of microtubule bundle reassembly and observed detailed features of the growing microtubules. This analysis was carried out in collaboration with the Brunner group (opposite), and will be published soon.

Microtubule bundling factors: To understand inter-microtubule bonds, we reconstructed microtubule arrays in deletion strains which affect the bundling function. Ase1 (encoding a non-motor homodimer protein) deletion strain shows a loss of parallel organisation of the bundles despite some overlapping area where microtubules remain associated but with a reduced inter-microtubule spacing. The deletion of Klp2 (a kinesin-like protein which slides microtubules along pre-existing ones in a minus end directed manner) shows mostly bundles composed of two microtubules. In the double deletion ase1Δ/klp2Δ a microtubule overlap region still persists but in a parallel orientation. SPB-associated bundles are missing. In all these mutants strains inter-microtubule bridges can be detected, suggesting that other bundling factors still prevail.

Future projects and goals

A new project has been started in collaboration with the Knop (page 16) and Kaksonen (page 15) groups concerning the reconstruction of the microtubule bundles involved in the karyogamy process in the budding yeast mating pathway by using tomography and microtubule modelling. We will also investigate the cytoskeleton architecture in shmooing cells (polarisation process of mating cells). This will involve the analysis of both microtubule and actin bundles in shmooing cells and their connection with the plasma membrane and the SPB.



In conjunction with in-house and external research groups we are now starting a major project to reconstruct the *Xenopus* mitotic spindle. Cryofixation and preparation of samples for tomography acquisition are just the first steps. The large-scale reconstruction of such a huge structure, and detailed parts of it, will be performed using extensive montaging and joining of tomograms. In the course of this project we hope to elucidate the spindle microtubule architecture at fine resolution and in doing so, derive information about microtubule polarity in the midzone of the spindle. A precise analysis of the organisation of the spindle poles should also be carried out. A student and a postdoc are currently working on this long-term project.

3D model of Ase1 deletion strain showing the disorganised microtubule bundles (green), SPB (yellow), mitochondria (blue) and nucleus (turquoise). The plasma membrane and cell wall appear as semi-transparent in grey. (Picture by Helio Roque, postdoc).

Selected references

Colombelli, J., Tangemo, C., Haselman, U., Antony, C., Stelzer, E.H., Pepperkok, R. & Reynaud, E.G. (2008). A correlative light and electron microscopy method based on laser micropatterning and etching. *Methods Mol. Biol.*, 457, 203-13

Höög, J.L. & Antony, C. (2007). Whole-cell investigation of microtubule cytoskeleton architecture by electron tomography. *Methods Cell Biol.*, 79, 145-167

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

Toya, M., Sato, M., Haselmann, U., Asakawa, K., Brunner, D., Antony, C. & Toda, T. (2007). Gamma-tubulin complex-mediated anchoring of spindle microtubules to spindle-pole bodies requires Msd1 in fission yeast. *Nat. Cell Biol.*, 9, 646-653

Cell morphogenesis and spatial microtubule organisation

Damian Brunner

PhD 1995, University of Zürich.

Postdoctoral research at the Imperial Cancer Research Fund, London.

Group leader at EMBL since 2000. Joint appointment with the Developmental Biology Unit.

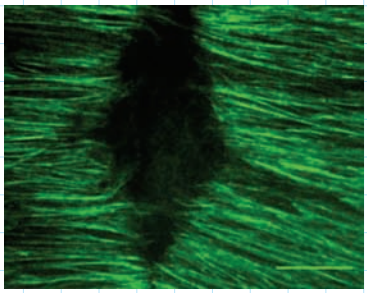
Previous and current research

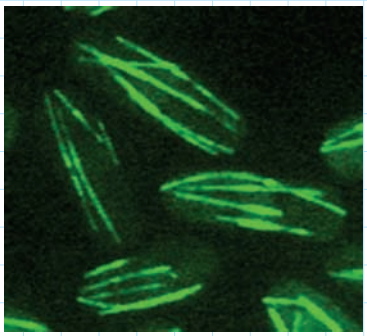
To create a defined morphology, cells need to polarise and correctly orient their polarity axis. Both processes require a defined intracellular order based on the specific sub-cellular arrangement of actin and microtubule filaments. Our investigations focus on the contribution of microtubules, whose 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 (figure 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.

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. 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, University of Colorado at Boulder), we discovered that Mal3p also binds and stabilises the microtubule lattice seam, explaining its weak localisation all along the microtubules. 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 reorganised 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 (figure 2). Surprisingly, these microtubules are essential exclusively for the final step, the fusion of the epithelium.





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 seven 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 (top): Interphase microtubule bundles in fission yeast cells. Figure 2 (bottom): Microtubule bundles in the epidermal cells during dorsal closure in Drosophila melanogaster.

Selected references

Foethke, D., Makushok, T., Brunner, D. & Nédélec, F. (2009). Force and length-dependent catastrophe activities explain interphase microtubule organization in fission yeast. *Molecular Systems Biology*, 5, 241

Bieling, P., Laan, L., Schek, H., Munteanu, E.L., Sandblad, L., Dogterom, M., Brunner, D. & Surrey, T. (2007). Reconstitution of a microtubule plus-end tracking system *in vitro*. *Nature* 45, 1100-115

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

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

Darren Gilmour

PhD 1996, Cambridge University.

Postdoctoral research at the Max Planck Institute for Developmental Biology, Tübingen.

Group leader at EMBL since 2004.

The role of collective cell migration during organ morphogenesis

Previous and current research

Morphogenesis is the generation of complex biological form through coordinated changes in the size, shape and positioning of groups of cells. The guided migration of cohesive groups of cells is a hallmark of embryonic morphogenesis. While such collective migrations determine the shape of most organ systems and are a common feature of wound repair, regeneration and cancer, they are still poorly understood.

The zebrafish lateral line primordium is a migrating cluster of some two hundred cells whose function is to generate and disperse mechanosensory organs throughout the embryonic skin. Cells in this moving tissue must multitask – they migrate, grow, divide and differentiate simultaneously. The lateral therefore provides a powerful model system for addressing how complex form arises through the interplay of basic cellular behaviours. In recent years we have developed a number of *in vivo* imaging and perturbation tools that allow this entire morphogenetic process to be addressed at sub-cellular resolution in the context of the intact, living embryo.

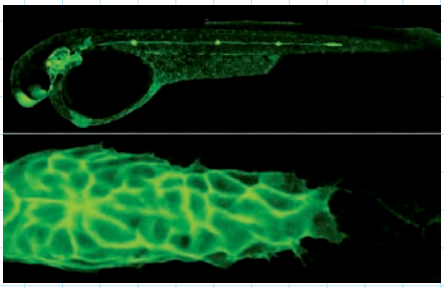


Figure 1: The zebrafish migrating lateral line organ allows collective migration to be easily studied in vivo.

Genetic screens have lead to the isolation of a number of signalling molecules required for primordium migration. The primordium is guided by the chemokine Sdf1 and its receptor Cxcr4, a signalling pathway that is known to regulate the invasive behaviour of many human tumours. Furthermore, cells within the primordium are assembled into rosette-like organ progenitors via a dynamic mesenchymal-epithelial transition that is driven through spots of FGF-ligand that repeatedly appear within the tissue as it migrates.

Future projects and goals

Our aim is to understand how changes in cell migration and morphology spread across moving tissues during organogenesis. We are developing quantitative imaging methods that allow us to precisely measure the activity of Cxcr4/Sdf1, FGF and other key chemical signalling systems with the aim of elucidating how local changes in activity drive differences in cell behaviour. As these signalling systems exert their effect via the cytoskeleton and cell cortex, we are also using a complementary, 'bottom-up' approach that addresses how local changes cytoskeletal dynamics regulate cell-cell interactions within tissues. Using biophysical tools such as laser ablation in combination with advanced 3D imaging, we hope to address the role of mechanical forces in coordinating cell behaviour. These quantitative data are being used to support the formulation of mathematically models that will accurately simulate this complex *in vivo* morphogenesis process.

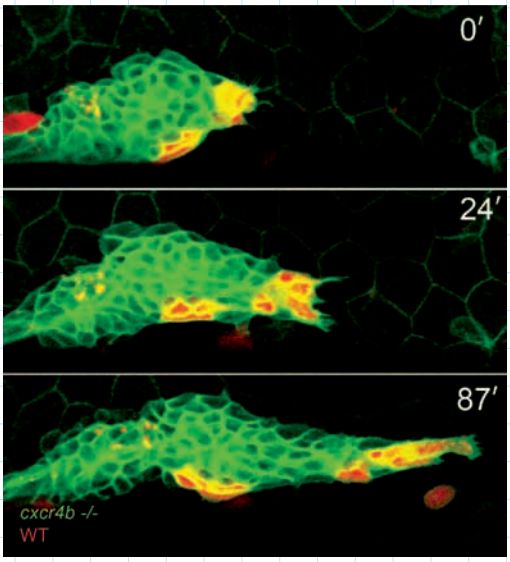


Figure 2: Transplanted wild-type cells (red) rescue the migration of cxcr4 mutant primordia (green).

Selected references

Lecaudey, V., Cakan-Akdogan, G., Norton, W.H. & Gilmour, D. (2008). Dynamic Fgf signaling couples morphogenesis and migration in the zebrafish lateral line primordium. *Development*, 135, 2695-2705

Pouthas, F., Girard, P., Lecaudey, V., Ly, T.B., Gilmour, D., Boulin, C., Pepperkok, R. & Reynaud, E.G. (2008). In migrating cells, the Golgi complex and the position of the centrosome depend on geometrical constraints of the substratum. *J. Cell Sci.*, 121, 2406-2414

Valentin, G., Haas, P. & Gilmour, D. (2007). The chemokine SDF1a coordinates tissue migration through the spatially restricted activation of Cxcr7 and Cxcr4b. *Curr. Biol,* 17, 1026-1031

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

Chromosome structure and dynamics

Christian Häring

PhD 2003, IMP Vienna.

Postdoctoral work at the University of Oxford.

Group leader at EMBL since 2007.

Previous and current research

Chromosomes undergo enormous changes over the course of the cell cycle. DNA replication generates two identical copies of every chromosome, the so-called sister chromatids, which remain tightly connected with each other. As cells get ready to divide, sister chromatid pairs individualise into compact rod-shaped structures and their kinetochores attach to the mitotic spindle. Once all sister kinetochores have attached in a bipolar fashion, the connection between sister chromatids is released to trigger their segregation towards opposite poles, followed by cytokinesis and chromosome decompaction. This process ensures that every daughter cell inherits a complete set of chromosomes. Errors during chromosome segregation lead to aneuploidy, a hallmark of most cancer cells and the leading cause for spontaneous miscarriages.

Even though the formation and segregation of mitotic chromosomes was first observed more than 125 years ago, the underlying mechanisms are still poorly understood. The overall aim of our research is to gain insight into the action of molecular machines that organise chromosomes prior to and during cell divisions. Recent research has identified two multi-subunit protein complexes called cohesin and condensin as central players in shaping and segregating chromosomes. While cohesin is holding sister chromatids together, condensin is a key component in maintaining chromatids in a stable compact form.

Both complexes are built of heterodimers of structural maintenance of chromosomes (SMC) and kleisin subunits that associate with additional proteins. The discovery that cohesin's kleisin subunit Scc1 connects the ABC ATPase domains of its Smc1 and Smc3 subunits to form a gigantic ring structure suggests that it might hold sister chromatids together by entrapping both sisters inside its ring (figure 1). Condensin might act similarly by entrapping different regions of the same chromatid within a ring structure to form loops of chromatin (figure 2).

We are investigating the molecular mechanisms of cohesin and condensin function using a combination of biochemistry, molecular biology, cell biology and, in collaboration, chemical and structural biology. In an independent project, we are exploring novel approaches to identify additional players that direct the formation of mitotic and meiotic chromosomes. For most of our studies we take advantage of the versatility of the budding and fission yeast model systems. The high degree of conservation between not only the proteins involved but also the general principles behind structuring chromosomes makes it very likely that discoveries made will be of universal significance for all eukaryotes.

Future projects and goals

Our major goal is to elucidate the fundamental molecular mechanics behind the organisation of mitotic chromosomes on different levels. We will initially focus on the following three questions:

- How does the condensin complex bind to chromosomes, how does it function on chromosomes, and how is its activity controlled?
- How does the interplay of condensin with DNA and other chromosomal proteins ultimately shape a mitotic or meiotic chromosome?
- What other key components are required for making a mitotic or meiotic chromosome?

Figure 1: Model of the cohesin ring holding sister chromatids together.

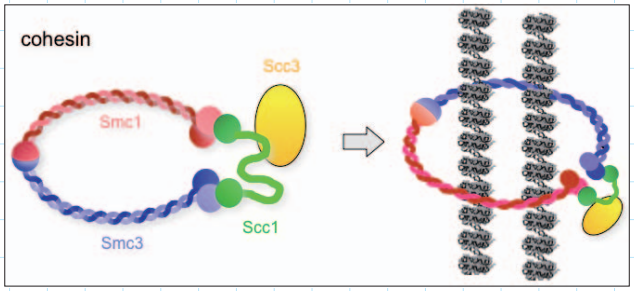
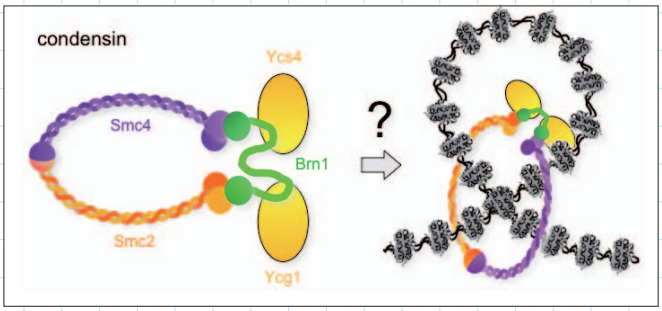


Figure 2: Model of the condensin ring structuring chromosomes.



Selected references

Haering, C.H., Farcas, A.M., Arumugam, P., Metson, J. & Nasmyth, K. (2008). The cohesin ring concatenates sister DNA molecules. *Nature*, 454, 297-301

Nasmyth, K. & Haering, C.H. (2005). The structure and function of SMC and kleisin complexes. *Annu. Rev. Biochem.*, 74, 595-648

Haering, C.H., Schoffnegger, D., Nishino, T., Helmhart, W., Nasmyth, K. & Lowe, J. (2004). Structure and stability of cohesin's Smc1-kleisin interaction. *Mol. Cell*, 15, 951-964

Gruber, S., Haering, C.H. & Nasmyth, K. (2003). Chromosomal cohesin forms a ring. *Cell*, 112, 765-777

Lars Hufnagel

PhD 2001, MPI for Dynamics and Self-Organisation, Göttingen.

Postdoctoral research at the Kavli Institute for Theoretical Physics, Santa Barbara, California.

Group leader at EMBL Heidelberg since 2007.

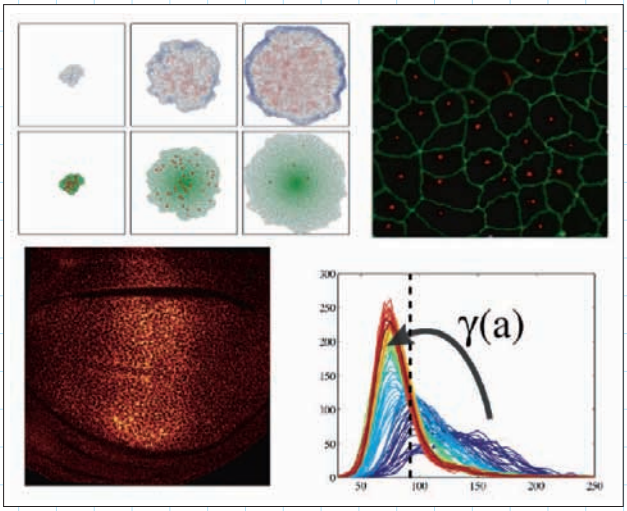
Dynamics of cell growth and tissue architecture

Previous and current research

Two fundamental processes must occur concurrently in tissues during animal development. Firstly, tissues must grow rapidly to generate the final adult size of the organism, and cells have to stop growing and dividing once the final size is reached. Secondly, the tissue needs to be specified and patterned with each cell adopting the appropriate fate and gene expression profile for its position. Both processes are intrinsically connected and need to be coordinated. Central to the formation of a tissue is the establishment, maintenance and remodelling of complex cell-cell interactions that supply mechanical integrity and stability. Tissue growth is a highly dynamic and heterogeneous process. It involves many spatial and temporal scales, and for a deeper understanding one has to integrate information on a single cell level with cell-cell interactions and population effects.

We have recently investigated the interplay between the growth of *Drosophila* wing imaginal discs and the formation of the Dpp morphogen gradient. Our results suggest a new scenario of size determination, where disc size is determined relative to the fixed morphogen distribution. Our model shows that a feedback of mechanical stress on cell growth can compensate for non-uniform distributions of growth-stimulating morphogens and insures uniform growth throughout the disc. Furthermore, we have formulated and analysed a model describing the interaction of morphogens with glypicans and have compared its prediction to measurements of the effect of glypican Dally-like (Dlp) overexpression on Wingless (Wg) morphogen signalling in *Drosophila* wing imaginal discs. The model explains the opposing effect that Dlp overexpression has on Wg signalling in the distal and proximal regions of the disc. Our model suggests that Dlp acts by allowing Wg to diffuse on cell surface while protecting it from loss and degradation, and that Dlp, rather than acting as Wg co-receptor, competes with receptors for morphogen binding.

Currently, we are investigating the role of mechanical constraints on cell growth, apoptosis, orientation of division, intra-tissue rearrangements and cell differentiation.



Bridging the scales from a single cell to the whole tissue by combining cell culture and organ growth experiments with modelling.

Future projects and goals

Our research interests are focussed on the control and regulation of cell proliferation, apoptosis and cellular rearrangement processes in developing tissues, with a specific emphasis on epithelial tissues and the role of mechanical interactions as a regulator.

We seek to characterise and quantify the spatiotemporal effects of mechanical stress, deformations and fluid flow-induced sheer stress on cell growth, gene expression and cellular polarity in two-dimensional epithelial tissues. To address this issue, we pursue an interdisciplinary approach combining classical biological techniques with detailed modelling methods from various fields, ranging from statistical physics to applied mathematics and computer science. Our research also relies on novel microscopy methods in conjunction with the development of sophisticated image analysis tools. Furthermore, the group continues its current research on *Drosophila* wing development and has a specific interest in the spread of pathogens in epithelial tissues.

Selected references

Brockmann, D. & Hufnagel, L. (2007). Front propagation in reaction-superdiffusion dynamics: taming Lévy flights with fluctuations. *Phys. Rev. Lett.*, 98, 178301

Hufnagel, L., Teleman, A.A., Rouault, H., Cohen, S.M. & Shraiman, B.I. (2007). On the mechanism of wing size determination in fly development. *Proc. Natl. Acad. Sci. USA*, 104, 3835-3840

Brockmann, D., Hufnagel, L. & Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439, 462-465.

Hufnagel, L., Kreuger, J., Cohen, S.M. & Shraiman, B.I. (2006). On the role of glypicans in the process of morphogen gradient formation. *Dev. Biol.*, 300, 512-522.

Brockmann, D., Hufnagel, L. & Geisel, T. (2005). Dynamics of Modern Epidemics. In 'SARS: A Case Study in Emerging Infections', McLean, A., May, R., Pattison, J. & Weiss, R. (eds), Oxford University Press

Hufnagel, L., Brockmann, D. & Geisel, T. (2004). Forecast and control of epidemics in a globalized world. *PNAS*, 101, 15124–15129

Dynamics of membrane trafficking

Marko Kaksonen

PhD 2002, University of Helsinki.

Postdoctoral research at the University of California, Berkeley.

Group leader at EMBL since 2006.

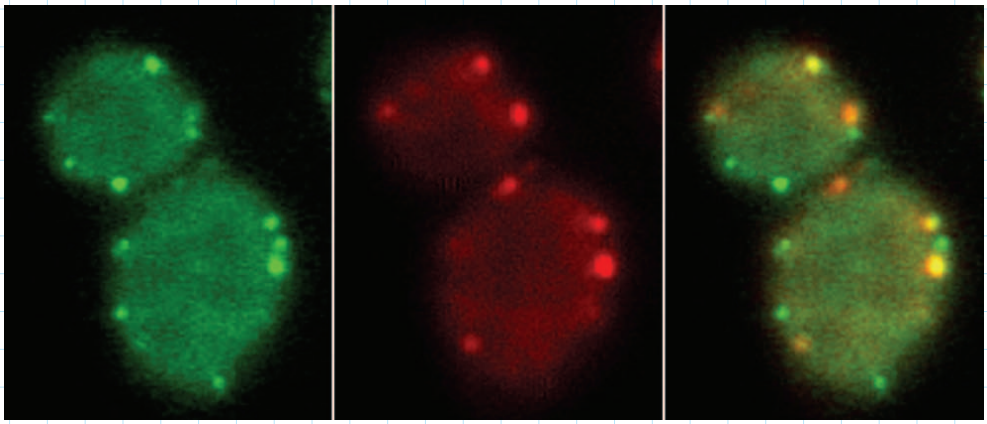
Previous and current research

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, with which 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.

Postdoctoral research at the MPI for Biochemistry, Munich and the Beatson Institute for Cancer Research, Glasgow.

Group leader at the MPI for Biochemistry, Munich.

Group leader at EMBL since 2001.

Systems biology of meiosis and mating in budding yeast

Previous and current research

Our group is interested in the various cellular processes that underlie the sexual cycle of budding yeast (mating and meiosis). In the past we have addressed the meiosis specific pathways that regulate spore morphogenesis with respect to spindle pole body function, membrane formation and morphogenesis and cytokinesis (figure 1). We mainly focussed on the processes that regulate spore morphogenesis in comparison to cell division by bud formation. Among other things, we concentrated on the regulation of spindle pole function in controlling vesicle fusion and in the initiation of spore morphogenesis and on membrane shaping of the spore.

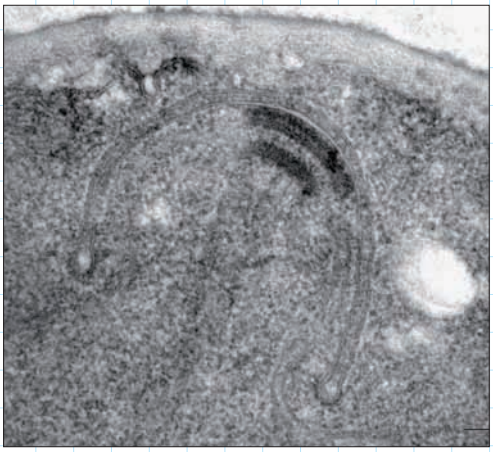


Figure 1: Electron micrograph of a forming spore. The picture shows a spindle pole body (SPB) that is in progress of forming a spore membrane.

Mating is another important aspect of the life cycle of yeast. How do yeast cells find a mating partner? We study the MAP kinase signal transduction pathway that underlies signal transduction during mating. We established Fluorescence (Cross-) Correlation Spectroscopy (FCCS) and FLIM (fluorescence lifetime imaging, figure 2) to work with yeast cells. These new quantitative imaging methods enable us to measure protein complex formation and to visualise the activity of the MAP kinases. This yields important new insights into the dynamics and the spatial organisation of the signalling process.

Future projects and goals

We continue to use quantitative microscopy approaches and subsequently expand our investigation to three interconnected MAP kinase signalling pathways by using semi-high throughput screening microscopy to quantify protein concentration, protein-protein interaction and protein localisation of all the major components involved. We consider both quiescent and signalling conditions. The goal is to enhance our understanding of the spatial and dynamic organisation of the signalling processes. This will help us to derive and further develop quantitative models of the processes that regulate signalling through theses pathways.

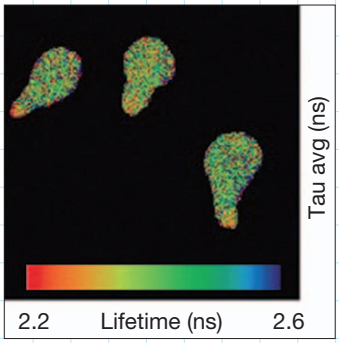


Figure 2: High relative Fus3 MAP kinase activity in the mating projection (shmoo) of pheromone stimulated yeast cells. Fus3 activity was detected using FLIM (in collaboration with Mark Hink and Philippe Bastiaens).

Our work on meiosis has gradually shifted to questions that relate to the role and function of genome recombination in meiosis. As a model, we use computer simulations of population of yeast-like genomes that undergo yeast-like life cycles. Here we address the role of meiosis and recombination and the impact of genome architecture on handling deleterious mutational load. To complement these approaches, we use yeast as a model for experimental evolutionary studies where we address the consequences of random mutations on fitness, and on the role of meiosis and recombination to purge deleterious load.

Furthermore, we study a novel yeast species with similar live-cycle properties as *S. cerevisiae*, but which has one notable and most interesting difference: this species appears not to recombine its genome during meiosis. We use genome sequencing and experimental approaches to address how this species performs meiosis I and to understand the impact of absent recombination on the evolution of the genome.

Selected references

Maier, P., Rathfelder, N., Maeder, C.I., Colombelli, J., Stelzer, E.H.K. & Knop, M. (2008). The SpoMBe pathway drives membrane bending necessary for cytokinesis and spore formation in yeast meiosis. *EMBO J.*, 27, 2363-2374

Maeder, C.I., Hink, M.A., Kinkhabwala, A., Mayr, R., Bastiaens, P.I. & Knop, M. (2007). Spatial regulation of Fus3 MAP kinase activity through a reaction-diffusion mechanism in yeast pheromone signalling. *Nat. Cell Biol.*, 9, 1319-1326

Maier, P., Rathfelder, N., Finkbeiner, M.G., Taxis, C., Mazza, M., Panse, S.L., 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

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

Cellular architecture

François Nédélec

PhD 1998, Université Paris 11.

Postdoctoral research at EMBL.

BioMS group leader since 2002.

Joint appointment with the Structural and Computational Biology Unit.

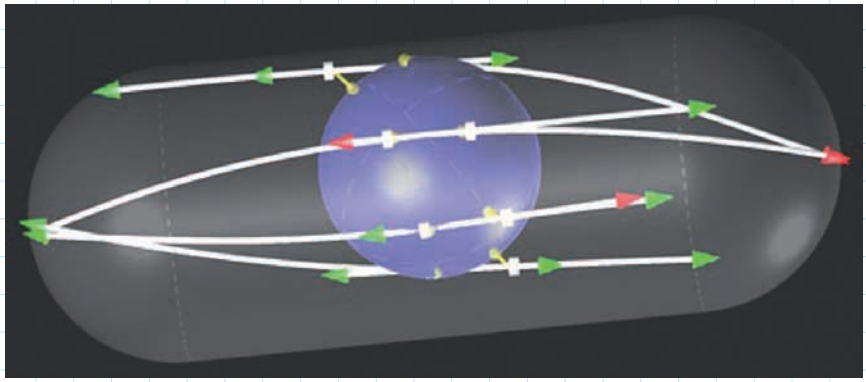
Previous and current research

Modern microscopy has shown us the dynamic nature of biological organisation. During cell division, for example, chromosome segregation 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 structure which is solid and yet highly dynamic at the same time: the main components – microtubules – are in rapid turnover. They 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 precisely the balanced forces necessary to position and segregate the chromosomes.

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 fulfil its biological function with the utmost accuracy?

Obviously, some kind of averaging is going on, but deciphering how multiple elements (proteins) contribute to a system's properties is not straightforward. It is a challenging problem for many reasons: 1) there are many different types of protein implicated; 2) elements are not present in so many copies, such as to allow a simple statistical averaging; and 3) most of their interactions are dynamic and sometimes poorly characterised.

Within the field of the cytoskeleton, we address these aspects 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. They are implemented in a simulation called *cytosim*, which is being applied to diverse problems of cytoskeletal organisation. Simulations are often used to validate or refute pre-existing ideas, but they can also be used in a more creative way: one can generate systematically various properties for the molecules, and automatically test their ability to form stable structures. The analysis of successful scenarios leads to the formulation of hypotheses, which can later be tested experimentally.



Simulation of the microtubule cytoskeleton in the fission yeast S. pombe.

Future projects and goals

We will study systems in which experiments and theory can be synergistically combined. We currently focus on chromosome-microtubule interactions using *Xenopus* egg extracts, and experimental system in which many parts of mitosis can be recapitulated. We are generally interested in modelling cellular processes in which the cytoskeleton serves a major role, such as the different stages of mitosis, the generation of cell shape in *S. pombe*, and the generation of asymmetry during cell division.

Selected references

Athale, C.A., Dinarina, A., Mora-Coral, M., Pugieux, C., Nédélec, F. & Karsenti, E. (2008). Regulation of Microtubule Dynamics by Reaction Cascades Around Chromosomes. *Science*, 322, 1243-1247

Jékely, G., Colombelli, J., Hausen, H., Guy, K., Stelzer, E., Nédélec, F. & Arendt, D. (2008). Mechanism of phototaxis in marine zooplankton. *Nature*, 456, 395-399

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

Kozlowski, C., Srayko, M. & Nédélec, Fr. (2007). Cortical microtubule contacts position the spindle in *C. elegans* embryos. *Cell*, 129, 499-510

Rainer Pepperkok

PhD 1992, University of Kaiserslautern.

Postdoctoral work at University of Geneva.

Lab head at the Imperial Cancer Research Fund, London.

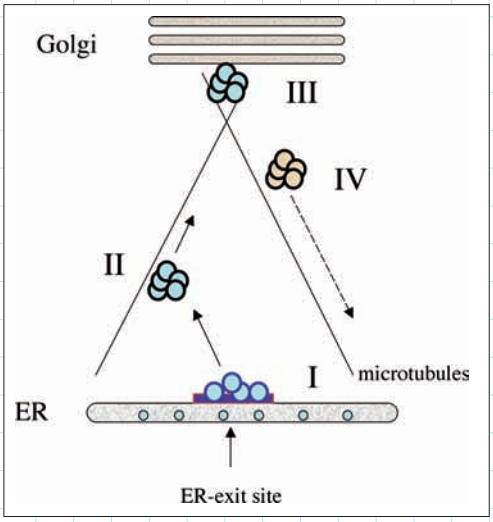
Team leader at EMBL since 1998.

Membrane traffic in the early secretory pathway

Previous and current research

Transport between the endoplasmic reticulum (ER) and the Golgi complex in mammalian cells involves at least four basic steps (see figure): 1) biogenesis of membrane bounded transport carriers at specialised domains (ER-exit sites) of the ER; 2) microtubule mediated transport of the carriers to the Golgi complex; 3) docking and fusion of the carriers with the Golgi complex; and 4) recycling of the transport machinery back to the ER. 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 and coordinated at the molecular level.

The specific questions we are presently addressing in this context are: 1) what are the mechanisms underlying the regulation of ER-exit sites biogenesis and function; 2) how are ER exit and microtubule mediated ER to Golgi transport coupled at the molecular level; 3) what are the mechanisms of Golgi biogenesis; and 4) which are the molecules regulating recycling of Golgi resident proteins to the ER.



The four steps involved in ER to Golgi transport in mammalian cells. (I): Biogenesis of COPII coated vesicles occurs at specialised ER exit sites of the ER. (II): COPII vesicles homotypically fuse to form larger vesicular tubular transport carriers (VTCs) that are transported to the Golgi complex along microtubules. (III): VTCs arrive at the Golgi complex and fuse to it to deliver their cargo. (IV): Transport machinery and misrouted proteins are return back to the ER by a distinct class of carriers.

To investigate this, we develop computer automated light microscopy approaches to directly visualise and quantify 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 mechanistic of the temporal and spatial regulation of the molecular interactions involved. Our combined data suggest that secretory cargo, lipids and the microtubule motor associated dynactin complex play a critical role in the stabilisation of the COPII vesicular coat complex to provide the time that is necessary for cargo selection and concentration at ER exit sites. In order to investigate the mechanisms of Golgi biogenesis we have developed an approach, in which we remove by laser nanosurgery the entire Golgi complex from living cells and subsequently analyse the "Golgi-less" karyoplast by time-lapse and electron microscopy. With this approach we could show that Golgi biogenesis in mammalian cells occurs *de novo* from ER derived membranes.

In order to identify putative molecules involved in this *de novo* Golgi biogenesis, we have developed and applied functional assays to assess the effect of knock-ins by cDNA over-expression and knockdowns by RNAi, on processes such as constitutive protein transport, Golgi integrity and function of vesicular coat complexes. To achieve the throughput that such genome-wide analyses require we have developed a fully automated high content screening microscopy platform including sample preparation, image acquisition and automated analysis of complex cellular phenotypes. We have applied this technology to genome-wide siRNA screens to identify and characterise comprehensively the genes and their underlying functional networks involved in secretory membrane traffic and Golgi integrity.

Future projects and goals

We will study the novel proteins, which we revealed in our screens to be involved in the early secretory pathway, in further detail at the systems level. 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

Simpson, J.C., Cetin, C., Erfle, H., Joggerst, B., Liebel, U., Ellenberg, J. & Pepperkok, R. (2007). An RNAi screening platform to identify secretion machinery in mammalian cells. *J. Biotechnol.*, 129, 352-365

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

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

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

Chemical cell biology

Carsten Schultz

PhD 1989, University of Bremen.

Postdoctoral research at the University of California, San Diego.

Habilitation 1997, Organic Chemistry, University of Bremen.

Group leader, MPI for Mol. Physiology, Dortmund.

Group leader at EMBL since 2001.

Previous and current research

Before joining EMBL, our research focussed on finding novel ways to stimulate chloride and water secretion of epithelial cells to help with the genetic disease cystic fibrosis (CF). Our compounds helped to investigate some of the underlying intracellular signalling pathways and provided drug candidates to eventually treat CF patients. In particular, we developed chemical methods to convert highly polar signalling molecules such as cyclic nucleotides, inositol phosphates and phosphoinositides to membrane-permeant, bioactivatable derivatives ('prodrugs').

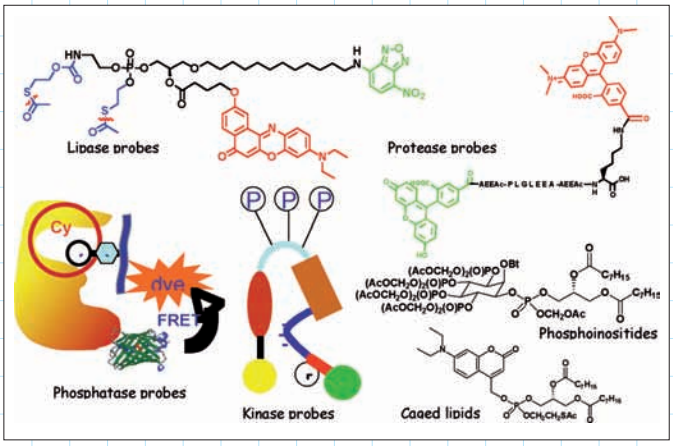
At EMBL, we are more interested in the basic signalling network underlying epithelial secretion. We developed a wide range of fluorescent reporter molecules, either genetically encoded or as small molecule fluorescent probes (see figure). With these sensors, we hope to provide a more complete picture of the signalling network and to help find compounds that might be beneficial for CF patients. The function of the probes is based on FRET or translocation and is suitable for imaging with spatial and temporal resolution. Currently, we use the approaches in Multiparameter Imaging, where 5-6 cellular events are monitored simultaneously (Piljić & Schultz, 2008a). In addition, we introduced a novel method to monitor the formation of enzyme-substrate complexes in living cells (Piljić & Schultz, 2008b). The effort is supported by a unique approach to model intracellular signalling networks. The imaging abilities are essential to validate these models and to support the emerging efforts towards systems biology at EMBL.

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 compounds in CF mouse. Small molecule fluorescent FRET probes are prepared to study intra- and extracellular enzyme activities with a focus on phospholipases and proteases, such as a probe to monitor matrix metallo proteinase 12 (MMP12) activity on the surface of macrophages, an enzyme crucial in the development of lung emphysema.

Future projects and goals

In 2009, our group joins the Cell Biology and Biophysics Unit at EMBL. In continuation of our previous efforts, we will focus predominantly on lipid signalling and lipid-controlled cell biology. To examine the effect of phospholipids, i.e. phosphoinositides, on endocytosis, we are preparing membrane- permeant phospholipids to specifically increase cellular phosphoinositide levels in a non-disruptive way. Very recently, we succeeded in synthesising photoactivatable derivatives to provide an even more controlled way for manipulating lipid levels in living cells. Vesicle trafficking and endocytosis is investigated in collaboration with the group of Rainer Pepperkok (opposite). To visualise lipid locations, we introduced the first method to fluorescently label lipids in living cells (Neef & Schultz, 2009).

Finally, we are interested in how the plasma membrane is repaired after physical impact, for which we combine fluorescence microcopy of tagged proteins with electron microscopy (correlative microscopy), the latter in collaboration with Claude Antony (page 10).



Several reporter and modulator molecules developed in our lab, including small molecule sensors for lipases and proteases, genetically encoded reporters for kinase and phosphatase activities, membrane-permeant and photoactivatable lipid molecules as well as lipid derivatives that can be fluorescently labelled in living cells.

Most projects rely on organic chemistry to produce the tools described above. The group therefore has a significant number of preparative chemists at the graduate student and postdoc level. The symbiosis of chemistry, biochemistry, and cell biology opens new doors and grants novel insights into how cells are functioning.

Selected references

Neef, A.B. & Schultz, C. (2009). Selective fluorescence labeling of lipids in living cells. *Angew. Chem. Int. Ed. Engl.*, 48, 1498-500

Jost, C.A., Reither, G., Hoffmann, C. & Schultz, C. (2008). Contribution of fluorophores to protein kinase C FRET probe performance. *Chembiochem.*, 9, 1379-1384

Piljic, A. & Schultz, C. (2008). Analysis of protein complex hierarchy in living cells. *ACS Chem. Biol.*, 3, 749-755

Piljic, A. & Schultz, C. (2008). Simultaneous recording of multiple cellular events by FRET. *ACS Chem. Biol.*, 3, 156-160

Skrahina, T., Piljic, A. & Schultz, C. (2008). Heterogeneity and timing of translocation and membrane-mediated assembly of different annexins. *Exp. Cell Res.*, 314, 1039-1047

Ernst H. K. Stelzer

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.

Optical nanotechnologies for relevant physiological approaches to a modern biology

Previous and current research

Modern biophotonics provides many technologies that operate in a nanodomain. The resolution of optical microscopes is in the range of 100nm, the precision of optical tweezers is a single nm, and laser-based nanoscalpels generate incisions 300nm wide and, in three dimensions, cause severing that is barely 700nm deep. Extremely efficient light microscopes require nanowatts of power to induce fluorescence emission.

Although many modern technologies could operate in 3D, they are mainly applied in a cellular context that is defined by hard and flat surfaces. On the other hand, it is well known that relevant physiological information requires the geometry, mechanical properties, media flux and biochemistry of a cell's context found in living tissues. A physiological context excludes single cells on cover slips. It is found in more complex 3D cell structures.

However, the observation and the optical manipulation of thick and optically dense biological specimens suffer from two severe problems: 1) the specimens tend to scatter and absorb light, so the delivery of the probing light and the collection of the signal light both become inefficient; 2) many biochemical compounds (most of them non-fluorescent) absorb light, suffer degradation of some sort and induce malfunction or even death.

The group develops and applies technologies for the observation of large and complex 3D biological specimens as a function of time. The technology of choice is the optical light sheet, which is fed into a specimen from the side and observed at an angle of 90° to the illumination optical axis. The focal volumes of the detection system and of the light sheet overlap. True optical sectioning and dramatically reduced 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), take advantage of modern camera technology and are compatible with essentially every contrast and specimen manipulation tool found in modern light microscopes.

Future projects and goals

It is our medium-term goal to integrate the optical nanotechnologies developed during the past years into our light sheet-based fluorescence microscopes (LSFM) and to apply them to complex biological objects.

We developed a technological basis that integrates LSFM 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.

Selected references

Keller, P.J., Pampaloni, F., Lattanzi, G. & Stelzer, E.H. (2008). Three-dimensional microtubule behavior in *Xenopus* egg extracts reveals four dynamic states and state-dependent elastic properties. *Biophys. J.*, 95, 1474-1486

Keller, P.J., Schmidt, A.D., Wittbrodt, J. & Stelzer, E.H.K. (2008). Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science*, 322, 1065-9

Keller, P.J., Pampaloni, F. & Stelzer, E.H.K. (2007). Three-dimensional preparation and imaging reveal intrinsic microtubule properties. *Nat. Methods*, 4, 843-846

Pampaloni, F., Reynaud, E.G. & Stelzer, E.H.K. (2007). The third dimension bridges the gap between cell culture and live tissue. *Nat. Rev. Mol. Cell Biol.*, 8, 839-845

Physical systems biochemistry of cytoskeleton dynamics and function

Thomas Surrey

PhD 1995, Eberhard-Karls University, Tübingen.

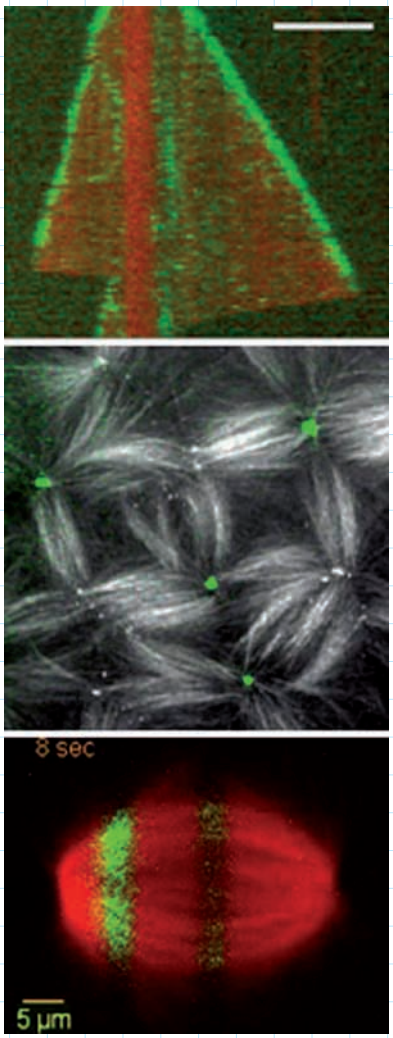
Postdoctoral research at Princeton University, USA and EMBL.

Staff Scientist 2001-2002.

Team leader at EMBL since 2002. Group leader since 2006.

Previous and current research

The cytoskeleton is responsible for the internal organisation of eukaryotic cells. Microtubules, motor proteins and associated proteins form a mechano-chemical network that determines the dynamic and adaptable nature of intracellular order. But how the collective behaviour of various differently moving motors and competing regulators of microtubule dynamics leads to specific organisations of the cytoskeleton is not understood. How do single molecules move in cells? What role does spatio-temporal control of activities play in the correct functioning of motor/microtubule networks? Can we construct minimal systems *in vitro* that display complex network dynamics with defined functionalities? And does such a synthetic approach help us to understand what is special about the functioning of mechano-chemical systems distant from thermodynamic equilibrium?



We address these questions using a combination of advanced light microscopy, biochemistry and quantitative cell biology. Our aim is to understand the behaviour of dynamic systems based on measured molecular properties. Therefore, we have studied how single fluorescently-labelled motors behave on single microtubules populated with competing molecules (Telley *et al.*, 2009, *Biophys. J.*). We have measured the movements of motors in intact mitotic spindles and have investigated how the biophysical properties of an essential mitotic motor are regulated by a kinase in its physiological context. We believe that *in vitro* reconstitutions of dynamic cytoskeleton behaviour from a minimal set of dynamically interacting proteins is a powerful approach for the dissection of systems behaviour. Microtubule end-tracking and self-organisation of networks consisting of microtubules and different motors (Surrey *et al.,* 2001, *Science*) are examples where system dynamics can be understood based on biochemical reconstitution combined with quantitative analysis.

Future projects and goals

In the future, we will continue to measure the biophysical properties of motors and microtubules both in their physiological context and *in vitro,* aiming at connecting single molecule physics with systems behaviour. We will develop tools that will allow us monitor and manipulate the spatio-temporal regulation of protein activities using chemical biology approaches in combination with advanced light microscopy. We will continue to generate more and more complex dynamic systems *in vitro* and to dissect their functions at a molecular level. Examples are microtubule end-tracking networks, mitotic spindles and cytoskeleton-membrane systems. Our goal is to understand how biological function of protein interaction networks is generated from the coordinated and regulated dynamic interactions of their components. In summary, we are interested in elucidating the design principles underlying intracellular organisation and dynamics using a combination of top-down and bottom-up approaches.

Top: time-space plot of microtubule end tracking (Bieling et al., *2007,* Nature*). Centre: self-organised network of microtubules and plus- and minus-motors (Surrey* et al., *2001,* Science*). Bottom: spindle with locally photoactivated motors (Uteng* et al., *2008,* J. Cell Biol.*).*

Selected references

Bieling, P., Kandels-Lewis, S., Telley, I.A., van Dijk, J., Janke, C. & Surrey, T. (2008). CLIP-170 tracks growing microtubule ends by dynamically recognizing composite EB1/tubulin-binding sites. *J. Cell Biol.*, 183, 1223-33

Uteng, M., Hentrich, C., Miura, K., Bieling, P. & Surrey, T. (2008). Poleward transport of Eg5 by dynein-dynactin in *Xenopus laevis* egg extract spindles. *J. Cell Biol.,* 182, 715-26

Bieling, P., Laan, L., Schek, H., Munteanu, E.L., Sandblad, L., Dogterom, M., Brunner, D. & Surrey, T. (2007). Reconstitution of a microtubule plus-end tracking system *in vitro*. *Nature*, 450, 1100-5

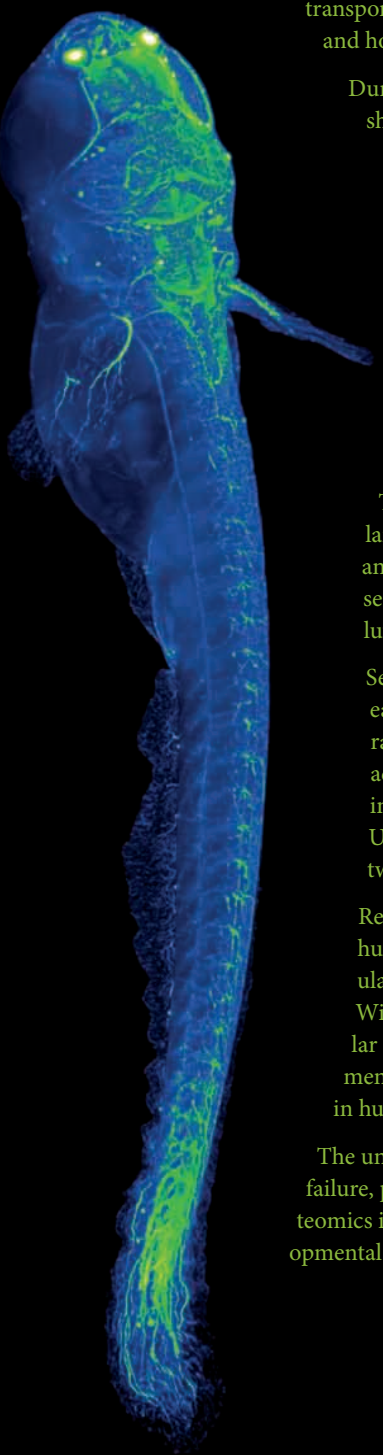
Surrey, T., Nédélec, F., Leibler, S. & Karsenti, E. (2001). Physical properties determining self-organization of motors and microtubules. *Science*, 292, 1167-71

Developmental Biology Unit

The development of living organisms requires the integration and precise coordination of all basic cellular and molecular processes in space and time. Live organisms are the physical manifestation of complex regulatory networks interacting with their environment.

Research in the Developmental Biology Unit is aimed at elucidating the basic principles and mechanisms underlying fundamental developmental processes, such as cell fate-specification and polarity, tissue morphogenesis, organogenesis and growth control. A major goal is to understand the regulatory cascades – hierarchies of gene expression choices – that control developmental decisions. Using selected animal and plant model organisms, our groups combine genetics, biochemistry, bioinformatics, high-throughput genomics, proteomics, and imaging to understand how cellular and molecular processes evolved and are coordinated in living organisms.

Cell polarity underlies many fundamental decisions in development, both in plants and animals. In many organisms, the first developmental events occur before the onset of zygotic transcription, under the control of by mRNAs and proteins asymmetrically localised in the egg cell. Understanding the mechanisms underlying cell polarisation, mRNA localisation and translational control in *Drosophila* development is a topic of research in the unit. Understanding the polarised transport of auxin in plants, which determines the positioning of lateral organs, such as leaves and flowers, and how this molecule specifies different cell types is also topic of research.



During development, progenitor cells are amplified and differentiate into tissues of characteristic shape and function. The expression of many differentiation factors is required for these morphological changes. Research in the unit aims to elucidate how cells in the early *Drosophila* embryo reorganize their content in response to the expression of key developmental transcription factors and, specifically, how tissue-specific gene expression controls patterns of protein and membrane trafficking, and how this trafficking regulates cell fate and behaviour.

Elucidating the temporal organisation, or timing, of embryonic development is another aim of research in the unit. Using the mouse model, the mechanisms controlling overall developmental rate at an organismal level, as well as the timing of individual patterning processes, including the dynamics of underlying signaling pathways, are being investigated. Analysis of novel mouse reporter lines using real-time imaging techniques allows visualisation of the activity and dynamics of signalling pathways over time, in the context of a developing embryo.

The marine annelid *Platynereis* is an ideal model for exploring the evolution of cell types, through large-scale expression profiling at cellular resolution and dissection of gene regulatory networks, and has already allowed elucidation of the evolutionary origin of the vertebrate hypothalamus. Research in the unit also aims to solve one of the remaining big mysteries in animal evolution: the evolution of the central nervous system (CNS).

Several groups in the unit seek to understand both normal development and its deviations in disease. During brain development, vast numbers of neurons are targeted for death and are cleared rapidly and efficiently by a resident lineage of phagocytes, the microglia. Most CNS pathologies are accompanied by activation of the phagocytic microglia, highlighting the importance of understanding the mechanisms underlying the function of these cells, both in healthy and diseased brains. Using advanced *in vivo* imaging combined with genetic approaches, the dynamic relationship between neurons and microglia in zebrafish is actively investigated.

Re-shuffling of regulatory inputs after chromosomal rearrangements is the likely cause of several human genetic disorders and may also link large structural variations widespread in humans to modulation of the quantitative, tissue-specific and temporal expression patterns of neighbouring genes. With a focus on the regulatory architecture of several developmental loci, understanding the molecular mechanisms that control functional interactions between genes and remote cis-regulatory elements and determining how they contribute to phenotypic variations during vertebrate evolution and in humans is an aim of research in the unit.

The unit's research has also led to development of mouse models for endocrine cancer, premature ovarian failure, polycystic kidney disease and obesity. The combination of genetics, expression profiling and proteomics is providing important insight into the molecular basis of these diseases and of their normal developmental counterparts.

Anne Ephrussi
Coordinator, Developmental Biology Unit

Cell polarity and RNA localisation

Anne Ephrussi

PhD 1985, Massachusetts Institute of Technology.

Postdoctoral research at Harvard University and Whitehead Institute, MIT, Cambridge, Massachusetts.

Group leader at EMBL since 1992. Coordinator of EICAT since 2005; Unit Coordinator since 2007.

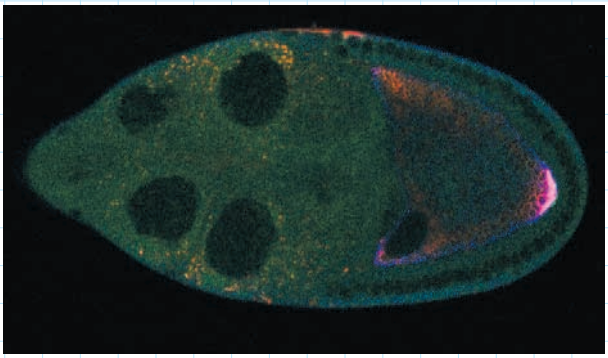
Previous and current research

Polarity is a main feature of eukaryotic cells, underlying many basic cellular functions and developmental processes. Cell polarisation involves the targeting of cytoskeletal structures, organelles, and molecules, including RNAs to specific subcellular locations. RNA localisation coupled with localised translational control is now recognised as a powerful, conserved and highly prevalent mechanism controlling the functional polarisation of cells, yet the mechanisms regulating these processes are still poorly understood.

In *Drosophila*, asymmetrically localised cell fate determinants in the oocyte specify the body axes and patterning of the future embryo. The key determinants, *bicoid*, *gurken* and *oskar*, are localised as mRNAs and locally translated, ensuring the spatial restriction of their activities. Proper cytoskeletal organisation and specific motor proteins are required for mRNA targeting. Using these RNAs as models, our research is concerned with understanding how RNA localisation and translational control are regulated in space and time.

Of particular interest is *oskar*, which regulates abdomen formation and induces germline formation in the fly. Ectopic *oskar* activity causes severe developmental defects, hence its tight spatial restriction is critical. This is achieved by RNA localisation-dependent translation: *oskar* translation is repressed during transport and activated when the mRNA reaches the posterior pole. Multiple mechanisms then cooperate to achieve tight anchoring of the mRNA and protein at the posterior.

The *Drosophila* oocyte is ideally suited for genetic, biochemical 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) cytoskeletal polarisation, (2) the assembly of the RNA transport complexes and their association with motors and the cytoskeleton mediating their movement, (3) spatial control of translation within cells.



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

Future projects and goals

Combining genetics, proteomics, biochemistry, and a broad spectrum of cell biological approaches, from electron microscopy to live cell imaging, we are investigating:

- the mechanisms underlying cell polarisation;
- the role of the cytoskeleton and motors in mRNA transport;
- the architecture of transport RNPs: the cis-acting RNA elements and interacting proteins, and how they assemble to form functional RNA transport complexes;
- the mechanisms coupling mRNA localisation and translational control;
- how Oskar protein nucleates formation of the polar granules, the germline granules of *Drosophila*.

Our goal is to understand the basic mechanisms underlying RNA transport and spatial control of translation, and how they cooperate to generate a correctly patterned embryo.

Selected references

Besse, F., Lopez de Quinto, S., Marchand, V., Trucco, A. & Ephrussi, A. (2009). Drosophila PTB promotes formation of high-order RNP particles and represses *oskar* translation. *Genes Dev*., 23, 195-207

Besse, F. & Ephrussi, A. (2008). Translational control of localized mRNAs: restricting protein synthesis in space and time. *Nat. Rev. Mol. Cell Biol., 9*, 971-80

Vanzo, N., Oprins, A., Xanthakis, D., Ephrussi, A. & Rabouille, C. (2007). Stimulation of endocytosis and actin dynamics by Oskar polarizes the *Drosophila* oocyte. *Dev. Cell,* 12, 543-555.

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.

Hachet, O. & Ephrussi, A. (2004). Splicing of *oskar* RNA in the nucleus is coupled to its cytoplasmic localization. *Nature,* 28, 959-963

Detlev Arendt

PhD 1999, Albert-Ludwigs-Universität, Freiburg.

Postdoctoral research at EMBL.

Team leader at EMBL since 2002. Group leader and Senior Scientist since 2007.

Academic Mentor, post-doctoral training since 2007.

Evolution of the central nervous system in Bilateria

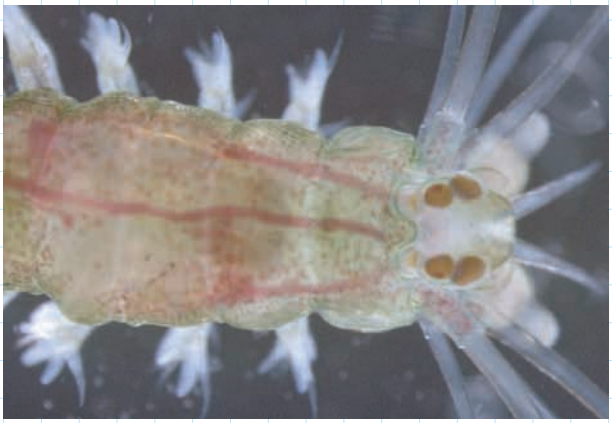
Previous and current research

We are intrigued by one of the remaining great mysteries in animal evolution: how did our central nervous system (CNS) 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, which 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 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.



Platynereis dumerilii *(Polychaeta, Annelida, Lophotrochozoa).*

Finally, we have also established neurobiological assay systems for larval swimming and for adult learning, combined with computer modelling of these and of other complex behavioural traits, in order to investigate the functions of conserved cell type and to gain insight into the neurobiology of marine planktonic life.

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.

Selected references

Jekely, G., Colombelli, J., Hausen, H., Guy, K., Stelzer, E., Nédélec, F. & Arendt, D. (2008). Mechanism of phototaxis in marine zooplankton. *Nature*, 456, 395-399

Arendt, D. (2008). The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* 9, 868-82

Denes, A.S., Jekely, G., Steinmetz, P.R., Raible, F., Snyman, H., Prud'homme, B., Ferrier, D.E., Balavoine, G. & Arendt, D. (2007). Molecular architecture of annelid nerve cord supports common origin of nervous system centralization in bilateria. *Cell*, 129, 277-288

Tessmar-Raible, K., Raible, F., Christodoulou, F., Guy, K., Rembold, M., Hausen, H. & Arendt, D. (2007). Conserved sensory-neurosecretory cell types in annelid and fish forebrain: insights into hypothalamus evolution. *Cell*, 129, 1389-1400

Timing of mammalian embryogenesis

Alexander Aulehla

MD 2002, Albert-Ludwigs-University, Freiburg, Germany

Research at the MD Anderson Cancer Center, Houston, USA and the Max-Planck Institute, Freiburg.

PhD 2008, Paris VI University.

Postdoctoral research at the Stowers Institute, Kansas City, USA, 2005-2009.

Group leader at EMBL since 2009.

Previous and current research

During an embryo's journey from a single cell to a complex organism, countless patterning processes unfold with remarkable precision, both spatially but also in respect to their temporal sequence, or timing. It is this temporal aspect of embryonic development that constitutes the focus of our research. How is developmental time measured in the embryo? How is the timing of patterning processes controlled and, importantly, globally synchronised? And finally, how does the timing itself influence the phenotype? We aim to approach these questions by studying the mechanisms controlling overall developmental rate, as well as by studying the timing of individual processes, including the dynamics of underlying signalling pathways.

One process that suits this approach particularly well is the formation of somites, the precursors of vertebrae. Somites, epithelial spheres that are generated from the paraxial mesoderm, form periodically in a head-to-tail sequence. The periodicity of this process (approximately 2 hours in mouse embryos) is thought to be linked to a molecular oscillator, termed the segmentation clock. In mouse embryos, it involves the oscillatory activity of several signaling pathways (Wnt, Notch and Fgf signalling) in the forming mesoderm. How these oscillations are generated in the first place and what ultimately controls and tunes the periodicity of these oscillations is unknown.

In order to analyse the mechanisms underlying oscillatory pathway activity, a prerequisite has to be fulfilled: the phenomenon of oscillations has to be made visible and thus amenable to quantification in the living embryo. We previously provided the proof of principle that fluorescence-based real-time imaging of segmentation clock activity in mouse embryos is feasible. We are now developing this approach further and are establishing a novel, versatile real-time reporter system that will allow us to visualise the dynamics of Wnt-signaling activity. This signaling pathway serves a multitude of evolutionary conserved functions during development and has been shown to play an essential role during somite formation. The real-time reporter system is designed to reflect Wnt- signaling activity both on transcriptional as well as translational level, directly in the context of developing mouse embryos. This will enable us to determine how the striking oscillations of Wnt-signalling activity are generated in the first place and moreover, to functionally test their role in embryonic patterning. We are particularly interested in identifying the factors, both intrinsic as well as extrinsic, that are responsible for setting the tempo of oscillations in the segmentation process. The insight gained from studying this specific oscillation phenomenon will be combined with our efforts that address the mechanisms that control the timing of overall development.

In situ *hybridisation of mouse embryo at day 9 of development. Uncx4.1 mRNA is visualized in formed somites, while Wnt3a mRNA is expressed in the posterior embryo.*

Future projects and goals

Using a combination of classical experimental embryology, mouse genetics, ES-cell technology and our expertise in real-time imaging of mouse embryos, our future goals are:

- Generation of a real-time imaging reporter system for Wnt-signaling oscillations in mouse embryos using embryonic stem cell technology;
- Discovering the mechanisms underlying Wnt-signaling oscillations during embryogenesis;
- Functional studies addressing the role of Wnt-signaling oscillations and the timing of somitogenesis;
- Identification of global mechanisms controlling the rate of development in mouse embryos.

Selected references

Aulehla, A., Wiegraebe, W., Baubet, V., Wahl, M.B., Deng, C., Taketo, M., Lewandoski, M., & Pourquie, O. (2008). A beta-catenin gradient links the clock and wavefront systems in mouse embryo segmentation. *Nat. Cell Biol.,* 10, 186-193

Aulehla, A., Wehrle, C., Brand-Saberi, B., Kemler, R., Gossler, A., Kanzler, B., and Herrmann, B.G. (2003). Wnt3a plays a major role in the segmentation clock controlling somitogenesis. *Dev. Cell,* 4, 395-406

Aulehla, A. & Johnson, R.L. (1999). Dynamic expression of lunatic fringe suggests a link between notch signaling and an autonomous cellular oscillator driving somite segmentation. *Dev. Biol.,* 207, 49-61

Stefano de Renzis

MD 1997, University Federico II, Naples.

PhD 2002, EMBL Heidelberg.

Postdoctoral work at Princeton University.

Group leader at EMBL since 2008.

Transcriptional control of protein and membrane trafficking during tissue morphogenesis

Previous and current research

During animal development the expression of many differentiation factors is concomitant with – and required for – any series of morphological changes, giving rise to final tissue and organ shape. Our goal is to understand how cells reorganise their content in response to the expression of key developmental transcription factors. We focus on how tissue-specific gene expression drives specific patterns of protein and membrane trafficking and how this regulates cell fate and behaviour. Tackling this problem requires an experimental system where changes in gene expression and intracellular trafficking are directly linked.

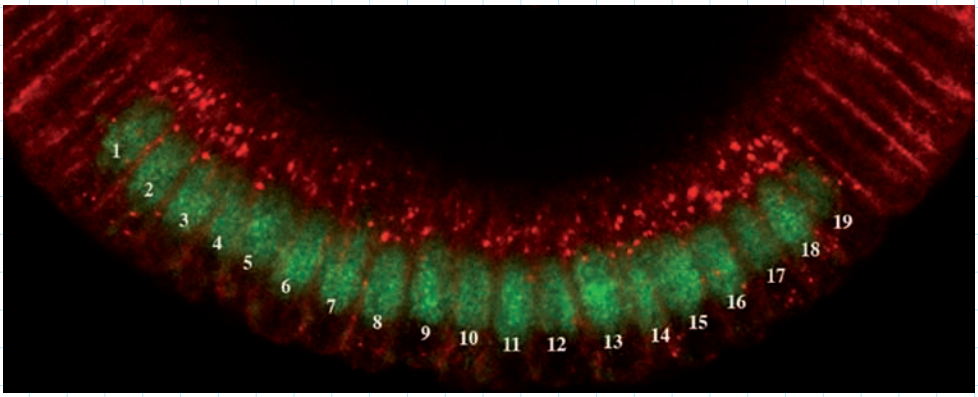
The early *Drosophila* embryo provides an excellent system. In about 60 minutes a syncytium of ~6000 nuclei completes the process of cellularisation, a particular form of cytokinesis involving a massive mobilisation of intracellular membranes. Concomitantly, the embryo undergoes extensive remodelling of gene expression characterised by the activation of the zygotic genome and degradation of previously supplied maternal transcripts (maternal to zygotic transition). This transition immediately precedes gastrulation when tissue differentiation becomes increasingly dramatic. Because zygotic transcription is required for cellularisation, it can directly influence the differentiation of the plasma membrane by differentially regulating the distribution of proteins and lipids in different cell types.

We have developed a system based on chromosomal rearrangements and microarrays that has allowed, for the first time, the identification of the entire set of zygotic genes active at cellularisation. This dataset represents a valuable resource for indicating specific gene functions and consequently the mechanisms of specific morphogenetic processes. We have applied this approach to identify the genes controlling the mesoderm specific activation of Notch trafficking (see figure). Importantly, mesoderm specific trafficking patterns are not limited to Notch and Delta. Many regulatory proteins involved in mesoderm invagination show also similar trafficking patterns. It is likely that the mesoderm specific modulation of intracellular membrane dynamics represents a general regulatory principle operating during mesoderm morphogenesis.

Future projects and goals

Using a combination of genetics and microarray approaches we will identify the cell biological basis underlying the pathways controlling changes in protein targeting in the mesoderm. The activity of specific regulators of the membrane transport machinery will be directly probed using video-microscopy as well as biochemical assays. In addition, we will characterise the activity of cis-regulatory sequences and transacting factors involved in timing the activation of the zygotic genome and degradation of maternal transcripts. We are particularly interested in characterising how pathways and genes identified above achieve early pattern expression. Our long-term goal is to analyse the differentiation of intracellular pathways also in other cell types and tissues. We wish to understand how machineries controlling intracellular trafficking are tuned during cell differentiation and how this differential tuning controls tissue morphogenesis.

Notch vesicular trafficking (red) corresponds precisely to cells expressing the transcription factor Snail (mesoderm).



Selected references

De Renzis, S. *et al.* (2007). Unmasking activation of the zygotic genome using chromosomal deletions in the *Drosophila* embryo. *PLoS Biology*, 5, 5

De Renzis, S. *et al.* (2006). Dorsal-ventral pattern of Delta trafficking is established by a Snail-Tom-Neuralized pathway. *Dev. Cell*, 2, 257-64

De Renzis, S. *et al.* (2002). Divalent Rab effectors regulate the sub-compartmental organization and sorting of early endosomes. *Nat. Cell Biol.*, 4, 124-33

Developmental patterning in plants

Marcus Heisler

PhD 2000, Monash University, Australia.

Postdoctoral research at the California Institute of Technology 2001-2007.

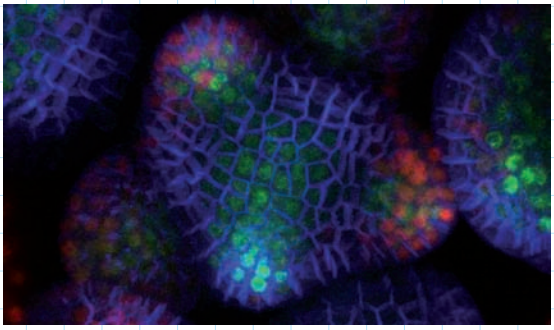
Senior Research Associate at the California Institute of Technology 2007-2009.

Group leader at EMBL since 2009.

Previous and current research

In addition to providing us with the air we breathe, the food we eat and much of the energy we use, plants exhibit a unique beauty associated with their strikingly symmetrical patterns of development. Multicellularity also evolved independently in plants giving us an opportunity to compare and contrast the developmental strategies used in different kingdoms. We are investigating plant development by focussing on the process of lateral organ formation (leaves or flowers) in the model species *Arabidopsis thaliana*. We are taking a broad approach that includes trying to understand organ positioning, differentiation and growth and how these different processes are coordinated. Experimentally, we have developed confocal-based methods to image growing plant tissues, enabling us to obtain dynamic high-resolution data (making full use of the different GFP spectral variants), which we can also incorporate directly into mathematical developmental models.

Recent work has revealed that primordial positions in plants are specified by local, high concentrations of an intercellular signalling molecule, auxin (indole-3-acetic acid). In turn, the formation of these auxin concentration maxima depends on a polar auxin transport system that directs auxin flux towards sites of primordial emergence. Polar auxin transport at the plant shoot apex plant is mediated by the auxin efflux carrier PIN1-FORMED1 (PIN1). This is member of a small family of membrane-bound auxin efflux proteins that are localised in a polar fashion to different sides of cells so that auxin efflux occurs in a directional manner. When primordia are specified, PIN1 localisation in meristem epidermal cells is on the sides of the cells facing towards the initiation site. A major goal of our research in the near and long-term future is to understand the mechanisms and signals responsible for coordinating and directing this polar localisation pattern underlying primordium positioning. Auxin not only induces primordium growth but also helps to regulate genes that help specify different organ cell types. Understanding how auxin induces different sets of genes in different domains of growing primordia is another focus for our lab. In particular we are interested in how the 'top' and 'bottom' or adaxial and abaxial cell types of primordia are specified and the downstream role of these cell types in controlling organ shape. Lastly, we are also working towards understanding the mechanical basis of morphogenesis by developing methods to mechanically perturb and track cells, quantify growth patterns and correlate these data with gene activities and the plant cytoskeleton to better understand how genes and mechanics relate to one another.

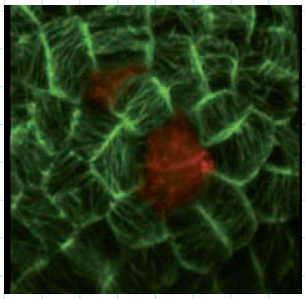


GFP labelling of two nuclear localised transcription factors (green and red) and a membrane protein (blue) in developing flower primordia at the Arabidopsis *shoot apex.*

Future projects and goals

There are many interesting questions we are pursuing, using any technique that seems appropriate, including:

- Understanding the patterning processes that specify adaxial and abaxial cell types;
- Understanding the mechanism by which adaxial and abaxial cell types regulate organ morphogenesis;
- Understanding the basis of supra-cellular patterning of the plant cytoskeleton and how it is coordinated with auxin transport;
- Understanding how polar auxin transport is patterned and its role in development.



Microtubules (green) form concentric alignments surrounding laser ablated plant cells (red).

Selected references

Gordon, S.P., Heisler, M.G., Reddy, G.V., Ohno, C., Das, P. & Meyerowitz, E.M. (2007). Pattern formation during *de novo* assembly of the *Arabidopsis* shoot meristem. *Development*, 134, 3539-48

Michniewicz, M., Zago, M.K., Abas, L., Weijers, D., Schweighofer, A., Meskiene, I., Heisler, M.G., Ohno, C., Zhang, J., Huang, F., Schwab, R., Weigel, D., Meyerowitz, E.M., Luschnig, C., Offringa, R. & Friml, J. (2007). Antagonistic regulation of PIN phosphorylation by PP2A and PINOID directs auxin flux. *Cell*, 130, 1044-56

Jonsson*, H., Heisler*, M.G., Shapiro, B.E., Meyerowitz, E.M. & Mjolsness, E. (2006). An auxin-driven polarized transport model for phyllotaxis. *Proc. Natl. Acad. Sci. USA*, 103, 1633-8

Heisler, M.G., Ohno, C., Das, P., Sieber, P., Reddy, G.V., Long, J.A. & Meyerowitz, E.M. (2005). Patterns of auxin transport and gene expression during primordium development revealed by live imaging of the *Arabidopsis* inflorescence meristem. *Curr. Biol.*, 15, 1899-911

Francesca Peri

PhD 2002, University of Cologne.

Postdoctoral research at the Max Planck Institute for Developmental Biology, Tübingen.

Group leader at EMBL since 2008.

Microglia: the guardians of the developing brain

Previous and current research

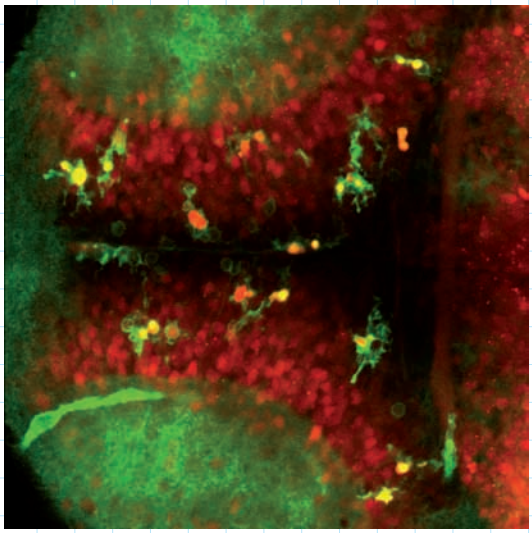
During brain development neurons are generated in great excess and only those that make functional connections survive, while the majority is eliminated via apoptosis. Such huge numbers of dying cells pose a problem to the embryo, as leaking cell contents damage the surrounding environment. Therefore, the clearance of dying cells must be fast and efficient and is performed by a resident lineage of 'professional' phagocytes, the microglia. These cells patrol the entire vertebrate brain and sense the presence of apoptotic and damaged neurons. The coupling between the death of neurons and their phagocytosis by microglia is striking; every time we observe dead neurons we find them already inside the microglia. This remarkable correlation suggests a fast-acting communication between the two cell types, such that microglia are forewarned of the coming problem. It is even possible that microglia promote the controlled death of neurons during brain development. Despite the importance of microglia in several neuronal pathologies, the mechanism underlying their degradation of neurons remains elusive.

The zebrafish *Danio rerio* is an ideal model system to study complex cell-cell interactions *in vivo*. As the embryo is optically transparent, the role of molecular regulators identified in large-scale forward and reverse genetic screens can be studied *in vivo*. Moreover, a key advantage of the system is that zebrafish microglia are extremely large, dynamic cells that form a non-overlapping network within the small transparent fish brain. Labelling microglia, neurons and organelles of the microglial phagocytotic pathway simultaneously in the living zebrafish embryos allows us to image for the first time the entire microglial population to study the interaction between neurons and microglia.

Future projects and goals

We plan to exploit the system at the cell biological level to understand how the microglia cells find, engulf and digest dying and sick neurons. We are generating methods for real time identification of apoptotic neurons and controlled-killing of neurons under the microscope to monitor the response of the surrounding microglia *in vivo*. We predict that dynamic changes in cell branching behaviour will reveal functionally significant interactions between microglia and the neighbouring cells. Moreover, we are currently using a number of forward and reverse genetic approaches to identify the molecules that control the interactions between these two cell types. A recent genetic screen has led to the isolation of several mutant lines where the response of microglia to neuronal apoptosis is affected, and we are cloning the underlying genes.

The aim of our work is to advance our understanding of microglial-mediated neuronal degeneration, a hallmark of many neuronal diseases as common as Alzheimer's and Parkinson's disease.



Microglia (green) and neurons (red) in the zebrafish embryonic brain.

Selected references

Peri, F. & Nusslein-Volhard, C. (2008). Live imaging of neuronal degradation by microglia reveals a role for v0-ATPase a1 in phagosomal fusion *in vivo*. *Cell*, 133, 916-927

Gene regulation and genome architecture in development and evolution

François Spitz

PhD 1997, Institut Cochin de Génétique Moléculaire, Paris.

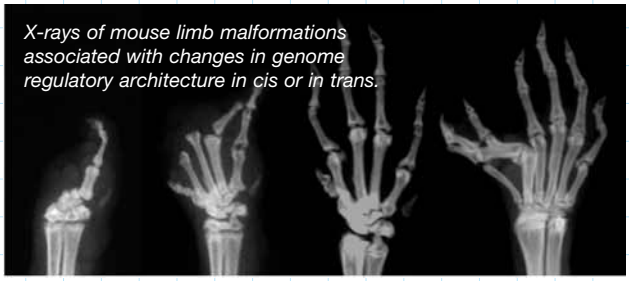
Postdoctoral research at the University of Geneva.

Group leader at EMBL since 2006.

Previous and current research

Embryonic development is organised by a large number of genes, for which expression levels and specificities have to be tightly regulated at the transcriptional level. In vertebrates, this is achieved with genes being regulated by batteries of cis-elements, which act in modular, synergetic and complementary manners. These regulatory elements are often found spread over large chromosomal domains that contain multiple genes with distinct functions and expression profiles, an intricate situation most probably due to regulatory tinkering during evolution. Several functional and computational approaches have been successful in predicting or identifying cis-regulatory elements for a number of genes. However, and in contrast, the molecular mechanisms that translate the intermingled arrays of gene and cis-regulatory elements – which compose vertebrate chromosomes into coherent gene-specific expression programmes – are poorly understood. The dramatic consequences of several chromosomal rearrangements found in humans illustrate that, besides the mere presence of cis-regulatory elements, the specific structure of a locus may contribute substantially to gene expression, underscoring the existence and importance of the genome regulatory architecture. More generally, the changes in this architecture associated with the widespread structural variations recently described in humans may also modulate the expression profile of the surrounding genes, leading to increased phenotypic diversity. However, in both cases, the molecular and regulatory mechanisms associated with this dimension of the genome are mostly unknown.

Our work aims to explore the nature of the vertebrate regulatory genome and identify the underlying mechanisms. As model systems, we focus on few megabase-large genomic loci, which organisation is extensively conserved within vertebrates, or where rearrangements have been associated with developmental genetic abnormalities in humans. We are using mouse transgenesis to identify the regulatory elements that control the specific activities of the multiple genes present in these loci, and state-of-the-art chromosomal engineering techniques to create series of mouse lines with deletions, duplications and inversions to reshuffle these loci in various ways, some replicating known human genetic disorders. By combining these approaches, we are starting to unravel the intricate regulatory organisation of these regions, as well as to get insights into the mechanisms that control and restrict the functional interactions between cis-regulatory elements and neighbouring genes.



X-rays of mouse limb malformations associated with changes in genome regulatory architecture in cis or in trans.

Future projects and goals

We aim to develop our understanding of functional and regulatory organisation of the mammalian genome, using genetic and chromosomal re-engineering approaches in mice as well as functional and computational genomics. In particular, if it is known that changes in chromatin structure, conformation and localisation within the nucleus are associated with different transcriptional activities, the hierarchy of these events and how they are determined by the genomic sequences are still poorly understood. The mouse lines we have generated, with differentially rearranged chromosomal domains, allow us to investigate the relationships between genome sequence, chromatin structure and gene expression by combining phenotypic and gene expression assays with chromatin profiling, conformation analysis and imaging.

We have recently developed efficient transposition and recombination strategies, enabling rapid generation of transgene insertions and chromosomal rearrangements *in vivo*. We aim to use these tools to further chart the regulatory architecture of the mammalian genome, and also to investigate the molecular and phenotypic consequences of structural variations found in humans.

We are also interested in comparing regulatory architecture between different species, to trace back its emergence during evolution and the associated evolutionary regulatory tinkering. We are particularly interested by the role of ancestral mobile elements, which appear to have been exapted into yet-to-be-determined functions (project funded by HFSP, in collaboration with Gill Bejerano, Stanford).

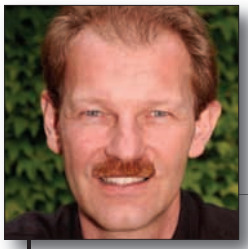
Selected references

Friedli, M., Nikolaev, S., Lyle, R., Arcangeli, M., Duboule, D., Spitz, F. & Antonarakis, S.E. (2008). Characterization of mouse Dactylaplasia mutations: a model for human ectrodactyly SHFM3. *Mamm. Genome*, 19, 272-278

Spitz, F. & Duboule, D. (2008). Global control regions and regulatory landscapes in vertebrate development and evolution. *Adv. Genet.*, 61, 175-205

Gonzalez, F., Duboule, D. & Spitz, F. (2007). Transgenic analysis of Hoxd gene regulation during digit development. *Dev. Biol.*, 306, 847-859

Spitz, F., Herkenne, C., Morris, M.A. & Duboule, D. (2005). Inversion-induced disruption of the Hoxd cluster leads to the partition of regulatory landscapes. *Nat. Genet.*, 37, 889-893



Mathias Treier

PhD 1994, University of Heidelberg.

Postdoctoral research at the University of California, San Diego.

Group leader at EMBL since 2000. Joint appointment with EMBL Monterotondo.

Mammalian organogenesis and physiology

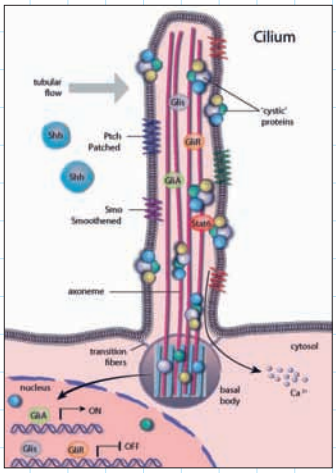
Previous and current research

The specification of cell types during organ development has been studied intensively over the last decade. The future challenge is to understand how these different cell types function in a concerted action within an organ to fulfill its physiological task, and ultimately how mammalian physiology is orchestrated to allow an organism to survive.

We employ mouse genetics to study various aspects of mammalian physiology, from the single cell stage to the complex interplay between organs that allow an organism to maintain energy homeostasis.

Stem/progenitor cell populations constitute the basic building units from which organs and whole organisms are created. We have identified with the transcriptional regulator, Sall4, one of the key players that is required to maintain the pluripotency state of embryonic stem cells. Sall4 is highly expressed in the inner cell mass (ICM) of a blastocyst which will give rise to the embryo proper and the primitive endoderm. We could demonstrate that Sall4 is essential for self-renewal of embryonic stem cells as well as progenitor cells of the primitive endoderm derived from the ICM. We are currently employing genetic and biochemical methods to understand the regulation and function of this important player in stem cell biology in greater detail.

At the organ level our research is mainly focussed on the kidney. We are interested in both the development of the organ as well as the physiological functions the kidney has to perform to maintain homeostasis at the organismal level. Many life-threatening inherited genetic disorders manifest themselves as malfunctions of the kidney, a particular example being polycystic kidney disease (PKD), which affects an estimated 13 million people worldwide regardless of sex, age, race or ethnic origins (www.pkdcure.org). Whereas PKD leads to an enlargement of the organ through uncontrolled growth, another debilitating kidney syndrome is nephronophthisis, in which the kidneys shrink. Both syndromes are believed to result from aberrant signalling of an ancient organelle present on most kidney cells called the cilium, the sensor that allows kidney cells to react to changes in physiological parameters within the blood and urine (see figure). With the Glis family of transcriptional regulators, we have identified molecular players that are involved in transmitting the signal from the cilium to the nucleus, allowing kidney cells to respond to changes in their environment. We are now investigating at the molecular level how signal transmission is regulated through post-translational modifications. Understanding cilia signalling in general will have implications for many other human diseases, like for example Bardet-Biedel syndrome, that are caused by malfunctioning of this organelle.



Schematic drawing of a cilium transmitting an extra-cellular signal to the nucleus.

The ultimate challenge in systems biology is to understand mammalian physiology. For any living organism, maintaining energy homeostasis is the central task for survival. We are particular interested in the neuronal circuits in the central nervous system (CNS) that are regulating energy balance. We have identified the brain-specific homeobox protein Bsx as an essential player in the regulation of food intake and locomotor activity, the two main components that determine energy homeostasis. We are currently investigating how higher brain centres interact with peripheral signals that signal satiety and hunger to regulate our drive to eat. In light of the obesity pandemic resulting in metabolic syndrome with its complications like type 2 diabetes, the understanding of the molecular mechanisms maintaining energy homeostasis has gained one of the highest priorities.

Future projects and goals

Transcriptional regulators will continue to be central to our investigation. Many can directly sense environmental cues and as a consequence alter the transcriptional readout from our individual genetic blueprint. With a series of mouse models for human diseases that we have created over the years, we are now in a position to dissect even complicated physiological questions at the organismal level. In parallel, we have started to look at how metabolism influences degenerative processes to open new avenues for pharmacological treatments in regenerative medicine.

Selected references

Uhlenhaut, N.H. & Treier, M. (2008). Transcriptional regulators in kidney disease: gatekeepers of renal homeostasis. *Trends Genet.*, 24, 361-371

Attanasio, M., Uhlenhaut, N.H., Sousa, V.H., O'Toole, J.F., Otto, E., Anlag, K., Klugmann, C. *et al*. (2007). Loss of GLIS2 causes nephronophthisis in humans and mice by increased apoptosis and fibrosis. *Nat. Genet.*, 39, 1018-1024

Sakkou, M., Wiedmer, P., Anlag, K., Hamm, A., Seuntjens, E., Ettwiller, L., Tschop, M.H. & Treier, M. (2007). A role for brain-specific homeobox factor bsx in the control of hyperphagia and locomotory behavior. *Cell Metab.*, 5, 450-463

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

Gene Expression Unit

The genome encodes the genetic blueprint that coordinates all cellular processes, which ultimately give rise to phenotype. The expression of genetic information is tightly regulated in both time and space at multiple steps, including at the transcriptional, post-transcriptional and post-translational levels. The Gene Expression Unit takes a systems biology approach to unravel these complex processes at all scales, integrating wet-lab and computational approaches.

In eukaryotes, many steps of gene expression, such as transcription and RNA processing, take place in the structurally complex environment of the nucleus and often involve remodelling of chromatin into active and inactive states. Messenger RNAs, once exported from the nucleus, undergo additional steps of regulation. Their translation results in the production of proteins, whose functions define the characteristics of different cell types, or cellular phenotypes. However, not all RNAs are translated. In recent years multiple types of non-coding RNAs have been discovered with diverse functionality. Genetic variation, affecting the function of genes at any level, results in abnormal phenotypes, often leading to diseases such as cancer. Groups within the unit are tackling various aspects of gene expression, often developing new enabling technologies.

A particular strength of the unit is its ability to address questions at different scales ranging from detailed mechanistic studies (using biochemistry and genetics), to genome-wide studies (using functional genomic, proteomic and computational approaches) to understand the processes leading from genotype to phenotype. Global, dynamic and quantitative measurements of biological molecules at all levels (DNA, RNA, proteins, cells, organisms, etc) as well as the integration of hypothesis and discovery-driven research characterise the unit. The synergy between computational and wet-lab groups provides a very interactive and collaborative environment to yield unprecedented insights into how genetic information is 'read' and mediates phenotype through molecular networks.

Eileen Furlong and Lars Steinmetz
Joint Coordinators, Gene Expression Unit

Eileen Furlong

PhD 1996, University College Dublin.

Postdoctoral research at Stanford University.

Group leader at EMBL since September 2002.

Senior scientist and joint Unit Coordinator of Gene Expression Unit since 2009.

Gene regulatory networks: Dissecting the logic

Previous and current research

Development is driven by the establishment of complex patterns of gene expression at precise times and spatial locations. Although a number of mechanisms fine-tune gene expression, it is initially established through the integration of signalling and transcriptional networks converging on enhancer elements, or cis-regulatory modules (CRMs). CRMs serve as regulatory circuits that integrate diverse transcriptional inputs leading to a specific spatio-temporal output of expression. Understanding how CRMs function is therefore central to understanding metazoan development and evolutionary change. Although there has been extensive progress in deciphering the function of individual CRMs, how these modules are integrated to regulate more global cis-regulatory networks remains a key challenge. Even in the extensively studied model organism, the *Drosophila* fruit fly, there are no predictive models for a transcriptional network leading to cell fate specification.

The main aim of our research is to understand how gene regulatory networks control development and how network perturbations lead to specific phenotypes. To address this we integrate functional genomic, genetic and computational approaches to make predictive models of developmental progression. We are particularly interested in the topology and function of developmental networks at a global level.

We use *Drosophila* mesoderm specification into different muscle primordia as a model system. The relative simplicity of the fly mesoderm, in addition to the number of essential and conserved transcription factors already identified, make it an ideal model to understand cell fate decisions at a systems level.

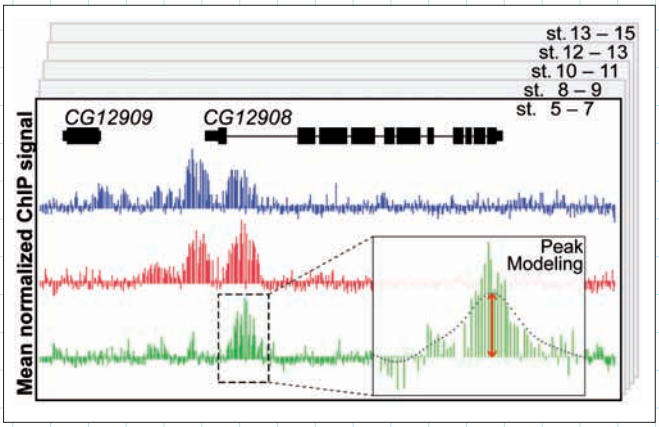


Figure 1: Genome-wide map of enhancer occupancy

Future projects and goals

We have constructed a comprehensive cis-regulatory network during consecutive stages of development, elucidating the combinatorial binding and temporal occupancy of thousand of cis-regulatory modules, which we are currently using to understand differ aspects of gene expression. Our future work will involve extending this network to other species to study the evolution of cis-regulatory networks at a genome-wide scale. We are also generating genetic tools to allow us investigate changes in chromatin remodelling during cell fate decisions. Our ultimate goal is to use this systems-level approach to make predictive models of embryonic development and the effect of genetic perturbations. Working with *Drosophila* allows us to readily test all predictions on network perturbations during embryonic development.

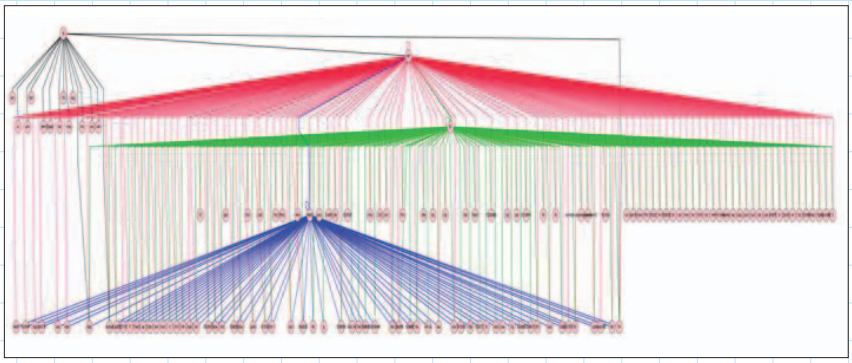


Figure 2: The transcriptional network for early mesoderm development

Selected references

Liu, Y.H., Jakobsen, J.S., Valentin, G., Amarantos, I., Gilmour, D.T. & Furlong, E.E. (2009). A systematic analysis of Tinman function reveals Eya and JAK-STAT signaling as essential regulators of muscle development. *Dev. Cell*, 16, 280-91

Bonn, S. & Furlong, E.E. (2008). cis-Regulatory networks during development: a view of *Drosophila*. *Curr Opin Genet Dev.*, 18, 513-20

Jakobsen, J.S., Braun, M., Astorga, J., Gustafson, E.H., Sandmann, T., Karzynski, M., Carlsson, P. & Furlong, E.E. (2007). Temporal ChIP-on-chip reveals Biniou as a universal regulator of the visceral muscle transcriptional network. *Genes Dev.*, 21, 2448-60

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

Functional genomics of complex traits

Lars Steinmetz

PhD 2001, Stanford University.

Postdoctoral research at the Stanford Genome Technology Center.

Group leader at EMBL since 2003. Academic Mentor, predoctoral training since 2008.

Joint Unit Coordinator since 2009.

Previous and current research

Individuals differ at thousands of DNA sequence positions. These differences interact with each other and with the environment in complex ways to give rise to heritable phenotypic variation. This is the basis of quantitative phenotypes such as body height, cancer, diabetes, crop yield and fungal virulence. Naturally occurring genetic variants also influence the onset and intensity of diseases, as well as their treatment susceptibility, thus providing an incentive for personalised medicine.

The overall aim of our research is to elucidate how genetic variation conditions complex phenotypes. To this end, we integrate experimental and computational biology approaches at multiple layers along the molecular processes linking genotype to phenotype. In particular we investigate the level of the genome, transcriptome and proteome.

Some selected projects include genotyping single-nucleotide polymorphisms across entire yeast genomes to infer meiotic recombination-activity distributions that define trait inheritance; studying the function of pervasive transcription of non-coding RNAs and the mechanisms of how they are generated; and analysing protein-interaction networks to enable the prediction of candidate disease genes for mitochondrial disorders.

Future projects and goals

We are developing new technologies to determine the phenotypic contribution for all sequence variants between two genomes in a single step. In addition, we are dissecting the genetic basis of sensitivity and resistance to malaria parasites in the mosquito, *Anopheles gambiae*. Ultimately, by integrating genetics, genomics, systems biology and computational modelling with high-throughput sequencing and microarrays, we aim to develop approaches that will enable personalised and preventative medicine across the world.

The group is associated with the Stanford Genome Technology Center at Stanford University.

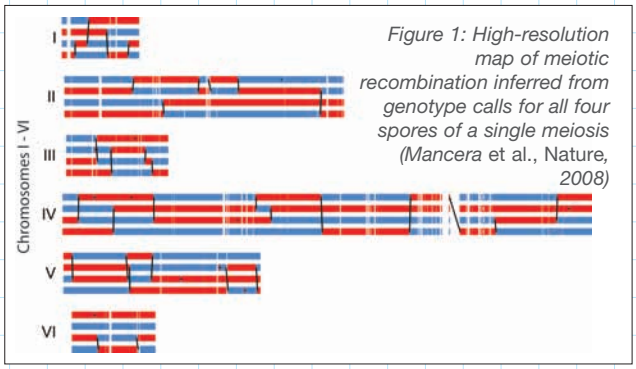


Figure 1: High-resolution map of meiotic recombination inferred from genotype calls for all four spores of a single meiosis (Mancera et al., Nature, *2008)*

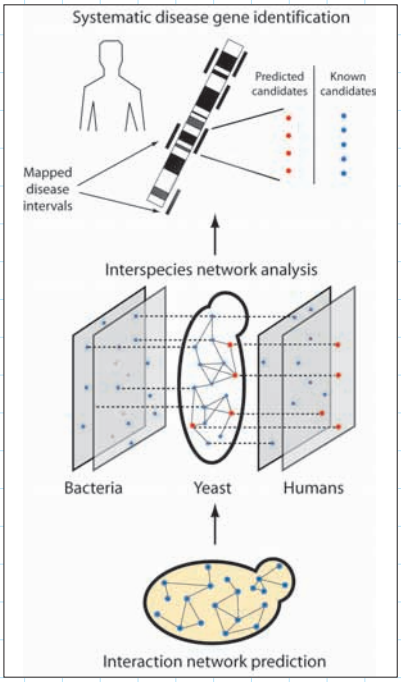


Figure 3 (right): From yeast protein-interaction datasets to the identification of candidate genes for human diseases (Perocchi et al., Mol. Biosyst., *2008)*

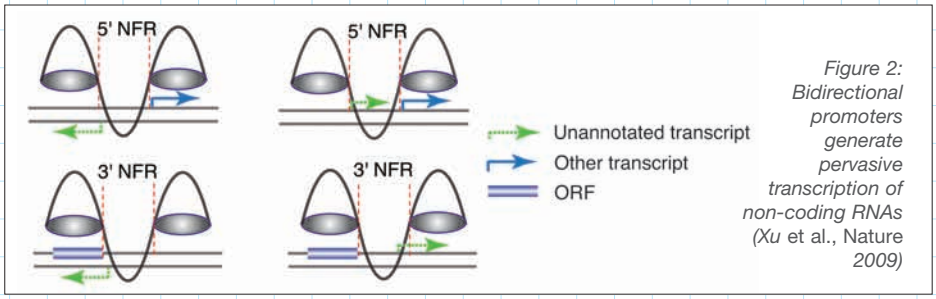


Figure 2: Bidirectional promoters generate pervasive transcription of non-coding RNAs (Xu et al., Nature *2009)*

Selected references

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M. & Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, 457, 1038-42

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W. & Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature*, 457, 1033-7

Mancera, E., Bourgon, R., Brozzi, A., Huber, W. & Steinmetz, L.M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature*, 454, 479-485

Sinha, H., David, L., Pascon, R.C., Clauder-Munster, S., Krishnakumar, S., Nguyen, M. *et al.* (2008). Sequential elimination of major-effect contributors identifies additional quantitative trait loci conditioning high-temperature growth in yeast. *Genetics*, 180, 1661-1670

Asifa Akhtar

PhD 1998, Imperial Cancer Research Fund, London.

Postdoctoral research at EMBL and Adolf-Butenandt-Institut, Munich.

Group leader at EMBL since 2001.

Mechanisms of transcription regulation through chromatin

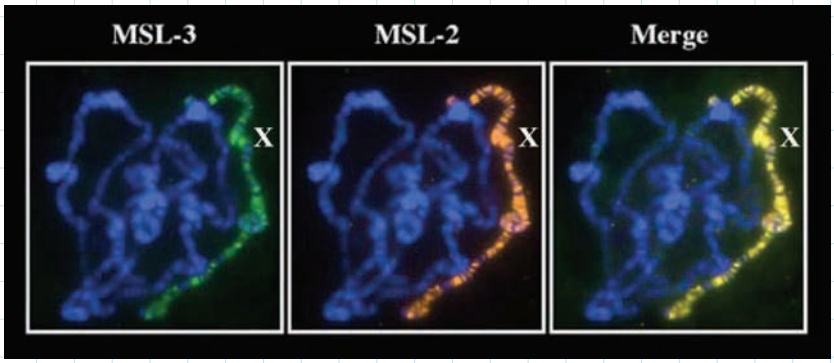
Previous and current research

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.



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.*

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.

Selected references

Kind, J., Vaquerizas, J.M., Gebhardt, P., Gentzel, M., Luscombe, N.M., Bertone, P. & Akhtar, A. (2008). Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell*, 133, 813-828

Kind, J. & Akhtar, A. (2007). Cotranscriptional recruitment of the dosage compensation complex to X-linked target genes. *Genes Dev.*, 21, 2030-2040

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

Computational Genetics

Wolfgang Huber

PhD 1998, University of Freiburg.

Postdoctoral research at IBM in San Jose, California and at the German Cancer Research Centre (DKFZ), Heidelberg.

Group leader at EMBL-EBI since 2004. Joint appointment with the Gene Expression Unit.

Group leader in the Gene Expression Unit since 2009.

Previous and current research

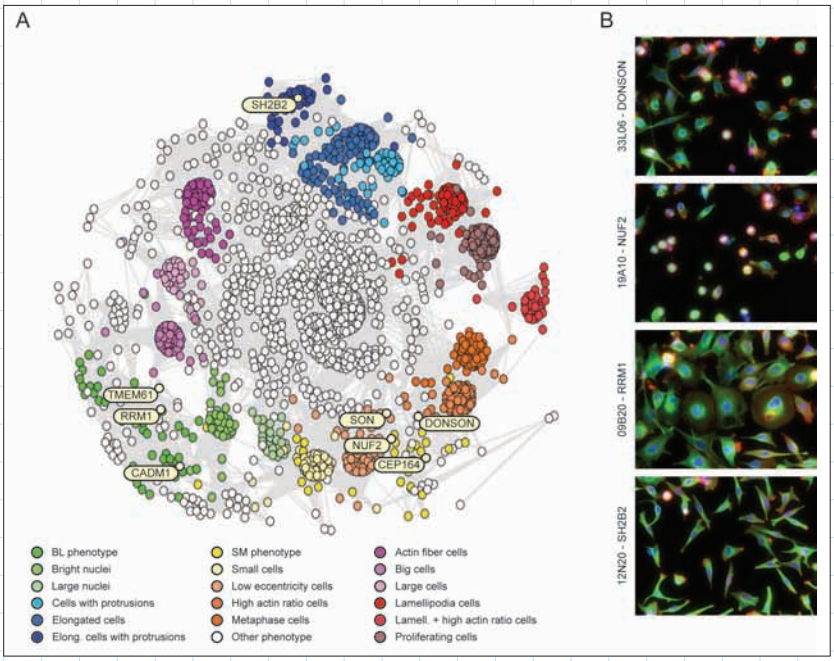
The group studies genotypes and phenotypes on a genome-wide scale: how do variations in the genomes of individuals shape their complex phenotypes? To this end, we develop computational methods in statistics, signal and image processing, and probability models.

We work with experimental labs in systems genetics and functional genomics to design and analyse genome-wide experiments whose aim is to unravel the mechanisms of genetic inheritance, gene expression, molecular interactions, signal transduction and how they shape phenotypes. Most phenotypes, including human diseases, are complex, i.e., they are governed by large sets of genes and regulatory elements. Our aim is to map these complex networks and eventually devise strategies for designing phenotypes by engineering combinatorial perturbations.

Our research is stimulated by new technologies, and we employ data from high-throughput sequencing (ChIP-seq, RNA-seq, genotyping, polymorphism discovery), tiling microarrays, large scale cell based assays, automated microscopy, as well as the most advanced methods of computational statistics. We are a regular contributor to the Bioconductor project (www.bioconductor.org).

Future projects and goals

One of the most exciting questions in biology is the predictive modelling and engineering of phenotypic outcomes based on individual genomes. To get there, we need a better understanding of cellular regulation and physiological processes through advances in experimental technologies for the manipulation and observation of genetic model systems, and in computational biology for understanding the data and model building. Of particular interest to us are systematic genetic assays for phenotypic consequences of DNA sequence and copy number variation and of drug perturbation; as well as high-content phenotyping using automated microscopy. To make these advances fruitful for predictive models of biological systems, we aim to stay at the forefront of developments in data analysis, statistical software and mathematical modelling. An emphasis lies on project-oriented collaborations with experimenters.



Genome-wide phenotypic similarity map. (A) Each of the 1,839 nodes represents an siRNA perturbation in HeLa cells whose shape and morphology was monitored by automated microscopy. Representative images for four siRNA perturbations, for the target genes DONSON, NUF2, RRM1 and SH2B2, are shown in panel B. Nodes in the map are linked by a grey edge when they are phenotypically similar. The graph is a two-dimensional representation of the phenotypic diversity observed after a genome-wide siRNA perturbation screen.

Selected references

Xu, Z. *et al.* (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature,* 457, 1033-7

Hahne, F. *et al.* (2008) Bioconductor Case Studies (Use R series). Springer

Lin, S.M. *et al.* (2008). Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res.,* 36, Article e11

Mancera, E. *et al.* (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature,* 454, 479-485

Tödling, J. & Huber, W. (2008). Analyzing ChIP-chip data using Bioconductor. *PLoS Computational Biology,* 4, e1000227

Maja Köhn

PhD 2005 MPI for Molecular Physiology, Dortmund.

Postdoctoral work at Harvard University, Cambridge, Massachusetts.

Group leader at EMBL since 2007.

Investigation of phosphatases using chemical biology tools

Previous and current research

Protein dephosphorylation by protein phosphatases (PPs) is fundamental to a vast number of cellular signalling processes and thus to physiological functions. Impairment of these processes contributes to the development of human diseases such as cancer and diabetes. The investigation of phosphatases is challenging, mainly due to their broad substrate specificity and the lack of tools to selectively study particular phosphatases. Therefore, knowledge of phosphatase function and substrate interaction is generally still quite limited. Our main interest is thus to control and investigate phosphatases with the help of chemical tools, based on phosphoinositide and peptide synthetic organic chemistry as well as protein semisynthesis, and also with molecular biology approaches. We are currently focussing on protein tyrosine phosphatases (PTPs) and dual specificity phosphatases (DSPs).

We are working on the design of inhibitors for PTPs based on chemical modification of protein/peptide substrates in a way that they a) cannot be dephosphorylated and b) show an increased half-life in the body. Thus, upon binding, the phosphatase cannot fulfil its function and is bound to the modified substrate with natural high affinity. In this way, one does not have to rely on the random discovery of effector molecules by, for example, exhaustive screening of large compound libraries. In addition, we are looking into other possible natural substrates of phosphatases such as phosphoinositides. Working on new synthetic strategies to simplify access to these compounds as well as their analogues is necessary in order to be able to control the function of lipid phosphatases in cells.

In case the substrate specificity of a PTP/DSP is not known, screening of focussed peptide libraries that are designed based on for example structural data of the phosphatase is necessary. When applying the above described strategy, the discovery of artificial or even natural substrates will lead to an inhibitor of the phosphatase of interest. Here, we are particularly interested in the PRL family of phosphatases, which is involved in several types of cancer. In addition to creating effector molecules using synthetic organic chemistry, we also apply protein semisynthesis and molecular biology approaches to obtain information about natural substrates, regulation and networks of these phosphatases.

Future projects and goals

We are interested in further developing chemical methods to stabilise (phospho-)peptides and in working on novel cell penetration concepts. We are planning on modifying phosphatases semisynthetically in order to control their function, not through effector molecules but intrinsically. Developing effector molecules for the highly non-specific PSTPs is a long-term goal. The activity of these phosphatases is controlled not only by specificity, but also by cofactors and cellular localisation, which adds to the challenge of finding tools to selectively target these phosphatases in the context of the cell.

Selected references

Jonkheijm, P., Weinrich, D., Köhn, M., Engelkamp, H., Christianen, P.C., Kuhlmann, J., Maan, J.C., Nusse, D., Schroeder, H., Wacker, R., Breinbauer, R., Niemeyer, C.M. & Waldmann, H. (2008). Photochemical surface patterning by the thiol-ene reaction. *Angew Chem Int Ed Engl.*, 47, 4421-4424

Köhn, M., Gutierrez-Rodriguez, M., Jonkheijm, P., Wetzel, S., Wacker, R., Schroeder, H., Prinz, H., Niemeyer, C.M., Breinbauer, R., Szedlacsek, S.E. & Waldmann, H. (2007). A microarray strategy for mapping the substrate specificity of protein tyrosine phosphatase. *Angew Chem Int Ed Engl.*, 46, 7700-7703

Watzke, A., Köhn, M., Gutierrez-Rodriguez, M., Wacker, R., Schroder, H., Breinbauer, R., Kuhlmann, J., Alexandrov, K., Niemeyer, C.M., Goody, R.S. & Waldmann, H. (2006). Site-selective protein immobilization by Staudinger ligation. *Angew Chem Int Ed Engl.*, 45, 1408-1412

Studying the origin and genomic impact of copy-number variants by genomics and computational biology

Jan Korbel

PhD 2005, EMBL Heidelberg, Humboldt University Berlin.

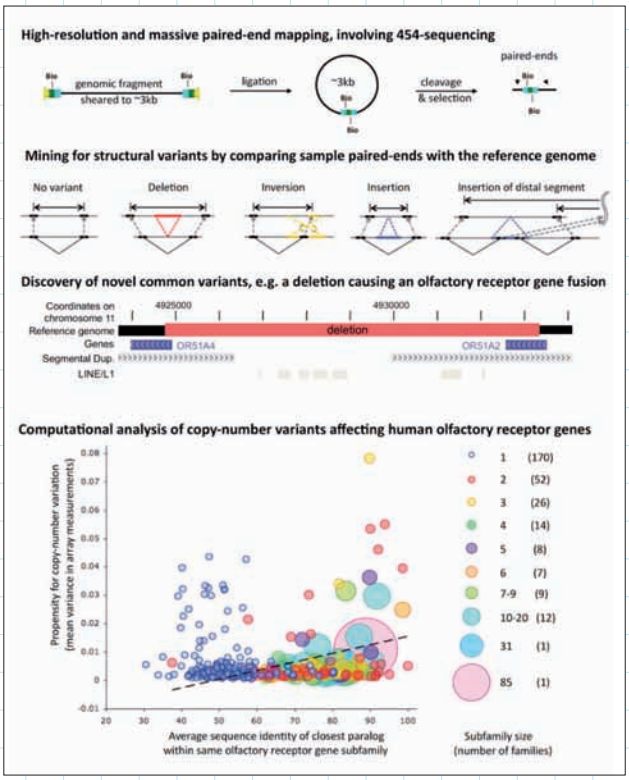
Postdoctoral research at Yale University, New Haven, CT.

Group Leader at EMBL since October 2008. Joint appointment with EMBL-EBI.

We are an interdisciplinary group interested in genome variation, with young researchers from various different scientific disciplines. Specifically, our group aims to combine experimental and computational approaches to study the mutational formation-processes and the functional impact of genomic variation in humans, in particular genomic structural variants (SVs). SVs, frequently referred to as copy-number variants (CNVs), are >1kb deletions, duplications, insertions, and inversions responsible for most genetic variation in the human genome.

Previous and current research

We recently developed high-resolution and massive paired-end mapping (PEM), an approach involving next-generation DNA sequencing of the end-stretches of 3kb genomic fragments and computational mapping of these against a reference genome to identify SVs at near-base pair resolution. We use PEM and other computational and experimental approaches, e.g. tiling microarray-based methods, to determine the extent of SVs in the genome. Furthermore, these approaches help us to deduce the molecular mechanisms causing SV-formation and to study the evolutionary and functional impact of SVs. For example, we recently determined the extent of copy-number variants affecting the olfactory receptor gene family and obtained insights into the evolution and diversity of this large human gene family. By computationally mining the sequenced breakpoint junctions of SVs, striking variations in the gene content of healthy individuals were identified, including novel fusion genes. In addition, our sequence analyses revealed that the molecular mechanisms causing SV-formation mainly include retrotransposition and meiotic as well as DNA repair-associated recombination processes.



Future projects and goals

The extent to which genomes of healthy individuals differ due to SVs, the impact of SVs on the phenotype and the mutational processes underlying SV-formation are presently poorly understood. We hypothesise that SVs commonly lead to phenotypic variation, e.g. by perturbing tightly regulated cellular processes. We will study the formation mechanisms and the functional impact of SVs in the human genome and in model organisms using experiments, and also by using data-mining approaches.

As member of the 1000 Genomes project, our group develops novel computational approaches for genotyping SVs following the sequencing of a thousand (or more) human genomes. Our approaches and those developed by collaborating groups within the consortium are, in the future, expected to help facilitate the design of disease association studies based on massively parallel DNA sequencing technology that may, for example, help elucidate the phenotypic impact of chromosomal aberrations associated with disease.

Combining genomics and computational biology enables deciphering the genomic impact of structural variation.

Selected references

Hasin, Y., Olender, T., Khen, M., Gonzaga-Jauregui, C., Kim, P.M., Urban, A.E., Snyder, M., Gerstein, M.B., Lancet, D. & Korbel, J.O. (2008). High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. *PLoS Genet.*, 4, e1000249

Korbel, J.O., Urban, A.E., Affourtit, J.P., Godwin, B., Grubert, F., Simons, J.F., Kim, P.M., Palejev, D., Carriero, N.J., Du, L., Taillon, B.E., Chen, Z., Tanzer, A., Saunders, A.C., Chi, J., Yang, F., Carter, N.P., Hurles, M.E., Weissman, S.M., Harkins, T.T., Gerstein, M.B., Egholm, M. & Snyder, M. (2007). Paired-end mapping reveals extensive structural variation in the human genome. *Science*, 318, 420-426

Korbel, J.O., Urban, A.E., Grubert, F., Du, J., Royce, T.E., Starr, P., Zhong, G., Emanuel, B.S., Weissman, S.M., Snyder, M. & Gerstein, M.B. (2007). Systematic prediction and validation of breakpoints associated with copy-number variants in the human genome. *Proc. Natl. Acad. Sci. USA*, 104, 10110-10115

Jeroen Krijgsveld

PhD 1999, University of Amsterdam, The Netherlands.

Postdoc at Utrecht University, The Netherlands and Harvard Medical School, Boston, USA.

Assistant Professor, Utrecht University, The Netherlands

Team leader at EMBL since 2008.

Quantitative proteomics

Previous and current research

Our research is centred around mass spectrometry-based proteomics. Mass spectrometry coupled to liquid chromatography has matured to the extent that thousands of proteins can be identified, so for simple organisms we can now start thinking of studying entire proteomes. For more complex organisms, including humans, complementary strategies are still required targeted at specific classes of proteins/peptides by pre-fractionation or selective enrichment. Our interest is in the expansion of this 'proteomic toolbox' and its integration into the larger domains of molecular biology and biochemistry. We focus particularly on quantitative techniques in mass spectrometry using stable isotope-labelling to study protein dynamics in a biological context.

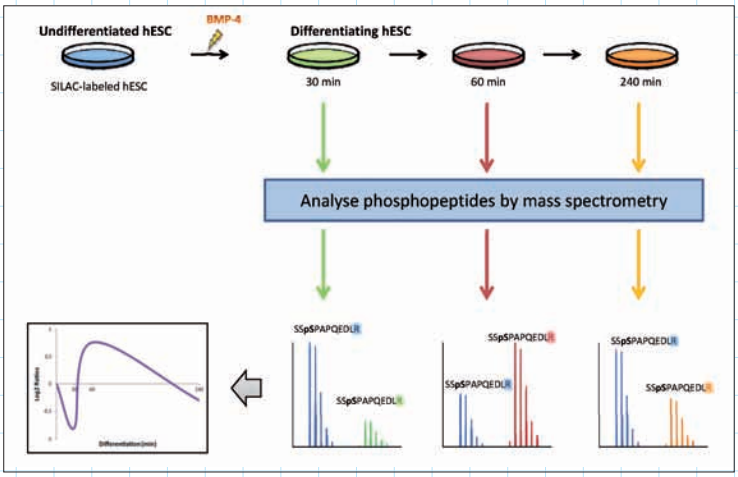
Our tools include stable isotope-labelling for protein quantitation (e.g. SILAC and chemical approaches), enrichment strategies for specific classes of proteins (membrane proteins, phosphopeptides) and separation techniques for detailed coverage of even very complex samples (SCX, nanoflow HPLC, peptide isoelectric focussing (IEF)). Finally, in our newly equipped lab we have state-of-the-art mass spectrometers (a Maxis electrospray Qq-Tof and HCT ion trap) as well as bioinformatic data flows for protein identification and quantitation.

Our biological interest is in developmental biology and in the underlying mechanisms in transcriptional regulation. Currently we are studying the dynamics of protein expression during fruit fly development by comparing distinct embryonic stages. To do this in a quantitative fashion, we have developed tools to metabolically label flies with stable isotopes. In conjunction with gene expression data this provides mechanistic insight in regulation of individual or functionally related proteins and helps explain regulatory mechanisms in early embryonic development. Another major effort is in application of quantitative proteomic techniques to stem cell pluripotency and differentiation. Currently we are studying the dynamics of phosphorylation upon inducing differentiation of human stem cells (see figure). This should benchmark an initial blueprint of the phosphorylation network activated during differentiation, while providing a starting point for further exploration in other cell lines and organisms.

Future projects and goals

Our future work can be divided in three major areas:

- We will apply quantitative proteomic techniques for global analysis of protein expression during embryonic development (stem cells, fruit flies).
- We will focus on the identification of protein-protein and protein-DNA complexes to understand regulatory principles of transcription under various biological conditions.
- In addition we will use targeted mass-spectrometric approaches tracking defined sets of proteins over time employing MRM (multi-reaction monitoring) experiments. We will be specifically focussed on applying this to the dynamics of posttranslational modifications, particularly phosphorylation.



Experimental workflow for the quantitative study of protein phosphorylation during stem cell differentiation. hESC that are labelled with heavy isotopes (SILAC) are induced to differentiate and mixed with non-labelled hESC. Selective isolation of phosphopeptides and mass spectrometric analysis results in the identification of thousands of peptides whose phosphorylation status can be quantified over time.

Selected references

Gouw, J.W., Tops, B.B., Mortensen, P., Heck, A.J. & Krijgsveld, J. (2008). Optimizing identification and quantitation of 15N-labeled proteins in comparative proteomics. *Anal Chem.*, 80, 7796-7803

Van Hoof, D., Pinkse, M.W., Oostwaard, D.W., Mummery, C.L., Heck, A.J. & Krijgsveld, J. (2007). An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nat. Methods*, 4, 677-678

Van Hoof, D., Passier, R., Ward-Van Oostwaard, D., Pinkse, M.W., Heck, A.J., Mummery, C.L. & Krijgsveld, J. (2006). A quest for human and mouse embryonic stem cell-specific proteins. *Mol. Cell Proteomics*, 5, 1261-1273

Krijgsveld, J., Ketting, R.F., Mahmoudi, T., Johansen, J., Artal-Sanz, M., Verrijzer, C.P., Plasterk, R.H. & Heck, A.J. (2003). Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nat. Biotechnol.*, 21, 927-931

Chromatin plasticity and gene regulation mediated by histone variants

Andreas Ladurner

PhD 2000, University of Cambridge.

Postdoctoral work at the Howard Hughes Medical Institute, University of California at Berkeley.

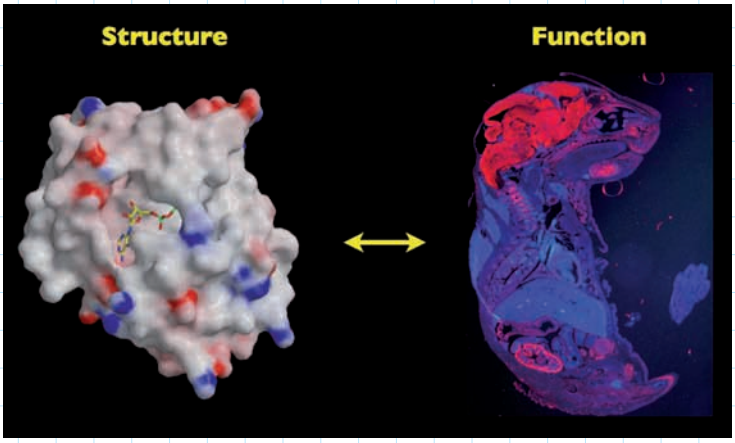
Group leader at EMBL since 2003. Joint appointment with the Structural and Computational Biology Unit.

Previous and current research

Chromatin is intimately involved in the regulation of gene expression. Dynamic changes in the composition or chemical modification of the proteins that wrap our genome critically affect the structure and thus biological function of the chromatin fibre. My lab identifies, characterises and exploits molecular mechanisms underlying plasticity in chromatin structure. Over the years, we have discovered the first example of a protein module capable of recognising a post-translational modification in a histone protein – bromodomains – and reported the first example of an endogenous metabolite that directly binds a histone variant, the macro domain. Our team studies how such post-translational modifications, metabolites and the turnover of histones contribute to chromatin plasticity and how these mechanisms together determine local and global transcriptional outcomes.

Our approach combines genetics, genomics, biochemistry, cell biology, structural biology, biophysics and model organisms to answer fundamental questions in chromatin biology and to identify novel paradigms of molecular recognition and biological regulation.

All living entities are propagated by cell division and the proteins that make up centromeric chromatin are essential in this complex and dynamic process. We have used genetic, biochemical and cell-biological approaches to identify the FACT complex as a cofactor for centromeric silencing in the fission yeast *S. pombe* and discovered a novel histone H3-H4 binding function in one of its domains. We have also identified a repeating short peptide motif in a variety of proteins that specifically bind and modify the function of the RNAi component Argonaute (so-called Ago hooks). We continue to characterise the direct molecular link between cellular NAD metabolism and chromatin structure mediated by the histone macroH2A1.1, and are now starting to obtain insight into the physiological significance of these novel interactions. Finally, we complement our studies in mammalian cells by targeting the poly-ADP-ribose pathway in the fruit fly *Drosophila*, which also serves as a model organism in our long-term systematic and mechanistic effort to dissect the molecular connections between epigenetics, chromatin-based gene regulation and organismal long-term memory formation, consolidation and maintenance.



Our lab pioneered the discovery that a chromatin component, the histone macroH2A1.1, directly binds the small molecule NAD metabolite O-acetyl-ADP-ribose, which is generated by the Sir2 family of histone deacetylases (left). We are now dissecting the potential physiological role of these novel interactions between cellular metabolism and gene regulation.

Future project and goals

- Chromatin-based gene regulation at the level of histone modifications, turnover and replacement;
- Molecular connections between histone macroH2A and cellular NAD metabolism;
- High-resolution structure determination of key chromatin plasticity components;
- Transcriptional and epigenetic basis for organismal long-term memory formation.

Selected references

Stuwe, T., Hothorn, M., Lejeune, E., Rybin, V., Bortfeld, M., Scheffzek, K. & Ladurner, A.G. (2008). The FACT Spt16 'peptidase' domain is a histone H3-H4 binding module. *Proc. Natl. Acad. Sci. USA*, 105, 8884-8889

Li, A.G., Piluso, L.G., Cai, X., Gadd, B.J., Ladurner, A.G. & Liu, X. (2007). An acetylation switch in p53 mediates holo-TFIID recruitment. *Mol. Cell,* 28, 408-421

Till, S., Lejeune, E., Thermann, R., Bortfeld, M., Hothorn, M., Enderle, D., Heinrich, C., Hentze, M.W. & Ladurner, A.G. (2007). A conserved motif in Argonaute-interacting proteins mediates functional interactions through the Argonaute PIWI domain. *Nat. Struct. Mol. Biol.*, 14, 897-903

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

Jürg Müller

PhD 1991, University of Zürich.

Postdoctoral research at the MRC Laboratory of Molecular Biology, Cambridge.

Junior group leader at the Max-Planck-Institute for Developmental Biology, Tübingen.

Group leader at EMBL since 2001.

Chromatin and transcription in development

Previous and current research

Our laboratory studies the biochemical mechanisms by which chromatin-modifying enzymes and chromatin-binding proteins regulate gene transcription. In particular, our work is focussed on the molecular mechanisms by which chromatin proteins encoded by the Polycomb group (PcG) and the trithorax group (trxG) of genes stably and heritably maintain expression states of target genes. PcG and trxG proteins are two evolutionary conserved sets of transcriptional regulators that control a plethora of developmental processes in both animals and plants. PcG proteins act as repressors that keep target genes inactive in cells where these genes should not be expressed, while trithorax group proteins promote transcription of the same target genes in other cells. Although the PcG/trxG system is best known for its role in maintaining spatially-restricted expression of developmental regulator genes in animals and plants, it is also used for processes ranging from X-chromosome inactivation in mammals to the control of flowering time in plants.

We study the PcG/trxG system in the model system *Drosophila* using an integrated approach that combines a variety of biochemical, biophysical, genetic and genomic assays. One focus of our research during recent years has been the biochemical purification of PcG protein complexes and working out their molecular mechanisms. We thus found that PcG protein complexes contain enzymatic activities that add or remove particular post-translational modifications at specific lysine residues in histone proteins in chromatin, including a histone methyltransferase and a histone deubiquitylase. Our analysis of PcG protein complexes also revealed that they contain subunits that allow these complexes to bind to specific post-translational modifications such as methylated lysines on histone proteins. From discovering these activities *in vitro*, we then proceeded to dissect how they regulate gene expression *in vivo*, by studying where PcG and trxG protein complexes bind to target genes in *Drosophila* and how their enzymatic activities modify target gene chromatin. By comparing the chromatin of target genes in wild-type and mutant *Drosophila* strains and by performing structure/function analyses of PcG proteins in *Drosophila*, we have obtained critical insight into the mechanisms by which these chromatin-modifying and -binding activities regulate gene transcription. For these studies we use a combination of detailed in-depth analyses at the single gene level and global analyses at the level of the entire genome.

Future projects and goals

Our long-term goal is to understand how transcriptional ON and OFF states are controlled by the Polycomb/trithorax system and how they are propagated through replication and cell division. The strength of our approach is the combination of *Drosophila* genetics and global genome-wide analyses *in vivo* with detailed in-depth biophysical and biochemical analyses *in vitro*. Whilst we will continue to use these strategies to study the Polycomb/trithorax system, we are moving on to study how chromatin-modifying and chromatin-binding proteins dynamically interact with nucleosomes of defined modification states at the single molecule level. A second recent focus is the functional dissection of the 'histone code' in *Drosophila* using genetic approaches.

Selected references

Oktaba, K, Gutierrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E.M. & Müller, J. (2008). Dynamic regulation by Polycomb group protein complexes controls pattern formation and the cell cycle in *Drosophila*. *Dev. Cell*, 15, 877-889.

de Ayala Alonso, A.G., Gutierrez, L., Fritsch, C., Papp, B., Beuchle, D. & Muller, J. (2007). A genetic screen identifies novel polycomb group genes in *Drosophila*. *Genetics,* 176, 2099-2108

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Kocher, T., Cohen, A., Stunnenberg, H.G., Wilm, M. & Muller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J.*, 26, 4078-4088

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

Structural and Computational Biology Unit

The Structural and Computational Biology Unit pursues an ambitious research programme in integrated structural systems biology. A wide spectrum of expertise allows the unit to tackle problems at different ranges of resolution, connecting atomic structures and dynamic information obtained by X-ray crystallography, NMR with medium-range resolution from single particle electron microscopy, and cellular imaging obtained by EM tomography and light microscopy. Biochemistry, chemical biology, single molecule fluorescence spectroscopy and computational biology complement the structural biology activities and integrate them into a comprehensive description of biological function.

Within the unit, 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 into a comprehensive spatial and temporal description of biological processes. In this way, groups in the Structural and Computational Biology Unit pursue a few common large projects that require the input of 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 the comprehensive structural and temporal description of an entire cell at almost molecular resolution by applying various 'omics' approaches to a small bacterium, by characterising its dynamic protein organisation and merging this molecular information to cellular, high-resolution tomograms.

Currently, the unit consists of twelve research groups with broad methodological experience. It covers electron microscopy (three groups), X-ray crystallography (two groups), NMR (one group), chemical biology (two groups) and computational biology (two groups and two teams). In addition, two groups based in different units have shared appointments with the unit (the Ladurner group, Gene Expression (page 39) and the Nédélec group, Cell Biology (page 17)).

The unit is very well equipped for experimental and computational work. Experimental facilities include a rotating anode and image plate detector for the collection of X-ray diffraction data, crystallisation robot and automated crystal visualisation, 800 MHz, 600 MHz and 500 MHz NMR spectrometers, several transmission electron microscopes and scanning micro-densitometers. There are also facilities for electron cryo-microscopy, cryo-3D tomography, light microscopy, CD and analytical ultracentrifugation, as well as for large scale growth of prokaryotic and eukaryotic cells. The whole computing environment of large central clusters and separate workstations is conveniently networked.

Peer Bork and Christoph Müller
Joint Coordinators, Structural and Computational Biology Unit

Peer Bork

PhD 1990, University of Leipzig.

Habilitation 1995, Humboldt University, Berlin.

At EMBL since 1991.

Joint Unit Coordinator since 2001.

Deciphering function and evolution of biological systems

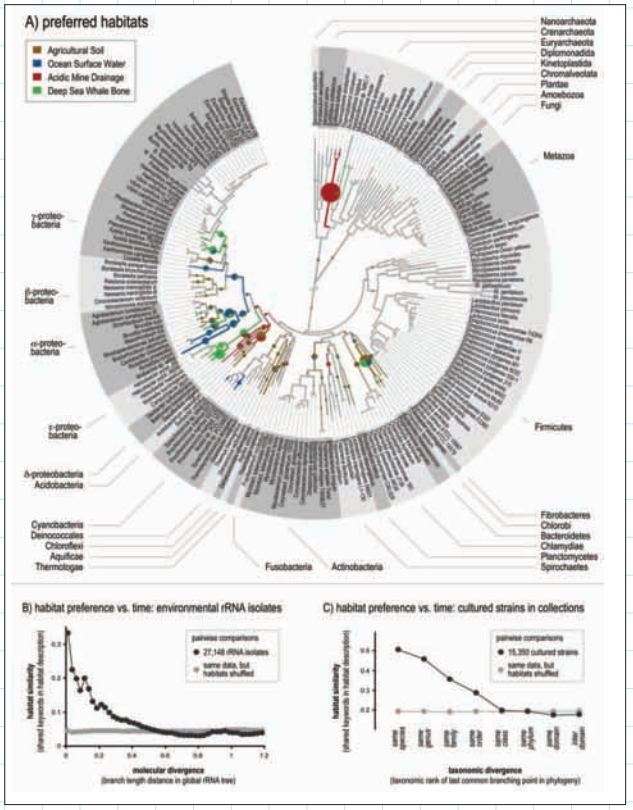
Previous and current research

The main focus of our 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:

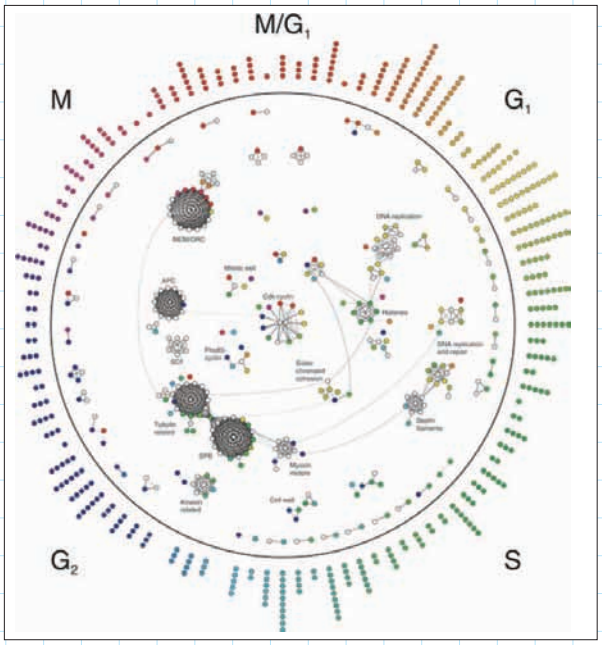
- genes, proteins and small molecules;
- networks and cellular processes;
- phenotypes and environments.

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 in Berlin and with the Molecular Medicine Partnership unit at the Heidelberg university.



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).*



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, *and references therein).*

Selected references

Campillos, M., Kuhn, M., Gavin, A.C., Jensen, L.J. & Bork, P. (2008). Drug target identification using side-effect similarity. *Science*, 321, 263-6

Raes, J. & Bork, P. (2008). Molecular eco-systems biology: towards an understanding of community function. *Nat. Rev. Microbiol.*, 6, 693-9

Sorek, R., Zhu, Y., Creevey, C.J., Francino, M.P., Bork, P. & Rubin, E.M. (2007). Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*, 318, 1449-52

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

Molecular mechanisms of transcriptional regulation and epigenetics

Christoph W. Müller

PhD 1991, University of Freiburg.

Postdoctoral work at Harvard University, Cambridge, Massachusetts.

At EMBL Grenoble since 1995. Joint Unit Coordinator at EMBL Heidelberg since 2007. Joint appointment with the Gene Expression Unit.

Previous and current research

Our group is interested in molecular mechanisms of transcriptional regulation in eukaryotes, where DNA is packaged into chromatin. In the context of chromatin, we are interested how different sequence-specific transcription factors assemble on DNA and how sequence-specific transcription factors interact with co-activators and general transcription factors to recruit RNA polymerases to the transcription start site. We are also studying the overall structure, architecture and inner workings of large molecular machines like RNA polymerases or remodelling factors involved in the transcription process. Finally, we like to gain insight how DNA sequence information and epigenetic modifications act together to regulate gene transcription.

To achieve these goals, we use structural information obtained by X-ray crystallography and electron microscopy combined with other biophysical and biochemical techniques. Systems currently under investigation include transcription factor/DNA complexes, yeast RNA polymerase III and multi-protein complexes involved in chromatin targeting, remodelling 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 of RNA polymerase III, TFIIIC and TFIIIB and their interactions during the RNA polymerase III recruitment process, transcriptional 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, MBT domains, 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 and remodelling complexes, by which mechanisms they are recruited, how they interact with chromatin templates and how their activities are regulated.

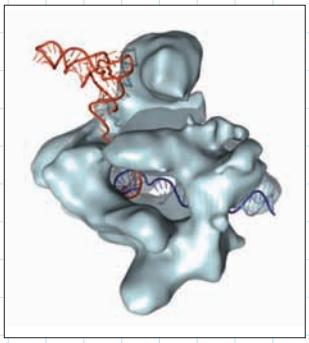


Figure 1 (above): Model of transcribing RNA polymerase III as determined by single particle electron microscopy. The incoming DNA is depicted in blue and white, while the transcribed RNA is depicted in red.

Future projects and goals

- Molecular insights into the recruitment of transcriptional regulator through the combination of DNA sequence-specific recognition and epigenetic modifications.
- Structural and functional analysis of macromolecular machines involved in transcription, chromatin remodelling and chromatin modification.
- Contributing to a better mechanistic understanding of eukaryotic transcription and epigenetics using biochemical and structural biology approaches.

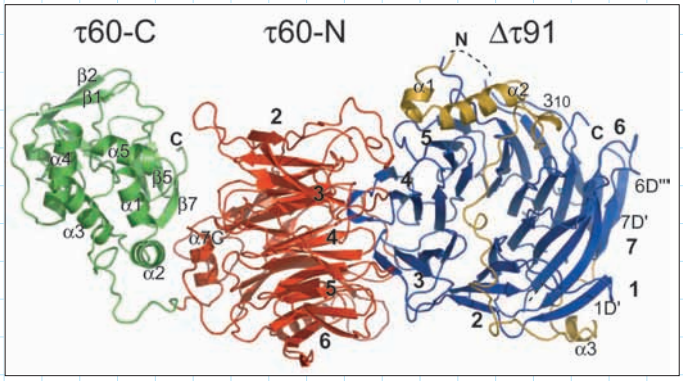


Figure 2 (right): Crystal structure of the τ60/Δτ91 subcomplex of yeast general transcription factor IIIC.

Selected references

Hamès, C., Ptchelkine, D., Grimm, C., Thevenon, E., Moyroud, E., Gérard, F., Martiel, J.L., Benlloch, R., Parcy, F., Müller, C.W. (2008). Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *EMBO J.*, 27, 2628-37

Fernandez-Tornero, C., Böttcher, B., Riva, M., Carles, C., Steuerwald, U., Ruigrok, R.W., 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, 813-23

Grimm, C., de Ayala Alonso, A.G., Rybin, V., Steuerwald, U., Ly-Hartig, N., Fischle, W., Müller, J., Müller, C.W. (2007). Structural and functional analyses of methyl-lysine binding by the malignant brain tumour repeat protein Sex comb on midleg. *EMBO Rep.*, 8, 1031-7

Mylona, A., Fernandez-Tornero, C., Legrand, P., Haupt, M., Sentenac, A., Acker, J., Müller, C.W. (2006). Structure of the τ60/Δτ91 subcomplex of yeast transcription factor IIIC: insights into pre-initiation complex assembly. *Mol. Cell*, 24, 221-232.

John Briggs

PhD 2004, Oxford University.

Postdoctoral research at the University of Munich.

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 during budding and fusion events. To explore this question we are studying a range of different cellular and viral specimens using cryo-electron microscopy and tomography.

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.

Cryo-electron microscopy techniques are particularly appropriate for studying vesicles and viruses 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 higher resolution structures obtained using cryo-electron microscopy and single particle reconstruction of purified complexes.

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 budding events? How do viruses hijack cellular systems for their own use? 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.

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

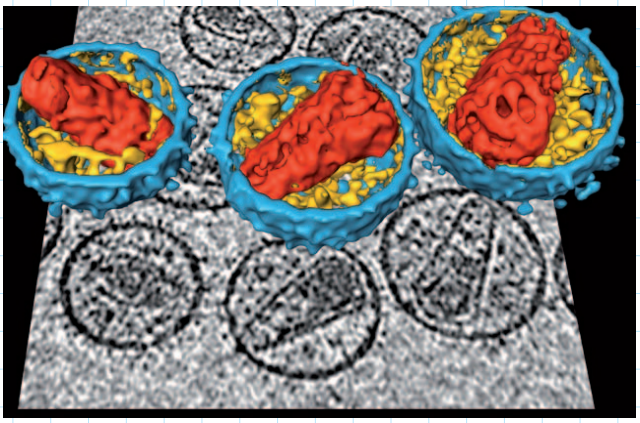
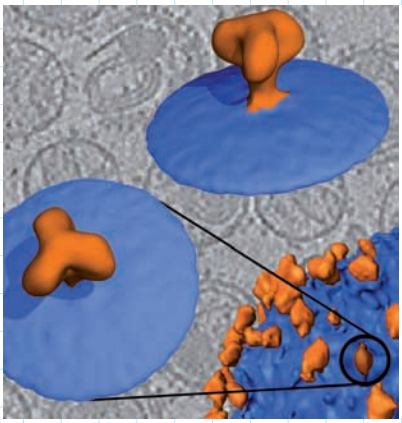


Figure 1: 3D reconstruction of HIV-1 virions using cryo-electron microscopy.



Selected references

Carlson, L.A., Briggs, J.A., Glass, B., Riches, J.D., Simon, M.N., Johnson, M.C., Muller, B., Grunewald, K. & Krausslich, H.G. (2008). Three-dimensional analysis of budding sites and released virus suggests a revised model for HIV-1 morphogenesis. *Cell Host Microbe*, 4, 592-599

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 3D reconstructions of authentic virions. *Structure*, 14, 15-20

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, 157-168

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

Functional mechanisms of complex enzymes in transcription regulation and anti-cancer drugs

Teresa Carlomagno

PhD 1996, University of Naples Federico II.

Postdoctoral research at Frankfurt University and Scripps Research Institute.

Group leader at the MPI for Biophysical Chemistry, Göttingen, 2002-2007.

Group leader at EMBL since 2007. Joint appointment with the Gene Expression Unit.

Previous and current research

We study the structure and dynamics of biomolecular complexes and catalytic RNAs in solution by nuclear magnetic resonance (NMR) spectroscopy in combination with a wide range of biochemical and biophysical techniques. Recent advances in the NMR methodology and instrumentation have allowed overcoming traditional size limitations and have made NMR a very powerful technique, in particular for the investigation of highly dynamic, partially inhomogeneous molecules and complexes.

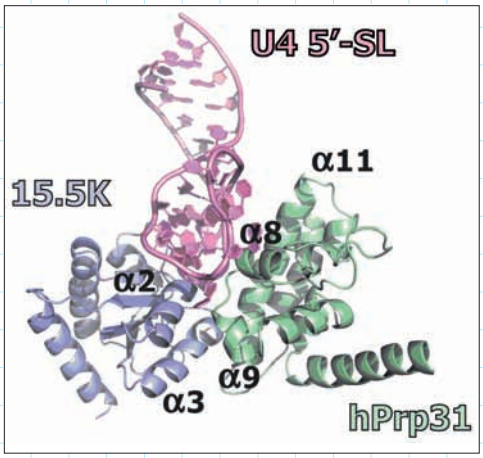
The laboratory focuses on studying: 1) The interaction of small drugs with cellular receptors; and 2) Structure-activity and dynamics-activity relationship of RNP complexes and catalytic RNAs involved in RNA processing.

Conformational switches occur in macromolecular receptors at all cellular levels in dependence of the presence of small organic molecules, which are able to trigger or inhibit specific cellular processes. We develop both computational and experimental tools to access the structure of large receptors in complex with function regulators. In particular we study the functional mechanisms of anti-cancer drug-leads, designed as inhibitors of kinases, proteasome and tubulin.

A second aim of our work is to describe the features of RNA-protein recognition in RNP complex enzymes and to characterise the structural basis for their function. Recently, we have determined the three-dimensional architecture of the ternary complex between the 15.5 kDa protein/U4 RNA and the hPrp31 protein, which is a constituent of the U4/U6 particle in the spliceosome (in collaboration with M. Wahl and R. Lührmann at the MPI, Göttingen; see figure). Currently, we are investigating the nucleolar multimeric box C/D RNP complex responsible for the methylation of the 2'-OH position in rRNA. 2'-O-methylation is one of the most relevant modifications of newly transcribed RNA as it occurs around functional regions of the ribosome. This suggests that 2'-O-methylation may be necessary for proper folding and structural stabilisation of rRNA *in vivo*. In another project, we collaborate with the group of Ramesh Pillai (page 95) at EMBL Grenoble to understand the structure of RNP complexes involved in the regulation of gene expression through small non-coding RNAs.

Future projects and goals

We use NMR spectroscopy to study how proteins and nucleic acids interact with each other and the structural basis for the activity of complex enzymes. In addition we dedicate our efforts to understand the activity of small-molecule inhibitors of cellular targets relevant in anti-cancer therapy. We use innovative NMR techniques to access the structure of large, dynamic multi-component complexes in combination with other structural biology techniques (SANS, X-ray and EM) and biochemical data. Our philosophy is to combine high-resolution structures of single-components of the complexes with both structural descriptors of the intermolecular interactions in solution and computational methods, to obtain an accurate picture of the molecular basis of cellular processes.



NMR-based docking model of the spliceosomal U4-RNA/15.5kD/hPrp31 complex, showing that the Nop-domain of the hPrp31 (green) recognises a composite platform formed by the U4 RNA stem II (pink) and by the 15.5K protein (blue). This complex fold is novel and classifies the Nop-domain as a genuine RNP recognition motif.

Selected references

Kirkpatrick, J.P., Li, P. & Carlomagno, T. (2009). Probing mutation-induced structural perturbations by refinement against residual dipolar couplings: application to the U4 spliceosomal RNP complex. *Chembiochem.*, in press

Li, P., Kirkpatrick, J. & Carlomagno, T. (2009). An efficient strategy for the determination of the 3D architecture of ribonucleoprotein complexes by the combination of a few easily accessible NMR and biochemical data: intermolecular recognition in a U4 spliceosomal complex. *J. Mol Biol.*, in press

Orts, J., Tuma, J., Reese, M., Grimm, S.K., Monecke, P., Bartoschek, S., Schiffer, A., Wendt, K.U., Griesinger, C. & Carlomagno, T. (2008). Crystallography-independent determination of ligand binding modes. *Angew Chem. Int. Ed. Engl.*, 47, 7736-40

Liu, S., Li, P., Dybkov, O., Nottrott, S. *et al*. (2007). Binding of the human Prp31 Nop domain to a composite RNA-protein platform in U4 snRNP. *Science*, 316, 115-20

Reese, M., Sanchez-Pedregal, V.M., Kubicek, K., Meiler, J. *et al*. (2007). Structural basis of the activity of the microtubule-stabilizing agent epothilone A studied by NMR spectroscopy in solution. *Angew Chem. Int. Ed. Engl.*, 46, 1864-8

Anne-Claude Gavin

PhD 1992, University of Geneva.

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

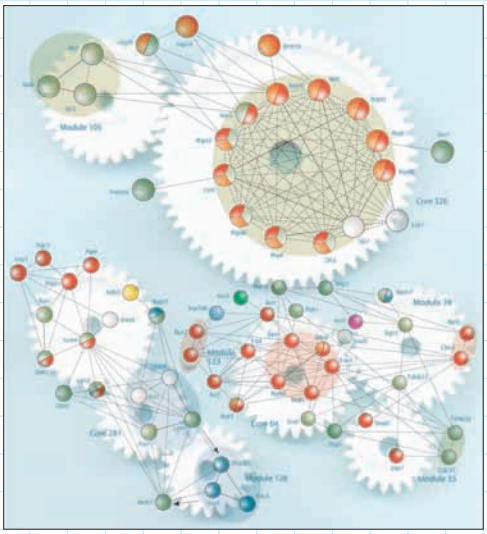
Previous and current research

How is biological matter organised? Can the protein and chemical worlds be matched to understand the cell's inner works? As our knowledge about the basic building blocks of eukaryotic genomes, proteomes and metabolomes grows, the challenge remains to understand 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 focussed 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, and follow three main lines of research to understand the principles that govern the assembly of these networks.

The charting of protein-protein interaction networks: Our knowledge of protein-protein interaction is still anecdotal; current estimations reveal that probably less than 10% have been characterised so far. We adopted 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 programme, which aimed 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 functioning of the whole cell. 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; this represents the second major research interest of the group.



Graphic: Petra Riedinger

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.

Selected references

Charbonnier, S., Gallego, O. & Gavin, A.C. (2008). The social network of a cell: Recent advances in interactome mapping. *Biotechnol. Annu. Rev.*, 14, 1-28

Kuhner, S. & Gavin, A.C. (2007). Towards quantitative analysis of proteome dynamics. *Nat. Biotechnol.*, 25, 298-300

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M. *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

Biological sequence analysis

Toby Gibson

PhD 1984, Cambridge University.

Postdoctoral research at the Laboratory of Molecular Biology, Cambridge.

Team leader at EMBL since 1986.

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 and elsewhere. We are probably best known for our involvement with the Clustal W and Clustal X programs that are widely used for multiple sequence alignment, working 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 >18,000 reported phosphorylation sites; and GlobPlot, a tool for exploring protein disorder.

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 to help users find functional sites in modular protein sequences. Short functional sites (e.g. figure 1) are used for the dynamic regulation of large cellular protein complexes and their characterisation is essential for understanding cell signalling. So-called 'hub' proteins that make many contacts in interaction networks are thought to have abundant regulatory motifs in large segments of IUP (intrinsically unstructured protein segments). Freely available ELM resource data is now used by many bioinformatics groups to improve prediction of linear motif interactions, e.g. the NetworKIN kinase-substrate predictor and the DILIMOT and SLIMFinder novel motif predictors.

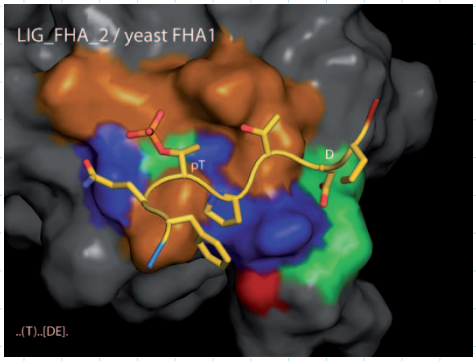


Figure 1 (above): Structure of a typical linear motif-ligand domain interaction. Here the Rad9 FHA domain is bound to a phosphothreonine peptide (pdb:1K3N). Annotated in ELM as LIG_FHA_2.

Future projects and goals

Computers are applied in molecular biology in the hope that, ultimately, they will inform experimental strategies. As an example, we have recently proposed new candidate KEN boxes, a sequence motif that targets cell cycle proteins for destruction in anaphase (figure 2). 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 are currently working to add structure and conservation filtering to ELM. We will apply the tools in the investigation of modular protein function and may deploy them in proteome and protein network 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.

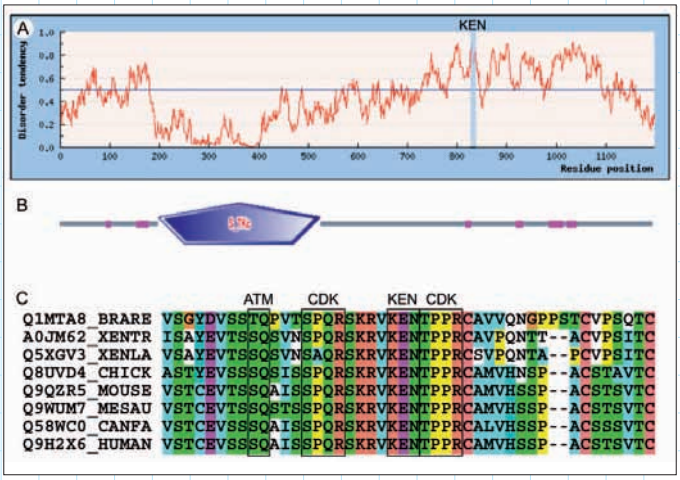


Figure 2: A candidate KEN box in the important cell cycle kinase Hipk2. The sequence segment is predicted to be natively disordered and has many conserved phosphorylation motifs as well as the KEN motif. (Michael et al.*, 2008)*

Selected references

Diella, F., Gould, C.M., Chica, C., Via, A. & Gibson, T.J. (2008). Phospho.ELM: a database of phosphorylation sites – update 2008. *Nucleic Acids Res.*, 36, D240-D244

Diella, F., Haslam, N., Chica, C., Budd, A., Michael, S., Brown, N.P., Trave, G. & Gibson, T.J. (2008). Understanding eukaryotic linear motifs and their role in cell signaling and regulation. *Front Biosci.*, 13, 6580-6603

Michael, S., Trave, G., Ramu, C., Chica, C. & Gibson, T.J. (2008). Discovery of candidate KEN-box motifs using cell cycle keyword enrichment combined with native disorder prediction and motif conservation. *Bioinformatics*, 24, 453-457

Perrodou, E., Chica, C., Poch, O., Gibson, T.J. & Thompson, J.D. (2008). A new protein linear motif benchmark for multiple sequence alignment software. *BMC Bioinformatics*, 9, 213

Edward Lemke

PhD, MPI for Biophysical Chemistry, Göttingen.

Research Associate, the Scripps Research Institute.

Group leader at EMBL since 2009. Joint appointment with Cell Biology and Biophysics Unit.

Structural light microscopy/single molecule spectroscopy

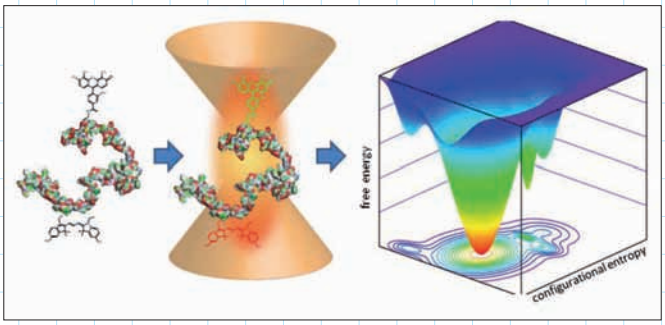
Previous and current research

Research in our laboratory combines modern chemical biology and biochemistry/molecular biology methods with advanced fluorescence and single molecule techniques to elucidate the nature of protein disorder in biological systems and disease mechanisms.

Currently, more than 50,000 protein structures with atomic resolution are available from the protein databank and due to large efforts (mainly crystallography and NMR) their number is rapidly growing. However, even if all 3D protein structures were available, our view of the molecular building blocks of cellular function will still be rather incomplete, as we now know that many proteins are intrinsically disordered, which means that they are unfolded in their native state. Interestingly, the estimated percentage of intrinsically disordered proteins (IDPs) grows with the complexity of the organism (prokaryotes ≈ 5% and eukaryotes ≈ 50%). In a modern view of systems biology, these disordered proteins are believed to be multi-functional signalling hubs central to the interactome (the whole set of molecular interactions in the cell). Their ability to adopt multiple conformations is considered a major driving force behind their evolution and enrichment in eukaryotes.

While the importance of IDPs in biology is now well established, many common strategies for probing protein structure are incompatible with molecular disorder and the highly dynamic nature of those systems. In addition, a mosaic of molecular states and reaction pathways can exist in parallel in any complex biological system, further complicating the situation to measure these systems. For example, some proteins might behave differently than the average, giving rise to new and unexpected phenotypes. One such example are the infamous Prion proteins, where misfolding of only subpopulations of proteins can trigger a drastic signalling cascade leading to completely new phenotypes. Conventional ensemble experiments are only able to measure the average behaviour of a system, discounting such coexisting populations and rare events. Ignoring such information can easily lead to generation of false or insufficient models, which may further impede our understanding of the biological processes and disease mechanisms.

In contrast, single molecule techniques, which probe the distribution of behaviours, can shed light on important mechanisms that otherwise remain masked. In particular, single molecule fluorescence (smF) studies allow probing of molecular structures and dynamics on the nanometer scale with high time resolution. Although not inherently limited by the size of a macromolecule, smF studies require site-specific labelling with special fluorescent dyes which still hampers the broad application and general use of this technique. It was recently demonstrated that amber nonsense suppression technology of genetically reprogrammed hosts is an especially powerful approach to overcome this limitation (Brustad *et al.*, 2008). Here, unnatural amino acids with unique chemical properties are conveniently site-specifically introduced into any protein site by the host organism itself, serving as manipulation sites. Our lab also continues to develop and apply such protein engineering tools to facilitate fluorescence studies of complex biological mechanisms.



Labelled proteins are excited using advanced laser techniques and emitted fluorescence photons are detected using home-built highly sensitive equipment. This strategy allows to study structure and dynamics of even heterogeneous biological systems.

Future projects and goals

Recent studies have shown that even the building blocks of some of the most complex and precise machines with an absolute critical role to survival of the cell, such as DNA packing and many transport processes, are largely built from IDPs. We aim to explore the physical and molecular rationale behind the fundamental role of IDPs by combining molecular biology and protein engineering tools with single molecule biophysics. Our long-term goal is to develop general strategies to study structure and dynamics of IDPs within their natural complex environments.

Selected references

Ferreon, A.C.M., Gambin, Y, Lemke, E.A., Deniz A.A. (2009). Single-Molecule Fluorescence illuminates a multi-conformational switch in α-synuclein. *Proc. Natl. Acad. Sci. USA*, doi:10.1073/pnas.0809232106

Brustad, E.M., Lemke, E.A., Schultz, P.G. & Deniz, A.A. (2008). A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. *J. Am. Chem. Soc.*, 130, 17664-5

Deniz, A.A., Mukhopadhyay, S. & Lemke, E.A. (2008). Single-molecule biophysics: at the interface of biology, physics and chemistry. *J. R. Soc. Interface,* 5, 15-45

Lemke, E.A., Summerer, D., Geierstanger, B.H., Brittain, S.M. & Schultz, P.G. (2007). Control of protein phosphorylation with a genetically encoded photocaged amino acid. *Nat. Chem. Biol.*, 3, 769-72

Structural bioinformatics

Rob Russell

PhD 1993, Oxford University.

Postdoctoral research at Imperial Cancer Research Fund, London.

Senior Investigator at SmithKline Beecham, London.

Group leader at EMBL since 2001. Joint appointment with EMBL-EBI.

Previous and current research

Our research is focussed broadly on the relationships between protein three-dimensional (3D) structure, function and evolution. A major interest over the past few 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.

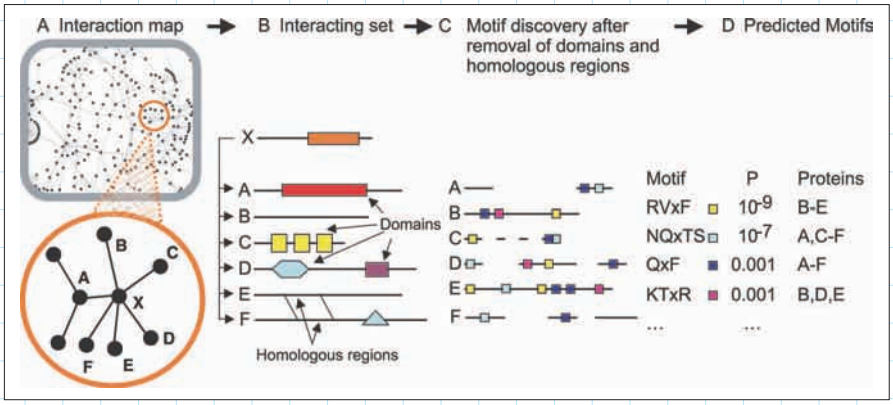
Protein interactions and complexes: Protein interaction networks are central to any understanding of cellular processes, and though many thousands are now known, few initiatives to uncover them pay much attention to one of the best sources of data available: complexes of known 3D structure. We thus study 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

Linding, R., Jensen, L.J., Ostheimer, G.J., van Vugt, M.A., Jørgensen, C., Miron, I.M., Diella, F., Colwill, K., Taylor, L., Elder, K., Metalnikov, P., Nguyen, V., Pasculescu, A., Jin, J., Park, J.G. *et al.* (2007). Systematic discovery of *in vivo* phosphorylation networks. *Cell*, 129, 1415-26

Gavin, A.C., Aloy, P., Grandi, P., Krause, R., Boesche, M. *et al.* (2006). Proteome survey reveals modularity of the yeast cell machinery. *Nature*, 440, 631-636

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 1995, Max-Planck-Institut für Medizinische Forschung, Heidelberg.

Postdoctoral research at the Max-Planck-Institut für Molekulare Physiologie, Dortmund.

Team leader at EMBL since 1999.

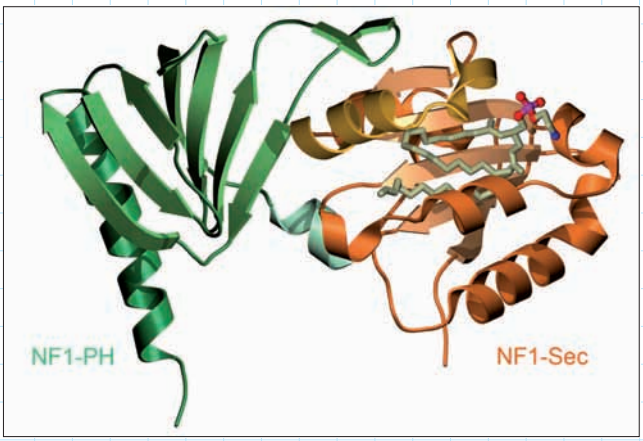
Signal transduction – disease proteins

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 a major focus is on neurofibromatosis type 1 (NF1), a genetic disease with an incidence of 1 in 3,500 newborns. NF1 patients have an increased tumour risk, may show a variety of developmental defects and frequently have learning disabilities. The NF1 gene encodes a huge protein (20 times larger than the oxygen carrier protein myoglobin), 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 hyperactive. The GAP activity of neurofibromin resides in a segment which represents only 10% of the protein and remains the only clearly defined biochemical function of the protein.

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 this approach, 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, binding cellular glycerophospholipids.



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.

Future projects and goals

A major goal is to arrive at a 3D model of neurofibromin. In addition to the 'divide and conquer' strategy, we are gradually returning to the 'conquer only' approach by trying to overexpress the full-length neurofibromin in various eukaryotic hosts. With its availability we will also consider electron microscopy to study its structure. 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.

Further projects of the laboratory include signalling by eukaryotic and prokaryotic protein kinases, novel phosphoryltransfer systems, structural neurobiology and the regulation of retroviral transcription.

Selected references

Hothorn, M., Neumann, H., Lenherr, E.D., Wehner, M., Rybin, V., Hassa, P.O., Uttenweiler, A., Reinhardt, M., Schmidt, A., Seiler, J., Ladurner, A.G., Herrmann, C., Scheffzek, K., Mayer, A. (2009). Catalytic core of a membrane-associated eukaryotic polyphosphate polymerase. *Science*, in press

Anand, K., Schulte, A., Vogel-Bachmayr, K., Scheffzek, K. & Geyer, M. (2008). Structural insights into the cyclin T1-Tat-TAR RNA transcription activation complex from EIAV. *Nat. Struct. Mol. Biol.*, 15, 1287-1292

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

Data integration and knowledge management

Reinhard Schneider

PhD 1994, University of Heidelberg.

Postdoctoral research at EMBL.

Co-founder and Chief Information Officer at LION bioscience AG.

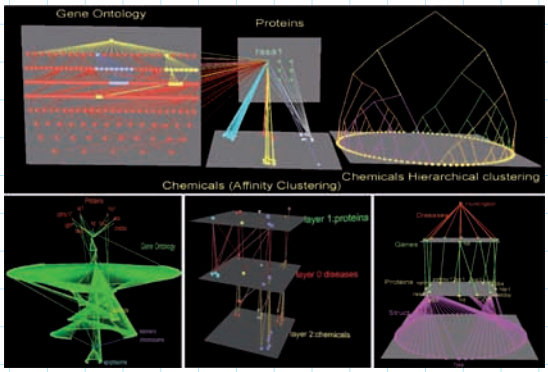
Chief Executive Officer at LION bioscience Research Inc., Cambridge, MA.

Team leader at EMBL since 2004.

Previous and current research

Today it is widely recognised that a comprehensive integration of data can be one of the key factors to improve productivity and efficiency in the biological research process. Successful data integration helps researchers to discover relationships that enable them to make better and faster decisions, thus considerably saving time and money.

Over the last 20 years, biological research has seen a very strong proliferation of data sources. Each research group and 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.



Examples of the graphical features of Arena3D. Heterogeneous data types can be visualised in a 3D environment and a range of layout and cluster algorithms can be applied.

The current systems biology approaches are generating data sets with rapidly growing complexity and dynamics. One major challenge is to provide the mechanism for accessing the heterogeneous data and to detect the important information. We develop interactive visual data analysis techniques using automatic data analysis pipelines. The combination of techniques allows us to analyse otherwise unmanageable amounts of complex data.

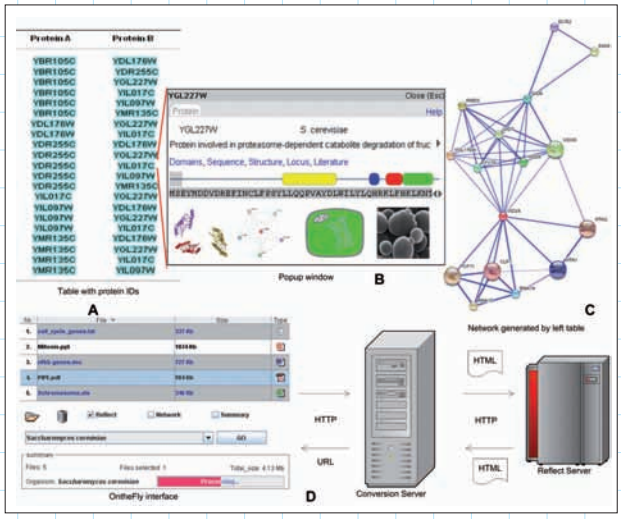
The principal aim of the group is to capture and centralise the knowledge generated by the scientists in the several divisions, and to organise that knowledge such that it can be easily mined, browsed and navigated. By providing access to all scientists in the organisation, it will foster collaborations 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 scientific data portals;
- Providing access to, and developing further, data-mining tools (e.g. text-mining);
- Visualisation environment for systems biology data.

Future projects and goals

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



OnTheFly and Reflect server. Figure (A,B,C) shows an annotated table (A) of an PDF full text article, the generated popup window with information about the protein YGL227W (B), and an automatically generated protein-protein interaction network (C) of associated entities for the proteins shown in part (A). Part (D) shows the architecture and functionality.

Selected references

Pavlopoulos, G.A., O'Donoghue, S.I., Satagopam, V.P., Soldatos, T.G., Pafilis, E. & Schneider, R. (2008). Arena3D: visualization of biological networks in 3D. *BMC Syst. Biol.*, 2, 104

Erhardt, R.A., Schneider, R. & Blaschke, C. (2006). Status of text-mining techniques applied to biomedical text. *Drug Discov. Today*, 11, 315-325

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

Directors' Research

Directors' Research is unlike other EMBL units in that it covers two thematically distinct research groups, headed by the Director General and Associate Director of EMBL. As the DG and AD have substantial management responsibility for all the units of EMBL, their laboratories are administratively separated from the other units.

The Mattaj Group has studied 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. Their studies have demonstrated 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. Currently they are focussed on identifying the factors that are involved in NE assembly and their modes of action. 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 post-transcriptional regulation of gene expression and in mammalian iron metabolism with research on diseases that result from disturbances in both areas. Their work on post-transcriptional 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 their exploration of the IRE/IRP network in mammalian iron homeostasis. 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 RanGTPase as a spatial regulator

Iain Mattaj

PhD 1979, University of Leeds.

Postdoctoral work at the Friedrich Miescher Institute and the Biocenter, Basel.

Group leader at EMBL since 1985. Programme Coordinator since 1990. Scientific Director of EMBL 1999–2005. Director General since 2005.

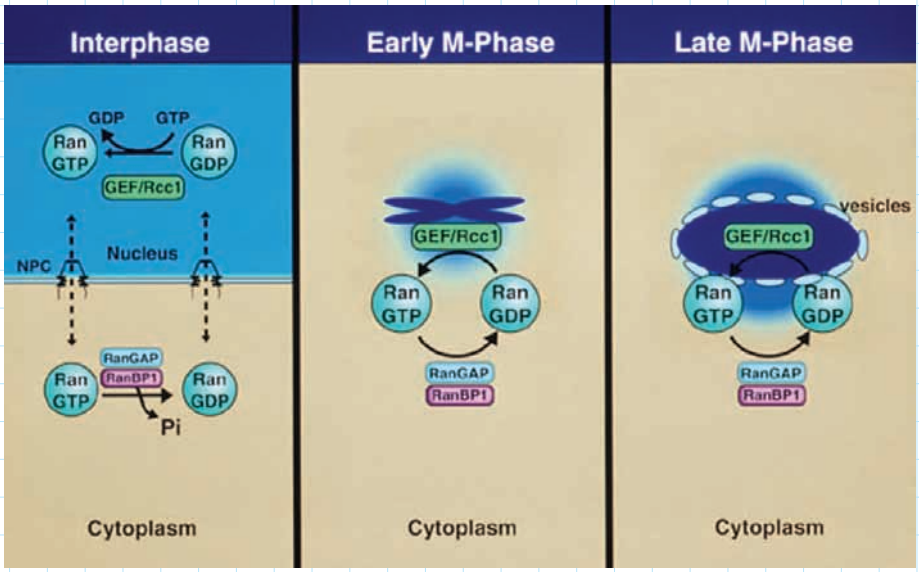
Previous and current research

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 focussing 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 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., Antonin, W., Jaedicke, A., Sachse, M., Santarella, R., Haselmann, U. & Mattaj, I. (2008). A role for gp210 in mitotic nuclear-envelope breakdown. *J. Cell Sci.*, 121, 317-328

Franz, C., Walczak, R., Yavuz, S., Santarella, R., Gentzel, M., Askjaer, P., Galy, V., Hetzer, M., Mattaj, I.W. & Antonin, W. (2007). MEL-28/ELYS is required for the recruitment of nucleoporins to chromatin and postmitotic nuclear pore complex assembly. *EMBO Rep.*, 8, 165-172

Gorjanacz, M., Klerkx, E.P., Galy, V., Santarella, R., Lopez-Iglesias, C., Askjaer, P. & Mattaj, I.W. (2007). *Caenorhabditis elegans* BAF-1 and its kinase VRK-1 participate directly in post-mitotic nuclear envelope assembly. *EMBO J.*, 26, 132-143

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

Cytoplasmic gene regulation and molecular medicine

Matthias W. Hentze

MD 1984, University of Münster.

Postdoctoral training at the NIH, Bethesda.

Group leader at EMBL since 1989; Senior Scientist since 1998. Co-Director of the EMBL/University of Heidelberg Molecular Medicine Partnership Unit since 2002. Associate Director since 2005.

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 (figure 1). We use mostly biochemical approaches and mammalian, yeast and *Drosophila* model systems.

Within the Molecular Medicine Partnership Unit (MMPU), we are investigating the posttranscriptional processes of nonsense-mediated decay (NMD) and 3' end processing and their importance in genetic diseases (with Andreas Kulozik). We also study the role of miRNAs in cancer and other diseases (with Andreas Kulozik and Martina Muckenthaler).

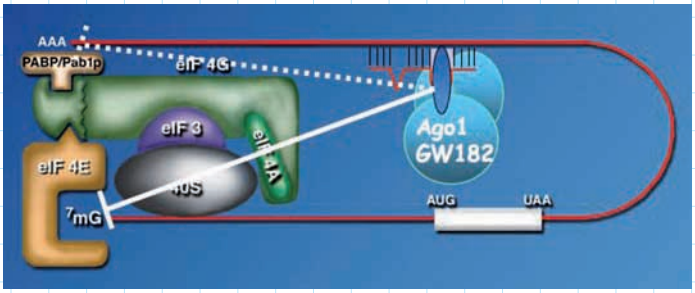


Figure 1: A two-hit model explaining miR-mediated repression.

Our second major interest is the systems biology of mammalian iron metabolism (figure 2). This work includes the system-wide exploration of the functions of the IRE/IRP regulatory network. Within the MMPU (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.

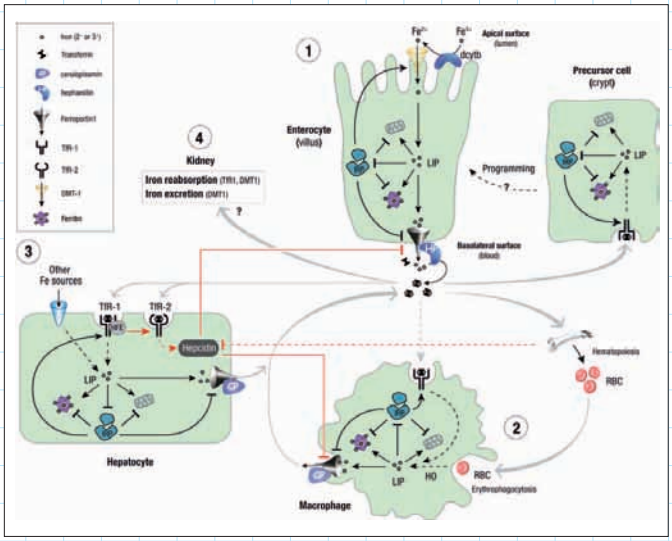


Figure 2: Systems biology of mammalian iron metabolism.

For research themes and projects of the teams in the MMPU, see The Molecular Medicine Partnership Unit (MMPU), University Hospital Heidelberg and www.embl.org/research/partners/mmpu/index.html.

Selected references

Galy, B., Ferring-Appel, D., Kaden, S., Grone, H.J. & Hentze, M.W. (2008). Iron regulatory proteins are essential for intestinal function and control key iron absorption molecules in the duodenum. *Cell Metab.*, 7, 79-85

Thermann, R. & Hentze, M.W. (2007). *Drosophila* miR2 induces pseudo-polysomes and inhibits translation initiation. *Nature*, 447, 875-878

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

Hentze, M.W., Muckenthaler, M.U. & Andrews, N.C. (2004). Balancing acts: molecular control of mammalian iron metabolism. *Cell*, 117, 285-297

Core Facilities

The Core Facilities at EMBL in Heidelberg were established soon after the Scientific Advisory Committee and EMBL Council approved the new scientific directions for the 2001-2005 Scientific Programme. EMBL Council was generous enough to provide some extra funding in the context of the new indicative scheme for this initiative. The main idea behind this decision was to try to provide a number of high level support teams that would help EMBL's scientific community by providing easy access to well-equipped facilities, both in terms of human resources and state of the art equipment.

The basic concept for the Core Facilities had been tested some years earlier (starting in 1998) when the Advanced Light Microscopy Facility (ALMF) was created. The ALMF was considered as the 'gold standard' for the establishment of the new facilities. The support activities need to be tailored to the demands of the community and the staff members hired for such activities need to clearly understand their role. The support activities also need to evolve with the science carried out in the units of the laboratory. Having functioning Core Facilities proves to be advantageous for the scientists using them, not only in terms of workload and efficiency aspects but also in terms of cost of projects. The continuous investment in state-of-the-art technologies allows EMBL scientists and external visitors from our member states to have access to a set of advanced technology platforms.

Today, the facilities are Advanced Light Microscopy, Genomics, Proteomics, Protein Expression and Purification, Electron Microscopy, Flow Cytometry and Chemical Biology.

Christian Boulin
Unit Coordinator of Core Facilities and Services

Rainer Pepperkok

PhD 1992, University Kaiserslautern.

Postdoctoral research at University of Geneva.

Lab Head at the Imperial Cancer Research Fund, London.

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 regularly in-house and international courses to teach advanced light microscopy methods.

Major projects and accomplishments

- The ALMF presently manages 18 top-of-the-line microscope systems and five 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 reached 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*
- Lambert Instruments*
- Leica Microsystems*
- Olympus Europe*
- Perkin Elmer*
- Scientific Volume Imaging
- Visitron*

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

Selected references

Simpson, J.C., Cetin, C., Erfle, H., Joggerst, B., Liebel, U., Ellenberg, J. & Pepperkok, R. (2008). An RNAi screening platform to identify secretion machinery in mammalian cells. *J Biotechnol.,* 129, 352-65

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

Pepperkok, R. & Ellenberg, J. (2006). High-throughput fluorescence microscopy for systems biology. *Nat. Rev. Mol. Cell Biol.,* 7, 690-6

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

Chemical Biology Core Facility

Joe Lewis

PhD 1991, Institute of Molecular Pathology, Vienna.

Postdoctoral research at EMBL.

Group leader at the Wellcome Trust Centre for Cell Biology, Edinburgh.

Group and Global HCV Project Leader at Anadys Pharmaceuticals, Heidelberg.

MBA 2009, Essec and Mannheim Business School. Facility Head at EMBL since 2004.

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, the DKFZ (German Cancer Research Center, Heidelberg) and the University of Heidelberg to provide the infrastructure and expertise to open up small molecule development to research groups at all three 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 institute to institute, several of which have completed screening and are now in the lead optimisation phase.

Services provided

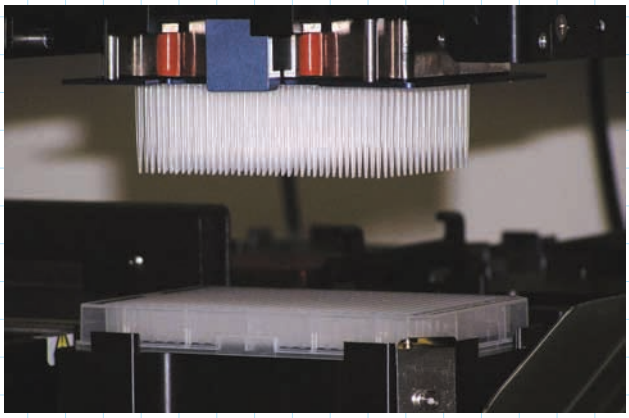
The new Chemical Biology Core Facility screening library is composed of 79,000 compounds. The selection focussed on compound catalogues from three leading vendors in the field. Each vendor offers access to significantly larger collections with low redundancy and highly competitive prices coupled with attractive options for resupply and follow-up synthesis services. The selected compounds were checked for drug-likeness, structural and shape diversity, novelty and compliance with medicinal chemistry requirements. Individual compound selection was done by picking representative compounds around selected scaffolds. A scaffold-based selection offers the advantage of high information screening: since the structural space around each scaffold is covered appropriately, any hit compounds from a high-throughput screen can be rapidly followed up by selecting similar compounds to enable initial structure-activity relationships to be discerned. This will help in the prioritisation of the hit compounds for further medicinal chemistry optimisation.

Further services include:

- Selection of appropriate assay technology platforms;
- Developing assays for medium-throughput screening;
- Assisting in the design of secondary specificity assays;
- Compound characterisation;
- Managing compound acquisition through our chemistry partners;
- Computation screening using Surflex.

Partners

- Technology partners: Perkin Elmer, IDBS.
- Chemistry partners: Tripos Inc., AMRI, Chembridge and Enamine.



Parallel pipetting of samples in 384-well format.

Selected references

Polycarpou-Schwarz, M., Muller, K., Denger, S., Riddell, A., Lewis, J.D., Gannon, F. & Reid, G. (2007). Thanatop: a novel 5-nitrofuran that is a highly active, cell-permeable inhibitor of topoisomerase II. *Cancer Res.*, 67, 4451-4458

Sehr, P., Pawlita, M. & Lewis, J. (2007). Evaluation of different glutathione S-transferase-tagged protein captures for screening E6/E6AP interaction inhibitors using AlphaScreen (R). *J. Biomol. Screen.*, 12, 560-7

Lewis, J.D. (2007). The role of the chemical biology core facility at EMBL: a vision for a European roadmap. *ACS Chem Biol.*, 2, 21-3

Neduva, V., Linding, R., Su-Angrand, I., Stark, A., de Masi, F., Gibson, T.J., Lewis, J.D., Serrano, L. & Russell, R.B. (2005). Systematic discovery of new recognition peptides mediating protein interaction networks. *PLoS Biol.*, 3, e405

Claude Antony

PhD 1984, Université Paris VI.

Postdoctoral research at EMBL 1987-1989.

Group Leader at CNRS 1994-2003.

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, in particular the newly installed electron tomography set-up. Techniques can be applied and adapted to various projects across the units to access EM resolution at the level of cell organisation. The facility also trains new users to make best use of our advanced equipment and develops new approaches and methods in EM applications to cellular and developmental biology.

Major projects and accomplishments

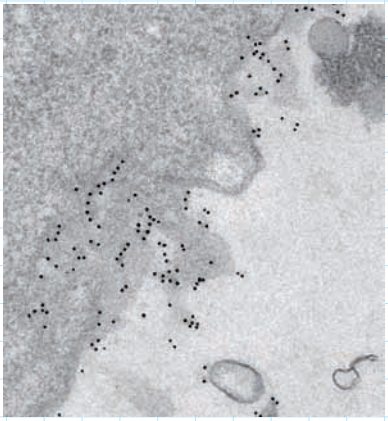
Our new Electron Tomography equipment includes a new microscope and computing set-up with effective programs for 3D reconstruction and image modelling. Mostly destined to perform cellular tomography of plastic embedded samples, the new microscope is a FEI F30 (300 kV microscope with a Field Emission Gun and FEI Eagle 4K camera, running serialEM as acquisition software (Univ. of Boulder, CO)). It is also equipped with a cryoholder to support cryo-EM investigations. This microscope is managed by specialised EM engineers who hold the knowledge for tomography data acquisition and processing, teach users the various applications for cellular structure modelling and introduce them to the F30 microscope.

Correlative microscopy technology (a collaboration between EMCF and ALMF, pages 58 and56) has been established with conventionally fixed cells grown on coverslips (Colombelli *et al.*, 2008, *Methods in Molecular Biology*). We are now developing a similar method adapted to cells grown on sapphire coverslips which are destined to be cryofixed by high-pressure freezing after LM visualisation.

Other projects include main investigations on correlative microscopy and the study of membrane repair (Schultz group, page 19; see figure); *de novo* formation of mammalian Golgi complex (Pepperkok group, page 18); tomography reconstruction of *Drosophila* embryo dorsal closure (Frangakis (ex-EMBL) and Brunner groups, page 11); on Denge virus replication and assembly sites (R. Bartenschlager, University of Heidelberg).

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 protocols;
- Electron Tomography, image acquisition (F30) and data processing for plastic embedded samples;
- Assisting users in choosing the right methods and protocols for their project;
- Organising courses and lectures on EM methods in cell biology.



Membrane fragments labelled against annexin 4 (picture by Charlotta Funaya).

Technology partners

- FEI Company: Supplier of advanced electron microscopes, including the new tomography microscope.
- Leica-microsystems is the constructor of our HPFreezer EMPACT2, a portable machine with an optional attachment, the Rapid Transfer System (RTS), which permits easy loading of the samples and allows correlative light and electron microscopy. They are also our supplier for the various ultramicrotomes units we use for sample plastic- or cryo-sectioning. Labtec and Abrafluid are providers for our HPM-010 device that we use for cryofixation of large biological specimens.

Selected references

Colombelli, J., Tangemo, C., Haselman, U., Antony, C., Stelzer, E.H., Pepperkok, R. & Reynaud, E.G. (2008). A correlative light and electron microscopy method based on laser micropatterning and etching. *Methods Mol Biol.*, 457, 203-13

Di Ventura, B., Funaya, C., Antony, C., Knop, M. & Serrano, L. (2008). Reconstitution of Mdm2-dependent post-translational modifications of p53 in yeast. *PLoS ONE*, 3, e1507

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

Toya, M., Sato, M., Haselmann, U., Asakawa, K., Brunner, D., Antony, C. & Toda, T. (2007). Gamma-tubulin complex-mediated anchoring of spindle microtubules to spindle-pole bodies requires Msd1 in fission yeast. *Nat. Cell Biol.*, 9, 646-653

Flow Cytometry Core Facility

Andrew Riddell

BSc Hons 1992, Paisley University.

PgDip, 1993, Caledonian University, Glasgow.

Work at the MRC LMB and CIMR and Hutchison/MRC, Cambridge.

Facility Head at EMBL since 2003.

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 uses 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.

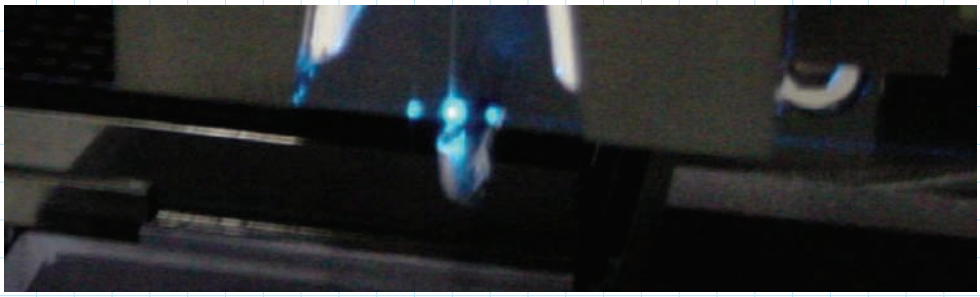
Services provided

- 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

Polycarpou-Schwarz, M., Muller, K., Denger, S., Riddell, A., Lewis, J.D., Gannon, F. & Reid, G. (2007). Thanatop: a novel 5-nitrofuran that is a highly active, cell-permeable inhibitor of topoisomerase II. *Cancer Res.*, 67, 4451-8

Füssl, A., Schleifenbaum, A., Goritz, M., Riddell, A., Schultz, C. & Kramer, R. (2006). Cellular uptake of PNA-terpyridine conjugates and its enhancement by Zn(2+) Ions. *J. Am. Chem. Soc.*, 128, 5986-7

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

Vladimír Beneš

PhD 1994, Czech Academy of Sciences, Prague.

Postdoctoral research at EMBL.

Facility Head since 2001.

Genomics Core Facility

The Genomics Core Facility (www.genecore.embl.de) is EMBL's in-house genomics service centre equipped with state-of-the-art technologies required for functional genomics analyses and operated by highly-qualified staff. As training is an inseparable part of GeneCore activities, its staff is involved in tutoring individual researchers as well as in organising practical courses on subjects such as quantitative real-time RT-PCR (qPCR), gene expression profiling and data analyses.

Major projects and accomplishments

The acquisition of new-generation sequencing technology is vital to keep EMBL at the forefront of European research and it was decided to acquire the Genome Analyzer (Solexa) by Illumina (pictured below). The system became operational in Spring 2008 and is used for following analyses in a single-read or a paired-end modus:

- Location analysis of nucleic acids-protein interactions – ChIP-Seq, CLIP-Seq
- Transcriptome sequencing – mRNA-Seq
- Discovery of small non-coding RNAs – ncRNA-Seq
- DNA methylation profiling – MeDIP-Seq
- *De novo* sequencing and resequencing

For analysis of large volumes of data generated with new-generation sequencing technology GeneCore has expanded collaboration with a German company Genomatix and obtained their Genome Analyzer and Mapping Station.

A group of small non-coding RNA molecules, microRNAs (miRNAs), is now recognised as an important player in regulation of gene expression. Detailed analysis of miRNAs is technically demanding because their mature form is only ~22 frequently highly homologous nucleotides long. With Matthias Hentze's group (page 54), GeneCore has worked on the development of the locked-nucleic-acids based oligonucleotide array miChip. This system, whose performance is further refined, enables profiling of mature miRNAs from human, mouse and other organisms. This valuable platform is complemented with a complete panel of qPCR TaqMan miRNA assays for human and mouse.

Also during 2008 the DNA microarray team dealt with increased number of samples submitted for genome-wide studies of DNA sequences bound by regulatory or modifying proteins, so called location analysis, on high-density tiling arrays (ChIP-on-chip).

Services provided

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

- project-focussed: microarrays (home-made, commercial), liquid handling robotics;
- sample-focussed: DNA sequencing, Bioanalyzer; miRNA qPCR profiling;
- access to instruments and complete support: qPCR, NanoDrop, PCR cyclers, microarray spotters & scanners, high-capacity vacuum concentrator

GeneCore offers sample processing on four different Affymetrix platforms. In addition to traditional '3'-end' arrays, exon arrays enable the study of alternative splicing patterns of individual transcripts at an unprecedented depth. In 2008 we implemented gene arrays, a new format launched by Affymetrix recently for analysis of whole-transcript expression; and tiling arrays are used for annotation independent transcriptome profiling. In 2008, GeneCore processed around 1,600 samples in total. Acquisition of a highly sensitive Agilent scanner further enhanced the facility's capacity to process spotted microarrays.

Three qPCR instruments we manage are primarily used for gene transcript quantification to corroborate microarray results and for detailed DNA occupancy profiling after chromatin immunoprecipitations. In 2008 we helped EMBL researchers analyse about 200,000 qPCR assayed points.

Selected references

Castoldi, M., Schmidt, S., Benes, V. *et al.* (2008). miChip: an array-based method for microRNA expression profiling using locked nucleic acid capture probes. *Nat. Protoc.*, 3, 321-329

Denger, S., Bahr-Ivacevic, T., Brand, H. *et al*. (2008). Transcriptome profiling of estrogen-regulated genes in human primary osteoblasts reveals an osteoblast-specific regulation of the insulin-like growth factor binding protein 4 gene. *Mol. Endocrinol.*, 22, 361-379

Kangaspeska, S., Stride, B., Metivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F. & Reid, G. (2008). Transient cyclical methylation of promoter DNA. *Nature*, 452, 112-115

Verga Falzacappa, M.V., Vujic Spasic, M., Kessler, R. *et al.* (2007). STAT3 mediates hepatic hepcidin expression and its inflammatory stimulation. *Blood,* 109, 353-8.

Protein Expression and Purification Core Facility

Hüseyin Besir

PhD 2001, Max Planck Institute of Biochemistry, Munich.

Postdoctoral research at Roche Diagnostics, Penzberg, and the Max Planck Institute of Biochemistry, Munich.

Facility Head at EMBL since 2006.

Our facility produces and purifies proteins from *E. coli*, insect and mammalian cells and sera using a variety of chromatographic methods. Following each purification, we can perform biophysical analyses to ensure the quality of the purified in terms of correct folding and stability. Our group also develops or evaluates new techniques or advanced protocols for protein production and purification. Most of our efforts are dedicated to develop time-saving solutions for these activities. 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 for general use which helps to considerably reduce the expenses of our users.

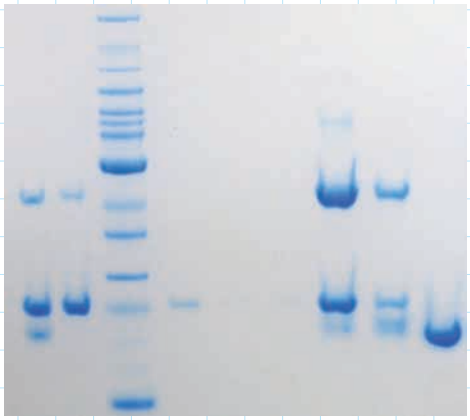
Major Projects and Accomplishments

We have evaluated new variants of our widely used pETM-series expression vectors for *E. coli* that can now be used for sequence and ligation-independent cloning (SLIC). We have adapted vectors for insect and mammalian cells for the same cloning protocol. Using a single PCR product with the gene of interest, you can integrate the insert into all of the vectors due to the universal overlaps that are present in the linearised vectors and the PCR product. A lethal gene insert in the original template vectors inhibits the growth of false positive colonies which reduces the number of clones to test for the correct insert. With this new vector set, one can test the expression of a gene in *E. coli*, insect (Baculovirus) and mammalian expression system in parallel and avoid redesigning of inserts for restriction-based cloning.

We have established new vectors for expression of fusion protein based on small proteins called Sumo1 and Sumo3 and their highly specific protease SenP2. In most of our expressions, SUMO-fusion proteins showed high expression yields. In cases of initially insoluble product, we could develop a protocol for proteolytic cleavage of the urea-denatured fusion protein with the robust protease under conditions where other proteases (e.g. TEV, 3C) show a poor performance. We could obtain pure, untagged proteins that were otherwise difficult to express or purify and could be used as antigens for immunisation, for example. We have constructed vectors with SUMO fusion tags for both SLIC and restriction enzyme-based cloning.

Services Provided

- Expression and purification of proteins in *E. coli*, insect and mammalian cells;
- Maintaining collections of expression vectors and bacterial strains;
- Producing frequently used enzymes (TEV protease, 3C protease, Taq, Pfu, T7 RNA pol., LIF, Cre) and protein molecular weight markers for general use;
- Developing and testing 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).



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

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.

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

Ventzki, R., Stegemann, J., Martinez, L. & de Marco, A. (2006). Automated protein analysis by online detection of laser-induced fluorescence in slab gels and 3D geometry gels. *Electrophoresis*, 27, 3338-3348

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

Schrodel, A. & de Marco, A. (2005). Characterization of the aggregates formed during recombinant protein expression in bacteria. *BMC Biochem.*, 6, 10

Jeroen Krijgsveld

PhD 1999, University of Amsterdam, The Netherlands.

Postdoc at Utrecht University, The Netherlands and Harvard Medical School, Boston, USA.

Assistant Professor, Utrecht University, The Netherlands

Team leader at EMBL since 2008.

Proteomic Core Facility

To address the needs of the post-genomic era, the Proteomic Core Facility offers a complete functional proteomic line, including gel- and chromatography-based protein and peptide identification. Using a variety of mass spectrometric approaches, we provide services for in-house and external visitors, both in the areas of proteomics and structural biology.

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.

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.

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.

Selected references

Rüggeberg, S., Horn, P., Li, X., Vajkocyz, P. & Franz, T. (2008). Detection of a γ-carboxy-glutamate as novel post translational modification of human transthyretin. *Protein & Peptide Letters,* 15, 43-46

Ventzki, R., Ruggeberg, S., Leicht, S., Franz, T. & Stegemann, J. (2007). Comparative 2-DE protein analysis in a 3D geometry gel. *Biotechniques*, 42, 271, 273, 275 passim

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

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 and grew out of EMBL's pioneering work in providing public biological databases to the research community. Its comprehensive range of data resources includes the European Nucleotide Archive (DNA and RNA sequences); Ensembl (genomes); ArrayExpress (transcriptomics and gene expression data); UniProt (protein sequences and functional information); PDBe (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. As the coordinator of ELIXIR, an EU-funded project to agree upon the future bioinformatics infrastructure for Europe, we are working with scientists and funders throughout Europe to pave the way towards a more stable footing for Europe's core data resources.

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. Eight research groups aim to understand biology through the development of new approaches to interpreting biological data. These diverse approaches 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; large-scale analysis of regulatory systems and differentiation and RNA genomics. In addition, our services teams perform extensive research to enhance existing data resources and develop new ones.

The EBI also provides user training, both on-site, through its hands-on training programme, and off-site through the Bioinformatics Roadshow. More information is available at www.ebi.ac.uk/training.

Almost all of our groups offer PhD places through the EMBL International PhD Programme. EMBL-EBI's PhD students register with, and obtain their doctorates from, the University of Cambridge. For a list of those groups with PhD places available for 2010/2011, 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
Director, EMBL-EBI

Janet Thornton

PhD 1973, King's College & National Inst. for Medical Research, London.

Postdoctoral research at the University of Oxford, NIMR & Birkbeck College, London.

Lecturer, Birkbeck College 1983-1989.

Professor of Biomolecular Structure, University College London (UCL) since 1990.

Bernal Professor at Birkbeck College, 1996-2002.

Director of the Centre for Structural Biology at Birkbeck College and UCL, 1998-2001.

Director of EMBL-EBI since 2001.

Computational biology of proteins – structure, function and evolution

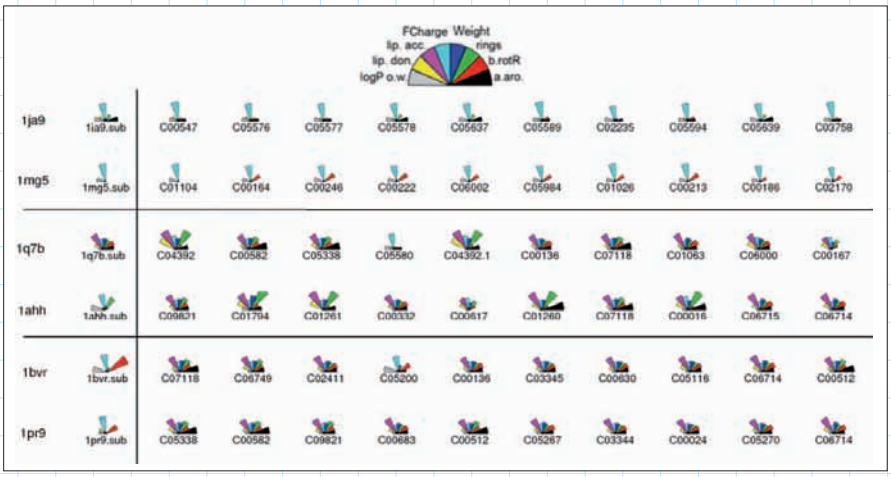
Previous and current research

The goal of our research is to understand more about how biology works at the molecular level, how enzymes perform catalysis, how these molecules recognise one another and their cognate ligands, and how proteins and organisms have evolved to create life. We develop and use novel computational methods to analyse the available data, 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 2008 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, their expression and control;
- 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. We will develop new computational tools to improve the handling of mechanisms and their reactions, which will allow improved chemistry queries across our databases. We are looking more closely at drug–protein interactions, membrane proteins (in collaboration with Professor David Jones at University College London) and allosteric effects. In the ageing project we are interested in tissue specificity and using human public transcriptome datasets to explore effects related to human variation and age.



We have used protein–ligand docking as a tool for protein function identification. The figure shows the physical chemical characterisation of the top ten hits from docking approximately 1,000 human metabolites to six members of the short chain dehydrogenase/reductase family of enzymes. The plots show eight 1D descriptors as colours, where the size of the sector reflects the value of the descriptor. These descriptors are: LogP, # H-bond donors; # H-bond acceptors, Molecular Weight, Charge, ~ rings, # rotatable bonds, ~ aromatic atoms. The first column shows plots for the 'known' cognate substrate for comparison. The plot highlights that in the first two rows all ten top hits are similar and resemble the substrate. The middle two examples show hits which are all different to each other and different from the substrate. However these two enzymes are known to be promiscuous. In the bottom two examples, all the hits look alike but are different from the known substrate. This is probably due to the inaccuracies of the scoring functions, and these results improve if the energy is recalculated with more sophisticated energy functions.

Selected references

Favia, A.D. *et al.* (2008). Molecular docking for substrate identification: The short-chain dehydrogenases/reductases. *J. Mol. Biol.*, 375, 855-874

Laskowski, R.A. & Thornton, J.M. (2008). Understanding the molecular machinery of genetics through 3D structures. *Nat. Rev. Genet.*, 9, 141-151

Najmanovich, R. *et al.* (2008). Detection of 3D atomic similarities and their use in the discrimination of small molecule protein-binding sites. *Bioinformatics*, 24, i105-111

Holliday, G.L. *et al.* (2007). The chemistry of protein catalysis. *J. Mol. Biol.*, 372, 1261-1277

Differentiation and development

Paul Bertone

PhD 2005, Yale University.

At EMBL-EBI since 2005.
Group leader since 2006.

Joint appointment with the Gene Expression and Developmental Biology Units.

Previous and current research

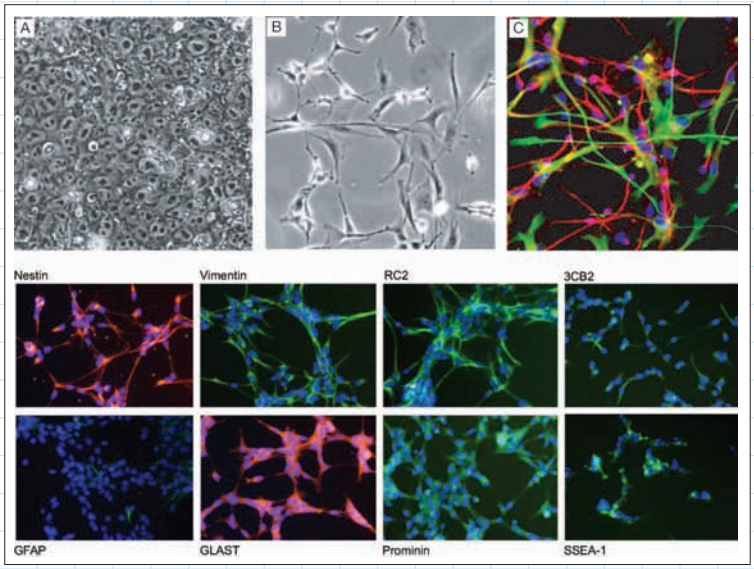
We investigate the cellular and molecular processes underlying mammalian stem cell differentiation, using a combination of experimental and computational approaches. Embryonic stem (ES) cells are similar to the transient population of self-renewing cells within the inner cell mass of the preimplantation blastocyst (epiblast), capable of pluripotential differentiation to all specialised cell types comprising the adult organism. These cells undergo continuous self-renewal to produce identical daughter cells, or can develop into specialised progenitors and terminally differentiated cells. Each regenerative or differentiative cell division involves a decision whereby an individual stem cell remains in self-renewal or commits to a particular lineage. Pluripotent ES cells can produce lineage-specific precursors and tissue-specific stem cells, 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 number 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 manifesting nuclear potency and self-renewal capacity of embryonic and tissue-specific stem cells. Despite these efforts, however, only a small number of components have been identified and large-scale characterisation of cellular commitment and terminal differentiation to specific cell types remains incomplete. Our research group applies the latest high-throughput technologies to investigate the functions of key regulatory proteins and their influence on the changing transcriptome. We focus on early lineage commitment of ES cells, neural differentiation and nuclear reprogramming. The generation of large-scale data from functional genomic and proteomic experiments will help to identify and characterise the 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 of vertebrate embryogenesis.

Future projects and goals

A long-term goal of this work is to elucidate accurate models 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 post-translational 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.



Top: Differentiation into neural stem (NS) cells from neural-rosette structures. A) ES cell primary culture, B, C) immunostaining for specific surface markers. Bottom: NS cells express markers characteristic of radial glia, permitting both accurate identification of differentiation stages and efficient FACS selection of homogeneous cell populations for genomic analysis. (Images: Steve Pollard, University of Cambridge; adapted from Conti et al.*, 2005).*

Selected references

Kind, J. *et al.* (2008). Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell*, 133, 813-828

Kirstetter, P. *et al.* (2008). Modeling of C/EBPα mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells. *Cancer Cell*, 13, 299-310

Anton Enright

PhD 2003, Cambridge University.

Postdoctoral work at Memorial Sloan-Kettering Cancer Center, New York.

Junior Faculty member at the Wellcome Trust Sanger Institute.

Group leader at EMBL-EBI since 2008.

Functional genomics and analysis of small RNA function

Previous and current research

Complete genome sequencing projects are generating enormous amounts of data. Although progress has been rapid, a large proportion of genes in any given genome are either un-annotated or possess a poorly characterised function. Our lab is entirely computational and our work involves the development of algorithms, protocols and datasets for functional genomics. Our goal is to predict and describe the functions of genes, proteins and in particular, regulatory RNAs and their interactions in living organisms.

Decoding microRNA function and regulation: The recent discovery of widespread translational regulation by microRNAs (miRNAs) highlights the enormous diversity and complexity of gene regulation in living systems and the need for computational techniques to help understand these systems. We developed the miRanda algorithm (www.microrna.org) for miRNA target detection in collaboration with the Computational Biology Center, at Memorial Sloan-Kettering Cancer Center in New York. Recently, we have predicted large-scale miRNA–target networks for mammalian, fish and insect genomes using the miRanda algorithm and cross-species sequence analysis as part of the miRBase database. The lab will continue to develop and improve methods for computational detection of miRNA target sites to investigate other possible aspects of miRNA target specificity, including sequence and structural motifs.

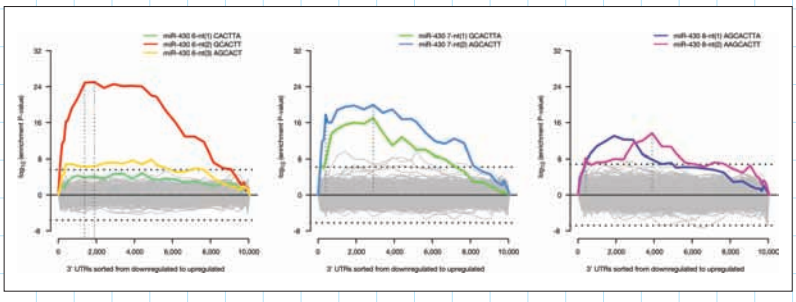
Analysing small RNA function: We work closely with experimental labs interested in the function of small RNAs in their system of interest by developing novel algorithms and techniques for analysis of primary data from such experiments (e.g. microarray). One example of this is the Sylamer algorithm (www.ebi.ac.uk/enright/sylamer) for associating miRNA or siRNA effects with gene expression data.

Studying regulatory RNAs in model systems: Through our experimental collaborations we work on understanding the role of RNA regulation in multiple diverse biological systems. These include: zebrafish development, mouse knock-out models, neuronal development, disease and cancer models and embryonic stem cells. Typically these experiments involve identification of miRNAs through profiling techniques followed by experimental perturbation of miRNAs of interest. High-throughput techniques such as microarrays and new technology sequencing are used to determine the effect of individual miRNAs in the system of interest.

Visualisation and network analysis: We retain an interest in both the analysis and clustering of biological networks (e.g. using Markov Clustering) and also in the visualisation of biological data. We continue to improve and maintain the BioLayout software for biological network visualisation and analysis.

Future projects and goals

We are interested in the evolution of regulatory RNAs and in developing phylogenetic techniques appropriate for short non-coding RNA. Our long-term goal is to combine regulatory RNA target prediction, secondary effects and upstream regulation into complex regulatory networks that may help us better understand the context of RNA in complex cellular networks.



Sylamer Results for the miR-430 microRNA in zebrafish. A clear signal is observed in gene-expression data for the seed-region of the miR-430 miRNA in wild-type versus mutant samples. The three panels show 6, 7 and 8nt motifs respectively.

Selected references

Van Dongen, S., Abreu-Goodger, C. & Enright, A.J. (2008). Detection of microRNA binding and siRNA off-targets from expression data. *Nature Methods*, 5, 1023-1025

Saini, H.K., Griffiths-Jones, S. & Enright, A.J. (2007). Genomic analysis of human microRNA transcripts. *Proc. Natl. Acad. Sci. USA*,104, 17719-17724

Giraldez, A.J. et al. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science*, 312, 75-79

Enright, A.J., Van Dongen, S. & Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.*, 30, 1575-1584

Enright, A.J. *et al.* (1999). Protein interaction maps for complete genomes based on gene fusion events. *Nature*, 402, 86-90

Evolutionary tools for sequence analysis

Nick Goldman

PhD 1992, University of Cambridge.

Postdoctoral work at National Institute for Medical Research, London, and University of Cambridge.

Wellcome Trust Senior Fellow 1995-2006.

Group leader at EMBL-EBI since 2002. Training Coordinator since 2004; Research and Training Coordinator since 2007.

Previous and current research

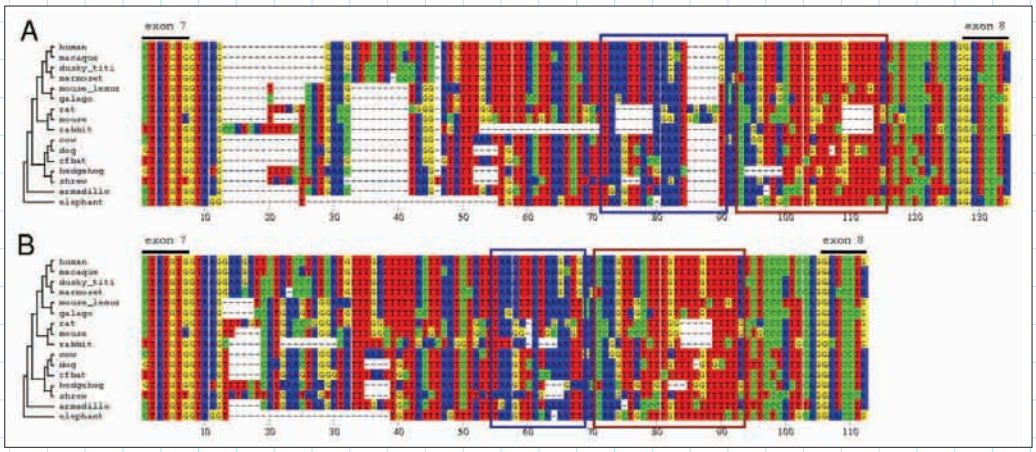
Research in the Goldman group concentrates on methods of data analysis that use evolutionary information in sequence data and phylogenies to infer the history of living organisms, to describe and understand processes of evolution, and to use this information to make predictions about the function of genomic sequence. One focus of the group is on comparative genomics and the bulk analysis of biological sequence data. Collaborations with major sequencing consortia remain fruitful, providing new data, challenges and a proving ground for new methods of sequence analysis. Intra-group collaborations between members developing theory and methods and those involved in the comparative analysis of genomic data remain a stimulating source of inspiration in all of our research areas.

The group has traditionally been strong in examining the theoretical foundations of phylogenetic reconstruction and analysis. In 2008 we have had a productive year, developing methods to infer and visualise evolutionary trees, and methods to use trees to improve sequence alignment. Our research aims to increase our understanding of the process of evolution and to provide new tools for biologists to elucidate the changing function of biological sequences.

Future projects and goals

The study of genome evolution continues to inspire us with novel problems in phylogenetic methodology. The complex nature of the non-independence of sequence data due to their evolutionary relatedness continues to generate statistically challenging problems, and the group is confident that we will continue to contribute to this theoretical field of research. We remain dedicated to retaining our interest in the practical applications of these methods in order to promote best practice in computational evolutionary and genomic biology, to keep in touch with the evolving needs of laboratory scientists and to continue to benefit from a supply of motivational biological questions where computational methods can help.

In 2009, next-generation sequencing technologies will produce almost as much sequence data as was produced in total prior to that date. For groups like ours that work on methods for sequence analysis, it is necessary to remember that the questions biologists want to ask of these data will change as more diverse experiments become practical. We have started to address this through our work on understanding the actual generation of data from next-generation sequencing platforms, and we will continue to work to allow the greatest benefit from modern molecular biology.



Screenshot of the PRANKSTER software for visualising multiple sequence alignments. The evolutionary tree relating the sequences, without which the evolutionary meaning of the alignment cannot be fully interpreted, is visible to the left, and various interesting insertion and deletion patterns are indicated (bottom).

Selected references

Jordan, G.E. & Piel, W.H. (2008). PhyloWidget: Web-based visualizations for the tree of life. *Bioinformatics,* 24, 1641-1642

Löytynoja, A. & Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science,* 320, 1632-1635

Taylor, M. S. *et al.* (2008). Rapidly evolving human promoter regions. *Nature Genetics,* 40, 1262-1263

Tress, M.L. *et al.* (2008). Determination and validation of principal gene products. *Bioinformatics,* 24, 11-17

Washietl, S., Machné, R. & Goldman, N. (2008). Evolutionary footprints of nucleosome positions in yeast. *Trends in Genetics,* 24, 583-587

Nicolas Le Novère

PhD 1998, Pasteur Institute, Paris.

Postdoctoral research at the University of Cambridge.

Research fellow, CNRS, Paris.

Group leader at EMBL-EBI since 2003.

Computational systems neurobiology

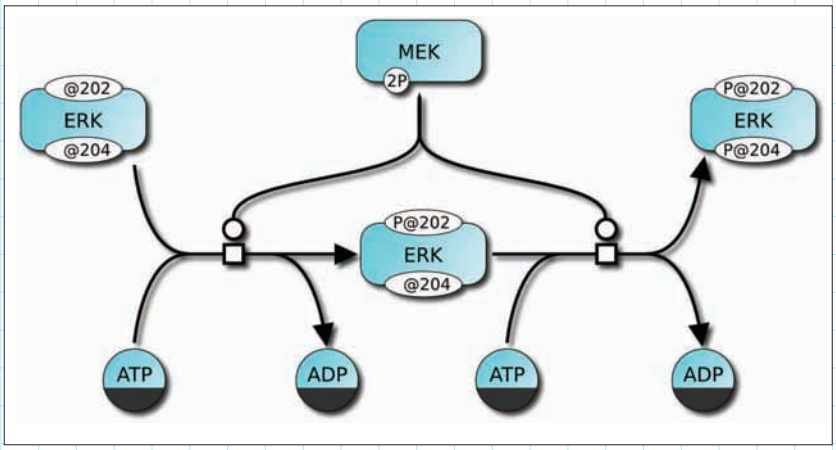
Previous and current research

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 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 are particularly interested in understanding the processes of cooperativity, pathway switch and bistability.

The group provides community services that facilitate research in computational systems biology. In particular, we are leading the efforts in encoding and annotating kinetic models in chemistry and cellular biology, including the creation of standard representations, the production of databases and software development. The Systems Biology Markup Language (SBML) is designed to facilitate the exchange of biological models between different types of software. As editors of the language, we also work on increasing its coverage, and are developing software to support SBML usage. The Systems Biology Graphical Notation is an effort to develop a common visual notation for biochemists and modellers. Moving from the form to the content, we are also developing standards for model curation (MIRIAM, MIASE), and controlled vocabularies to improve the models (the Systems Biology Ontology, the TErminology for the Description of DYnamics etc.). Finally, a model is only useful if it can be easily accessed and reused. BioModels Database (www.ebi.ac.uk/biomodels) is now the reference resource where scientists can store, search and retrieve published mathematical models of biological interest.

Future projects and goals

In forthcoming years, the activity of the group will continue along 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, TrkB, PI3K) and tackle problems such as the role of scaffolding proteins or the synchronisation of calcium waves and phosphorylation gradients. Building on the growth of the 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 the development of standards and resources for systems biology will continue, with the goal of completing the puzzle of representations and ontologies so as to efficiently integrate the different levels of description of biochemical and cellular processes, qualitative, quantitative and experimental.



An example of SBGN in action, showing the catalysis of ERK phosphorylation by MEK in the growth factor signalling pathway.

Selected references

Le Novère, N. *et al.* (2008). DARPP-32: Molecular integration of phosphorylation potential. *Cell. Mol. Life Sci.,* 65, 2125-2127

Stefan, M.I. *et al.* (2008). An allosteric model of calmodulin explains differential activation of PP2B and CaMKII. *Proc. Natl. Acad. Sci. USA,* 105, 10768-10773

Taylor, C.F. *et al.* (2008). Promoting coherent minimum reporting guidelines for biological and biomedical investigations: The MIBBI project. *Nat. Biotechnol.,* 26, 889-896

Laibe, C. & Le Novère, N. (2007). MIRIAM Resources: Tools to generate and resolve robust cross-references in Systems Biology. *BMC Syst. Biol.,* 1, Article 58

Genome-scale analysis of regulatory systems

Nicholas Luscombe

PhD 2000, University College London.

Postdoctoral work at Department of Molecular Biophysics & Biochemistry, Yale University.

Group leader at EMBL-EBI since 2005.

Joint appointment with the Gene Expression Unit.

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 and genetic investigations. In the past decade, the availability of genome sequences and development of new laboratory techniques has generated (and continues 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 poised to achieve this.

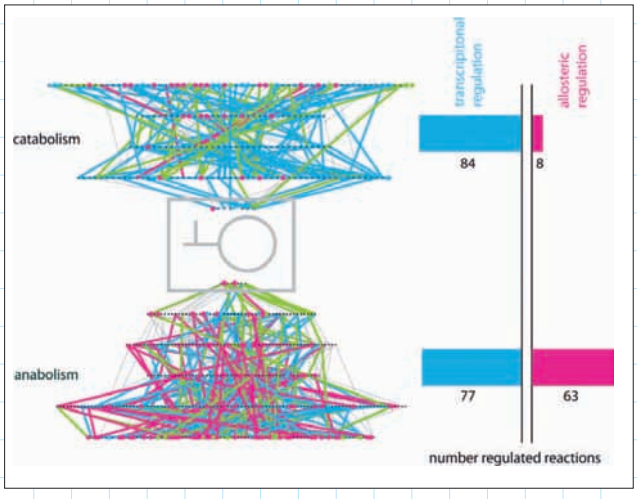
Much of our work so far has focussed on the regulatory system in the yeast *Saccharomyces cerevisiae*. By integrating diverse data sources – from genome sequence to the results of functional genomics experiments – we can study the regulatory system at a whole-organism level. We have also expanded our interests to understanding regulation in enterobacteria and humans.

Our current projects include:

- examining how the metabolic system is controlled at multiple levels through the feedback activity of small molecules;
- analysing the repertoire, usage and cross-species conservation of transcription factors in the human genome;
- wet/dry collaborations to uncover the regulation governing complex organismal behaviour;
- wet/dry collaborations to understand the epigenetic control of dosage compensation in animals.

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. A major focus continues to be our close interactions with research groups performing genome-scale experiments.



A network representation displays the E. coli *metabolic system. Nodes represent small molecules and edges depict enzymatic reactions. The reactions are coloured according to whether they are controlled transcriptionally (blue), allosterically (cyan) or by both methods (green). Allosteric feedback predominantly regulates anabolic pathways, whereas transcriptional feedback controls both anabolic and catabolic pathways.*

Selected references

Vaquerizas, J.M. *et al.* (2009). A census of transcription factors in the human genome: function, expression and evolution. *Nat. Rev. Genet.*, 10, 252-63

Hancock, V. *et al.* (2008). Transcriptomics and adaptive genomics of the asymptomatic bacteriuria *Escherichia coli* strain 83972. *Mol. Genet. Genomics*, 1-12

Kind, J. *et al.* (2008). Genome-wide analysis reveals MOF as a key regulator of dosage compensation and gene expression in *Drosophila*. *Cell*, 133, 813-828

Seshasayee, A.S. *et al.* (2008). Principles of transcriptional regulation and evolution of the metabolic system in *E. coli*. *Genome Res.*, 19, 79-91.

Dietrich Rebholz-Schuhmann

Master in Medicine, 1988, University of Düsseldorf.

PhD 1989, University of Düsseldorf.

Master in Computer Science, 1993, Passau.

Senior scientist at gsf, Munich, Germany and LION bioscience AG, Heidelberg.

Group leader at EMBL-EBI since 2003.

Facts from the literature and biomedical semantics

Previous and current research

Text mining comprises the fast retrieval of relevant documents from the whole body of the literature (e.g. Medline database) and the extraction of facts from the text thereafter. Text mining solutions are now becoming mature enough to be automatically integrated into workflows for research work.

Research in the Rebholz-Schuhmann group is focussed on fact extraction from the literature. It is our goal to automatically connect literature content to other biomedical data resources (e.g. bioinformatics databases) and to evaluate the results. Ongoing research targets the identification of relationships between genes and diseases, molecular interactions and other types of information. Over the past two years, the team has generated several public resources: a lexicon of biomedical terms, an ontology for gene regulatory events and recently an authoring service (PaperMaker).

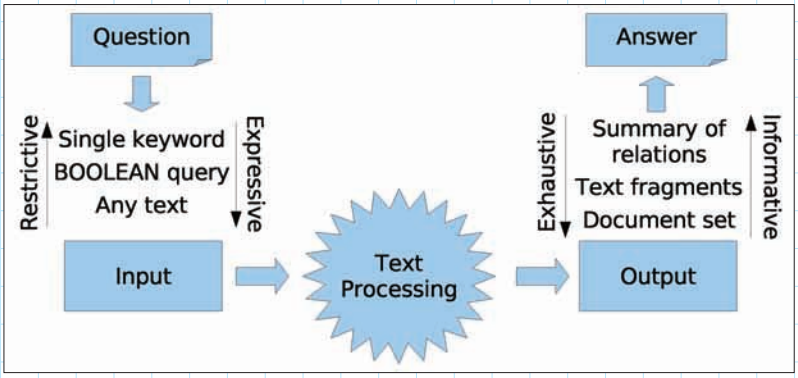
The work in the research group is split into different parts: 1) research work in named entity recognition and its quality control (e.g. UKPMC project, CALBC); 2) knowledge discovery tasks, e.g. for the identification of gene–disease associations; and 3) development of a modular IT infrastructure for information extraction (Whatizit). All parts are tightly coupled.

Future projects and goals

The following goals are priorities for the future. Firstly we will continue our ongoing research in term recognition and mapping to biomedical data resources to establish state-of-the-art text mining applications. We will develop this by focussing on automatic means to measure and evaluate existing options to identify the most promising solutions (UKPMC project, CALBC support action).

Secondly, we will invest further effort into the extraction of content from the scientific literature. Such solutions will be geared towards the annotation of diseases and the generation of fact databases. As part of this research we will investigate workflow systems where text mining supports bioinformatics information retrieval solutions. One solution is the integration of public biomedical data resources 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.



Overview of the categorisation of information retrieval tools on the basis of their input and output formats.

Selected references

Beisswanger, E. *et al.* (2008). Gene Regulation Ontology (Gro): Design principles and use cases. Paper presented at Studies in Health Technology and Informatics 2008, 136, 9-14

Jaeger, S. *et al.* (2008). Integrating protein-protein interactions and text mining for protein function prediction. *BMC Bioinformatics,* 9, Article S2

Kim, J.J. *et al.* (2008). MedEvi: Retrieving textual evidence of relations between biomedical concepts from Medline. *Bioinformatics,* 24, 1410-1412

Kim, J.J. & Rebholz-Schuhmann, D. (2008). Categorization of services for seeking information in biomedical literature: A typology for improvement of practice. *Brief. Bioinform.,* 9, 452-465

Rebholz-Schuhmann, D. *et al.* (2008). Text processing through web services: Calling Whatizit. *Bioinformatics*, 24, 296-298

PANDA proteins and the Apweiler research group

Rolf Apweiler

PhD 1994, University of Heidelberg. Germany.

Team leader at EMBL-EBI since 1997.

Previous and current research

The PANDA (Protein and Nucleotide Data) group was created in June 2007 by merging the former Ensembl (Birney) and Sequence Database (Apweiler) groups.

The activities of the PANDA group are focussed on the production of protein sequence, protein family and nucleotide sequence databases at EMBL-EBI. We maintain and host the EMBL Nucleotide Sequence Database, the Ensembl genome browser, the UniProt protein resource, and a range of other biomolecular databases. These efforts can be divided into three major groups: nucleotides, proteins, and chemoinformatics and metabolism. In addition to PANDA activities, the Apweiler group has a complementary research component.

The activities of the PANDA proteins teams are centred on the mission of providing public access to all known protein sequences and functional information about these proteins. The UniProt resource provides the centrepiece for these activities. Most of the UniProt sequence data is derived from translation of nucleotide sequences provided by the European Nucleotide Archive and Ensembl. All UniProt data undergoes classification provided by InterPro (see the report from Sarah Hunter, page 78). In addition, we add information extracted from the scientific literature and curator-evaluated computational analysis whenever possible. The combined InterPro literature annotation forms the basis for automatic annotation approaches to annotate all the sequence data without experimental functional data. Protein interaction and identification data is or will be provided to UniProt by the IntAct protein–protein interaction database and by the Protein Identification (PRIDE) database.

Ongoing research activities in the group include the development of methods to improve searching of large biological datasets, approaches to improve protein identification from mass spectrometry data, algorithms for genome-wide sequence comparison and the development of tools for the automatic annotation of proteins.

Future projects and goals

It is our intention to work on improved integration and synchronisation of all PANDA resources. 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 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 new methodologies to the annotation of UniProtKB/TrEMBL records.

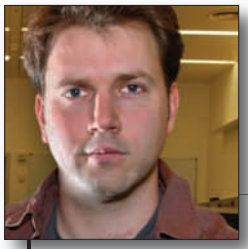
Selected references

Klie, S. *et al.* (2008). Analyzing large-scale proteomics projects with latent semantic indexing. *J. Proteome Res.*, 7, 182-191

Mueller, M. *et al.* (2008). Analysis of the experimental detection of central nervous system-related genes in human brain and cerebrospinal fluid datasets. *Proteomics*, 8, 1138-1148

The UniProt Consortium (2008). The Universal Protein Resource (UniProt). *Nucleic Acids Res.*, 36, D190-195

Mueller, M. *et al.* (2007). Annotating the human proteome: Beyond establishing a parts list. *Biochimica et Biophysica Acta*, 1774, 175-191



Ewan Birney

PhD 2000, Sanger Institute, Hinxton, Cambridge.

Team leader at EMBL-EBI since 2000.

PANDA nucleotides and methods for genome analysis

Previous and current research

Ewan Birney is joint head of the PANDA team, with Rolf Apweiler, and has a strategic oversight of the major DNA projects: Ensembl, Ensembl Genomes and the European Nucleotide Archive (ENA). These are large projects all dealing with DNA sequence information in a variety of forms, in particular in the annotation and interpretation of genomes. DNA sequence remains at the heart of molecular biology and hence bioinformatics and its use has grown significantly with the recent advent of ultra-high throughput DNA sequencing machines. In 2008 we have seen a striking growth in two areas – the use of these new machines for surveying natural variation in populations, in particular the human population and the more routine determination of genotypes from large disease cohorts, leading to associations between genetics and disease. The shift in technology and the repositioning of genomic information as a key organisation principal has meant that there have been significant changes to the way our DNA archival services operate and more focus on coordinating with genomic resources.

In addition, the Birney research group focusses on DNA sequence interpretation. There are two major themes to this research. The first is algorithm development. There have been a number of algorithmic developments in the Birney group, in particular on sequence alignment methods (Slater & Birney, *BMC Bioinformatics*), multiple alignments (Paten *et al.*, *Genome Research*) and on *de novo* assembly using short reads (Zerbino & Birney, *Genome Research*). The second is on data-driven discovery of important genomic features in the genome. This includes large projects, such as the ENCODE project (The ENCODE Consortium, *Nature*), which involves a large number of experimental groups focussing on the interpretation of genomic information, particular from non-coding DNA sequence. Integration across different data types provides new insights, for example, the surprising lack of correlation of conservation with experimentally-assayed function. There are also more specific, focussed projects, such as the exploration of cis-regulation in vertebrates (Ettwiller *et al.*, *Genome Biology*) in which specific new data discovery technique are developed to elucidate genomic function.

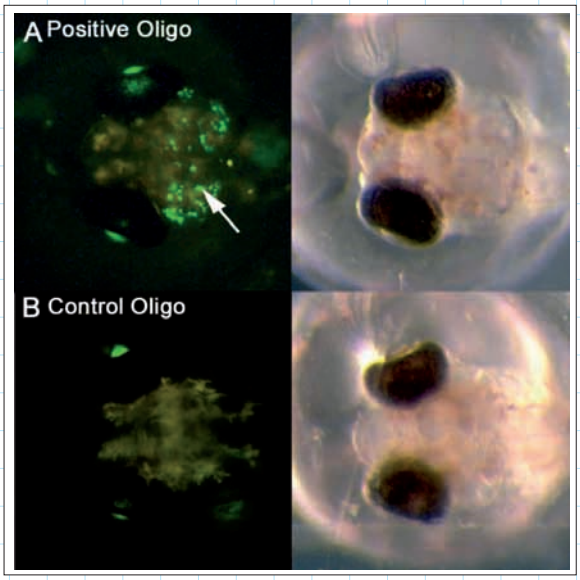


Figure showing the expression of synthetic enhancers designed using algorithms from the cis-regulatory research performed in the group. The arrows show tissue specific expression in medaka fish embryos from these ab initio *designed enhancers.*

Future projects and goals

Future research continues both of these themes – algorithm development and data-driven discovery, both relating to genomic DNA sequence, but will also add the use of intra-species variation (i.e. natural variation in a population) with molecular markers as a component. Leveraging the natural polymorphisms in different populations allows us to understand how molecular function varies between individuals, and how this variation is correlated to the genotype of each individual. In the context of the human genome, very often this is done in the context of specific diseases, so one has genotype, functional information and disease status. In other organisms (for example, rodents), one has more controlled phenotype measurement at the organism level, allowing more complex scenarios to be explored.

Selected references

Ettwiller, L. *et al.* (2008). Analysis of mammalian gene batteries reveals both stable ancestral cores and highly dynamic regulatory sequences. *Genome Biol.*, 9, R172

Paten, B. *et al.* (2008). Enredo and Pecan: genome-wide mammalian consistency-based multiple alignment with paralogs. *Genome Res.*,18, 1814-1828

Zerbino, D.R. & Birney, E. (2008). Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res.*, 18, 821-829

The ENCODE Consortium (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447, 799-816

Slater, G.S. & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, 6, 31

The Microarray Informatics Team

Alvis Brazma

PhD 1987, Computer Science, Moscow State University.

Postdoctoral research in New Mexico State University, La Cruses.

Team leader at EMBL-EBI since 2000.

Previous and current research

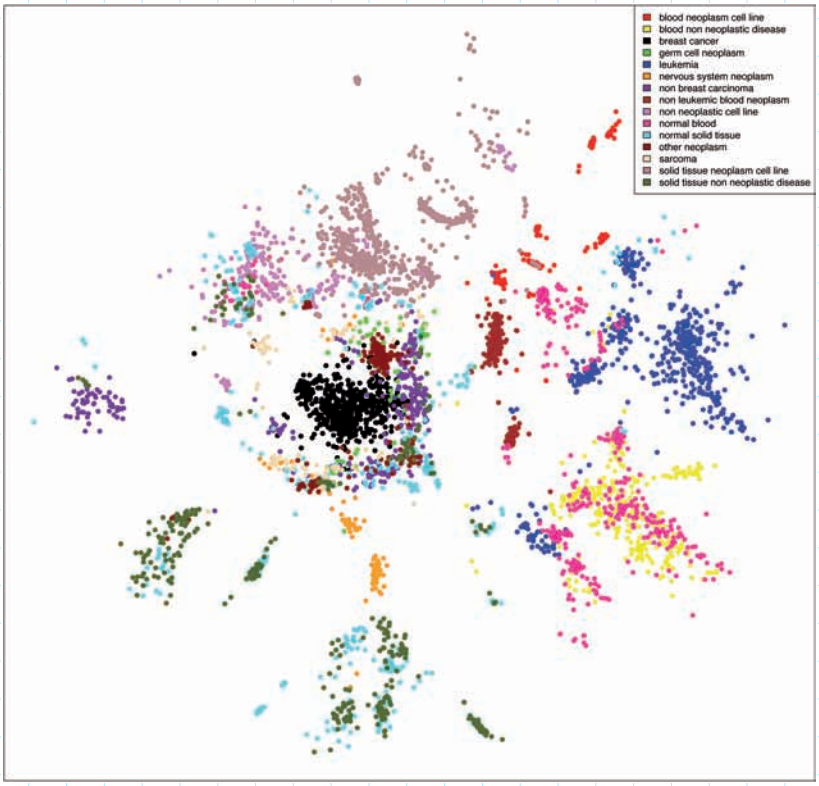
The Microarray Informatics team is working in four main directions:

- development of the ArrayExpress Archive and Atlas of Gene Expression;
- high-throughput data integration and analysis;
- development of algorithms for systems biology;
- biomedical informatics-related research and development projects.

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.*, 2000, Brazma *et al.*, 2001) and began work to establish the ArrayExpress database. As of February 2009 the ArrayExpress Archive holds data from approximately 200,000 microarrays. The ArrayExpress Atlas of Gene Expression allows the users to query for expression profiles of particular genes, tissues or disease states across multiple experiments. Our PhD students and postdocs focus mostly on integrative data analysis and on building models for systems biology (e.g., Rustici *et al.*, 2004, Schlitt & Brazma, 2006).

Future projects and goals

A biological system, such as a cell, tissue, organ or organism, can be in many different states, such as developmental stages, disease states, or physiological states. Different cell types can be considered as different biological states evolving from the progenitor cell state. This poses many questions; how many different biological states are there, what are the relationships between them, which tissue or cell types are more similar to each other and which are different, how is the biological state affected by a disease, how much does gene expression depends on environment, and how much on genotype? Finding answers to these questions is one of the most important goals of our group's research. Towards this goal we are building a comprehensive gene expression atlas for human and model organisms. The Gene Expression Atlas integrates data from tens of thousands of transcriptomics assays available in ArrayExpress. We will also continue large collaborative projects, such as integration of transcriptomics, proteomics and human genome variation data to understand the molecular mechanisms of disease, as well as building biomedical data analysis infrastructure to help us in answering these questions.



Visualisation of relationship transcriptomes of ~5,300 human samples categorised in 15 biological classes using Neighbor Retrieval Visualizer (NeRV; Venna & Kaski, 2007) developed by our collaborators in Helsinki University of Technology.

Selected references

Parkinson, H. *et al.* (2009). ArrayExpress update – from an archive of functional genomics experiments to the atlas of gene expression. *Nucleic Acids Res.,* 37, D868-872

Rustici, G. *et al.* (2007). Global transcriptional responses of fission and budding yeast to changes in copper and iron levels: a comparative study. *Genome Biol.,* 8, R73

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

Rustici, G. *et al.* (2004). Periodic gene expression program of the fission yeast cell cycle. *Nature Genetics,* 36, 809-817

Paul Flicek

DSc 2004, Washington University, St. Louis, Missouri.

At EMBL-EBI since 2005.

Team leader at EMBL-EBI since 2008.

Vertebrate genomics

Previous and current research

The Vertebrate Genomics team is a combined service and research group that creates and manages data resources focussing on genome annotation and human variation. The major service projects of the Vertebrate Genomics team are Ensembl, the European Genotype Archive, and the Data Coordination Centre for the 1000 Genomes Project. In support of these projects, we are developing the specialised, large-scale bioinformatics infrastructure required for each analysis. The team's research is on computational genome annotation with a particular focus on the integration of diverse data types such as extensive comparative sequencing, DNA–protein interactions, epigenetic modifications, and the DNA sequence itself.

Ensembl (www.ensembl.org) is a comprehensive genome information system featuring an integrated set of tools for genome annotation, data mining and visualisation of chordate genomes. As such, it is one of the fundamental database resources used to address questions in medical research and molecular biology. As of August 2008, there were 39 fully-supported genomes in Ensembl including human, mouse, chicken, five species of fish, a nematode, and several other mammalian, chordate and insect species.

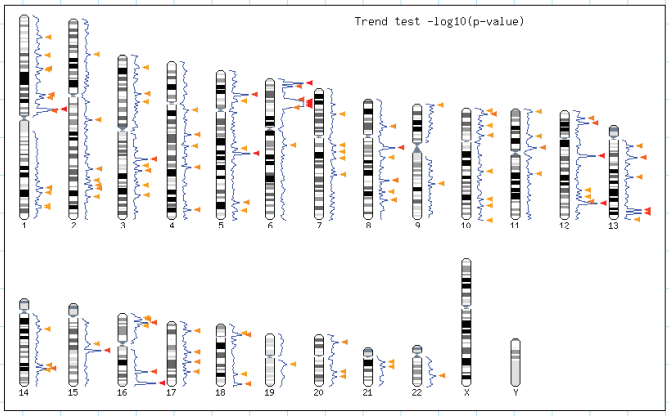
The European Genotype Archive (EGA) database provides a permanent archive for all types of personally identifiable genetic data including genotypes, genome sequence and associated phenotype data. The EGA contains both data collected from individuals whose consent agreements stipulate data release for specific approved research uses or bona fide researchers, as well as data approved for full public release.

The 1000 Genomes Project (www.1000genomes.org) aims to create a comprehensive and public catalogue of common human genetic variation in three populations by using next-generation sequencing technology. During 2008, the project conducted three pilot projects to assess the feasibly of creating a deep and accurate catalogue and develop the necessary tools to manage and analyse the data. The pilot projects included the sequencing of 180 individuals to 2x coverage; sequencing two trios consisting of a child and both parents to 20x coverage; and targeted sequencing of 1,000 genes in 1,000 individuals.

In collaboration with the NCBI, the Vertebrate Genomics team is one half of the 1000 Genomes Project Data Coordination Centre (DCC) and has co-leadership of the project's data flow group. Over the course of the year the project produced approximately 2 terabases of sequence (equivalent to 8.5 times the number of nucleotides in the EMBL-Bank sequence archive) at a rate approaching 30 gigabases per day. This data is collected by the DCC and made available to the 1000 Genome Project analysis group and interested researchers worldwide.

Future projects and goals

Next-generation sequencing methods are having a profound impact. For example, we have been investigating ways to use short read transcriptome data in our automatic annotation to support the substantial amounts of data we expect in the future. The availability of an increasing number of genome sequences is challenging the comparative genomics aspects of the team's work both in terms of scale and complexity. ENCODE and the 1000 Genomes Project will respectively provide significant new data into the functional genomics and variation resources. Future developments for the EGA include a suite of customised data mining tools, an analysis pipeline infrastructure supporting uniform analysis of the data in the archive, and the development (in collaboration with international partners) of standards for the exchange of genotype data including whole genome sequences.



An example GenomeView from the European Genotype Archive showing genomic regions that are significantly associated with type I diabetes.

Selected references

Flicek, P. *et al.* (2008). Ensembl 2008. *Nucleic Acids Res.*, 36 (Database issue): D707-D714

Johnson, D.S. *et al.* (2008). Systematic evaluation of variability in ChIP-chip experiments using predefined DNA targets. *Genome Res.*, 18, 393-403

Saar, K. *et al.* (2008). SNP and haplotype mapping for genetic analysis in the rat. *Nat. Genet.*, 40, 560-566

Warren, W.C. *et al.* (2008). Genome analysis of the platypus reveals unique signatures of evolution. *Nature*, 453, 175-183

Flicek, P. (2007). Gene prediction: compare and CONTRAST. *Genome Biol.*, 8, 233

The GO Editorial Office

Midori Harris

PhD 1997, Cornell University, Ithaca, NY.

Scientific Curator, Saccharomyces Genome Database, Stanford University, Stanford, CA.

GO Editor at EMBL-EBI since 2001.

Previous and current research

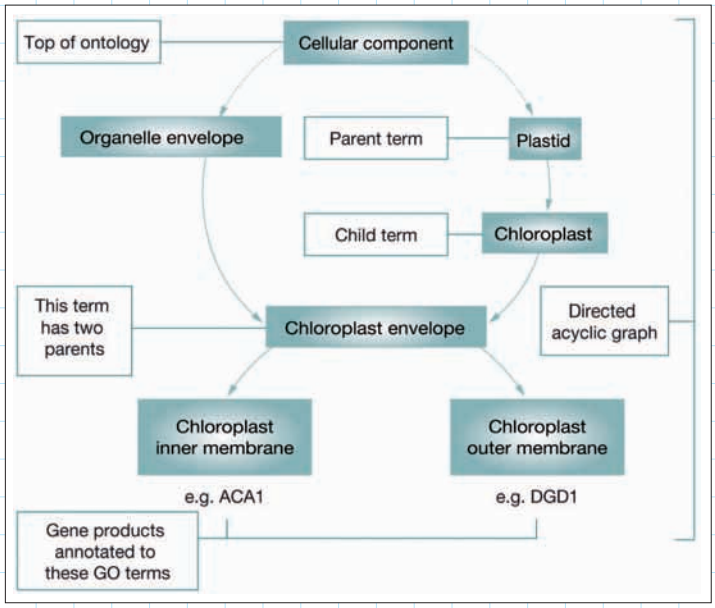
The Gene Ontology (GO) project (www.geneontology.org) is a collaborative effort to construct and use ontologies to facilitate the biologically meaningful annotation of genes and their products in a wide variety of organisms. At the EBI, the GO Editorial Office plays a key role in managing the distributed 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. The GO ontologies cover three key biological domains that are shared by all organisms:

- **molecular function** defines the tasks performed by individual gene products; examples include aminoacyl-tRNA ligase activity and translation elongation factor activity;
- **biological process** defines broad biological goals, such as signal transduction or ribosome assembly, that are accomplished by ordered assemblies of molecular functions;
- **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). The GO vocabularies have long defined 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). New relationships representing biological regulation have recently been added, as described below. The figure shows a portion of the GO cellular component DAG.



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 envelope. Any gene that is annotated to this term is automatically annotated to both chloroplast and envelope. Some terms and relationships have been omitted for clarity.

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. Improvements begun in 2008 on signal transduction, transcription, and other topics will therefore continue in 2009. The new regulates relationships will be used to create the first links between the biological process and molecular function ontologies, with additional types of links to follow. Work on recasting many complex process terms as explicit cross-products with orthogonal ontologies such as the ChEBI ontology and the cell ontology will also continue.

Selected references

The Gene Ontology Consortium (2008). The Gene Ontology (GO) Project in 2008. *Nucleic Acids Res.*, 36, D440-D444

Eilbeck, K. *et al.* (2005). The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.*, 6, R44

The Gene Ontology Consortium (2001). Creating the Gene Ontology resource: design and implementation. *Genome Res.*, 11, 1425-1433

Kim Henrick

PhD 1974, University of Western Australia, Perth.

Postdoctoral research at the Polytechnic of North London, Imperial College of Science and Technology, SERC Daresbury Laboratory, Cambridge Centre for Protein Engineering and National Institute for Medical Research.

Team leader at EMBL-EBI since 2001.

Macromolecular Structure Database Team

Previous and current research

The Macromolecular Structure Database (MSD; www.ebi.ac.uk/msd) holds detailed knowledge of protein structure and function of biological macromolecules. Access to this information is vital for many different users, for example, the identification of potential targets for therapeutic intervention as well as of lead structures for pharmaceutical use. We uniquely integrate the experimental data derived by 3D cryo-electron microscopy and electron tomography techniques, and derive the molecular biological assemblies of structures held in the PDB. Our aim is to provide integrated data resources that evolve with the needs of structural biologists; for all biologists seeking to understand the structural basis of life, for researchers looking for the causative agents of disease and diagnostic tools and for the pharmaceutical and biotech industries.

Through our membership of the Worldwide Protein Data Bank organisation (wwPDB; www.wwpdb.org) we are an equal PDB partner and work closely with the United States (RCSB) and Japanese (PDBj) partners to provide the single international archive for structural data. To reflect this status, we are in the process of migrating to a new name, the Protein Data Bank in Europe (PDBe), as recommended by the MSD Scientific Advisory Board.

The PDB archive is growing annually in value, importance and size. The current conservative estimate of the value of the entire PDB archive is in excess of $4.5 billion (just over €3 billion), with the annual research cost involved in elucidating these structures estimated to be $750 million (€530 million). More than 15 different funding agencies worldwide fund the wwPDB centres, generating an estimated total of $9 million (€6.4 million) per year. When these figures are compared, the costs of maintaining the archive represents approximately 1% of the annual research investment. The combined MSD services (providing access to 20 different views of PDB data) deliver 3.6 million hits per month, transferring about 238GB to approximately 53,000 users. In addition, users download around 1.7 million and 4 million files per month from the MSD and PDB FTP sites respectively. This year saw the PDB make the deposition of scientific evidence for a macromolecular structure mandatory (www.wwpdb.org/news.html). From 1 February 2008, structure factor amplitudes/intensities (for crystal structures) and restraints (for NMR structures) are now mandatory for PDB deposition. In accordance with this policy, the PDB ID should be included in publications and authors must agree to release the atomic coordinates and associated experimental data when the article is published.

Future projects and goals

To continue to meet the goals of the PDB as a critical global scientific resource we must evolve with and anticipate the needs of the scientific community. This includes the five- to ten-year vision to enrich the annotation of entries to include biological function as well as all aspects of the PDB services. We aim to maximise the efficiency and effectiveness of macromolecular structure data handling, and support the scientific community through the development and adoption of common deposition and annotation processes and tools. These shared processes and tools will enable balancing of deposition load across the wwPDB member sites and allow for shared maintenance and updates to the archive and tools in the future. The next deposition and annotation product will be a unified system, which will be used at all deposition sites. It is designed to provide significant reduction in curation time, assure consistency and will be extensible to accommodate hybrids and new structural determination methods. It will also provide users with portable validation tools and 'table 1' of their experiment.

From July 2009, the MSD/PDBe team will be led by Gerard Kleywegt.

Selected references

Golovin, A. & Henrick, K. (2008). MSDmotif: Exploring protein sites and motifs. *BMC Bioinformatics*, 9, Article 312

Haquin, S. *et al.* (2008) Data management in structural genomics: an overview. *Methods Mol Biol.*, 426, 49-79

Henrick, K. *et al.* (2008) Remediation of the protein data bank archive. *Nucleic Acids Res.*, 36, D426-D433

Lawson, C.L. *et al.* (2008). Representation of viruses in the remediated PDB archive. *Acta Crystallogr. Sect. D-Biol. Crystallogr.*, 64, 874-882

Markley, J.L. *et al.* (2008). BioMagResBank (BMRB) as a partner in the Worldwide Protein Data Bank (wwPDB): new policies affecting biomolecular NMR depositions. *J. Biomol. NMR*, 40, 153-155

The Proteomics Services Team

Henning Hermjakob

Dipl. Inf (MSc) 1996, Bioinformatics, University of Bielefeld.

Research assistant at the National Research Centre for Biotechnology (GBF), Braunschweig, Germany, in the Transfac Database team.

Team leader at EMBL-EBI since 2005.

Previous and current research

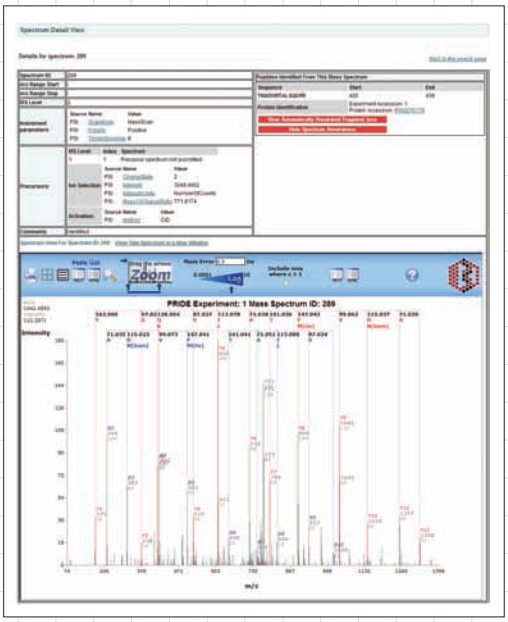
The Proteomics Services team develops tools and resources for the representation, deposition, distribution and analysis of proteomics and proteomics-related data. The team is a major contributor to the Proteomics Standards Initiative (PSI; www.psidev.info) of the international Human Proteome Organization (HUPO). We provide reference implementations for the PSI community standards, in particular the PRIDE protein identification database (www.ebi.ac.uk/pride) and the IntAct molecular interaction database (www.ebi.ac.uk/intact). On the next level of abstraction, we provide the Reactome database of pathways (www.reactome.org) in collaboration with Cold Spring Harbor Laboratory, New York.

As a result of long-term engagement with the proteomics community, journal editors and funding organisations, proteomics data deposition in PSI-compliant data resources such as IntAct and PRIDE is increasingly becoming a strongly recommended part of the publishing process. Accordingly, this has resulted in a rapid increase in the data content of our resources.

The Proteomics curation teams ensure consistency and appropriate annotation of all data; whether from direct depositions or literature curation, to provide the community with high-quality reference datasets.

Across a range of European projects (Apo-Sys, BioSapiens, FELICS, ENFIN, ProteomeBinders and Transfog) we contribute to the development of data integration technologies using the Distributed Annotation System (DAS) and web services. In particular, the successful Ontology Lookup Service (OLS; www.ebi.ac.uk/ols), Protein Identifier Cross-Reference Service (PICR; www.ebi.ac.uk/Tools/picr) and the DASTY DAS client (www.ebi.ac.uk/dasty) are under constant evolution and development.

The Proteomics Services team follows an open source, open data approach; all resources we develop are freely available.



PRIDE detail view for a submitted spectrum, including annotation of submitted fragment ions in the spectrum.

Future project and goals

In 2007, our molecular interaction activities resulted in a substantial set of published manuscripts, from the MIMIx guidelines via the PSI MI 2.5 format to the standard implementation in the IntAct database. In 2008, a similar breakthrough has been achieved in the domain of protein identifications, with three published MIAPE modules and the release of the mzML format for mass spectrometry data representation. For 2009, we plan to build on these standards and their implementation in PRIDE and IntAct, and initiate a regular exchange of proteomics data with international collaborators in the ProteomExchange and IMEx consortia (http://imex.sf.net).

We also plan to intensify data integration within and beyond the projects of the Proteomics Services team, in particular in the context of the EnVision platform. We also hope to achieve this in the context of closer integration between IntAct and Reactome, supplementing canonical pathway information with relevant molecular interaction data.

Finally, we will continue our successful collaboration with all PSI partners, in particular with journals and editors, to encourage data producers to make their data available to the community through public databases by utilising community-supported standards.

Selected references

Klie, S. *et al.* (2008). Analyzing large-scale proteomics projects with latent semantic indexing. *Journal of Proteome Research*, 7, 182-191

Martens, L., Palazzi, L.M. & Hermjakob, H. (2008). Data standards and controlled vocabularies for proteomics. *Methods Mol. Biol.*, 484, 279-286

Michaut, M. *et al.* (2008). InteroPORC: Automated inference of highly conserved protein interaction networks. *Bioinformatics*, 24, 1625-1631

Montecchi-Palazzi, L. *et al.* (2008). The PSI-MOD community standard for representation of protein modification data. *Nat. Biotechnol.*, 26, 864-866

Mueller, M. *et al.* (2008). Analysis of the experimental detection of central nervous system-related genes in human brain and cerebrospinal fluid datasets. *Proteomics*, 8, 1138-1148

Sarah Hunter

Bioinformatics MSc., 1998, University of Manchester.

Bioinformatics Database Administrator at Pharmacia & Upjohn, Stockholm.

Bioinformatics Group Coordinator at Biovitrum, Stockholm.

Team leader at EMBL-EBI since 2007.

The InterPro Team

Previous and current research

The InterPro team coordinates the InterPro and CluSTr projects and develops the software used by the Gene Ontology Annotation (GOA) group 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 UniProt Archive (UniParc) proteins using the InterProScan software, and displays the matches to the UniProt KnowledgeBase (UniProtKB) 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 by the Ensembl and Ensembl Genomes databases and in the GOA project to provide large-scale mapping of proteins to GO terms.

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.

Members of the team also develop and maintain the protein2GO tool used by the Gene Ontology Annotation (GOA) team and the QuickGO browser.

Future projects and goals

We are currently planning an overhaul of the InterPro web interface and web services so that more users will be able to easily access and interpret our data. Our intention is to allow for more complex querying and more navigable web pages; we also intend to provide more data via REST and SOAP-based web services. We are rewriting the InterProScan software package to improve its flexibility and modularity and bring it in line with our internal production pipelines. To be able to cope with the increasing number of new signatures being developed by InterPro consortium members, we are in the process of improving our internal curation tools to aid rapid signature annotation and integration.

Future plans for CluSTr include use of its coverage of unique sequence sets to identify potential conserved protein families for new InterPro signature building. As a consequence, it is likely we will have to improve the web services available to access the data.

Match view of protein Q05652, a probable Drosophila *serine/threonine-protein kinase. This view includes the matches displayed in the Domain Architecture, table and detailed graphical views. Structure matches are shown at the bottom as white striped bars, and links are provided to UniProtKB, DASTy, SPICE and GOA. Matches to unintegrated signatures are included.*

Selected references

Dunn, M.J. *et al*. (2008). EuPA achieves visibility - An activity report on the first three years. *Journal of Proteomics,* 71, 11-18

Hunter, S. *et al*. (2008). InterPro: the integrative protein signature database. *Nucleic Acids Res.*, doi: 10.1093/nar/gkn785

Jones, P. *et al*. (2008). PRIDE: New developments and new datasets. *Nucleic Acids Res.,* 36, D878-D883

Ensembl Genomes

Paul Kersey

PhD 2005, University of Edinburgh.

Team leader at EMBL-EBI since 2008.

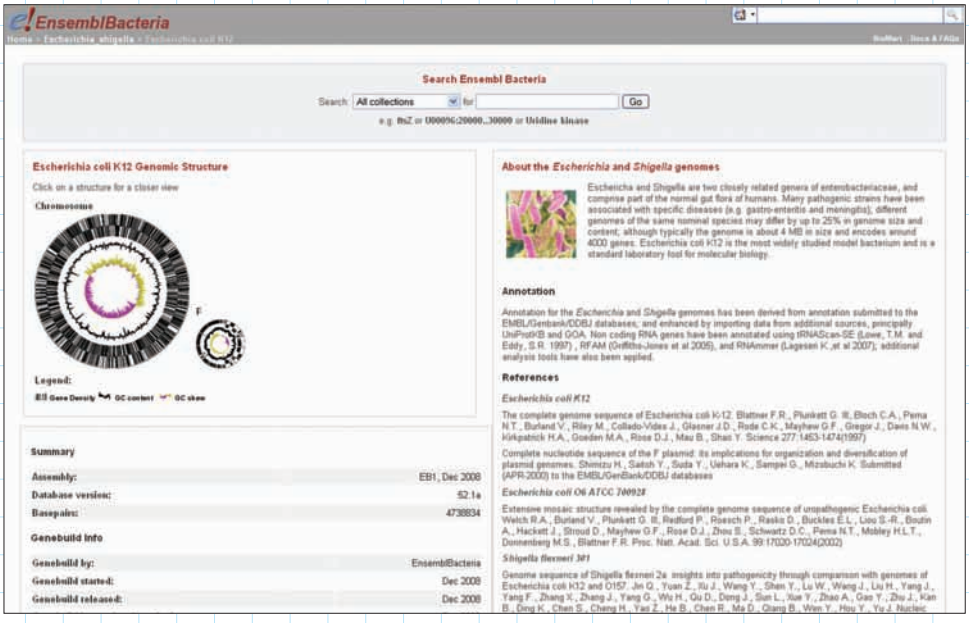
Previous and current research

The activities of the Ensembl Genomes team are focussed on the representation of genome and genome-related data from non-vertebrate species. A huge influx of data is expected for non-vertebrate species in the near future, driven by the use of ultra-high throughput sequencing technologies. The group's core strategy for handling this data is based on the belief that the demands of these species can be met through the re-use and extension of the Ensembl genome annotation, analysis and visualisation platform that has been used successfully for vertebrate genomes since 2000, which will become apparent with the public launch of five new sites (Ensembl Metazoa, Ensembl Protists, Ensembl Plants, Ensembl Fungi and Ensembl Bacteria) in the first half of 2009. At the same time, we are making a major effort to deepen our links with sections of the scientific community working on individual species. Our vision is that Ensembl Genomes will evolve as a project with two faces; to the wider scientific public, we will provide an integrative portal to genomes and related data; while simultaneously we will support an infrastructure and flexible toolkit to empower particular communities to manage genome annotation. Through the use of common technology, we can make it easier for communities to handle genome scale data, and to integrate their systems with the other key EBI resources.

Future projects and goals

2009 will see the launch of the Microme Project, a 13-member European partnership to develop a database and downstream services for metabolic pathways in bacteria. Based on the successful Reactome software infrastructure, Microme is especially timely as pathway-based inference is likely to become increasingly important in the annotation and functional interpretation of bacterial genome sequence, with the potential to suggest biotechnological solutions to problems such as energy and food production, waste decontamination and industrial catalysis.

We will continue to work with the UK and international research communities to establish collaborations to exploit the Ensembl infrastructure in new domains and bring fresh content to Ensembl Genomes. Currently, new proposals are being prepared for human and plant pathogen data, while we are seeking to deepen existing relationships with established resources in the areas of protist pathogens (EuPath DB) and plants (Gramene). The relationship between hosts, vectors and pathogens is of particular interest as Ensembl Genomes (and Ensembl) expand to include increasing numbers of complete disease systems, and we will be looking at ways of linking genotypes to interaction, phenotype and geographic data. We will also be working to develop the interface between genomic and metagenomic data, an area of increasing scientific interest and data production.



Ensembl Genomes. The homepage for the Escherichia/Shigella *class within the new Ensembl Bacteria site.*

Selected references

Field, D. *et al*. (2008a). The minimum information about a genome sequence (MIGS) specification. *Nat. Biotechnol*., 26, 541-547

Megy, K. *et al*. (2008). Genomic resources for invertebrate vectors of human pathogens, and the role of VectorBase. *Infection, Genetics and Evolution*, in press

Mulder, N.J. *et al*. (2008). *In silico* characterization of proteins: UniProt, InterPro and Integr8. *Mol. Biotechnol.,* 38, 165-177

Topalis, P. *et al*. (2008). How can ontologies help vector biology? *Trends Parasitol.,* 24, 249-252

Rodrigo Lopez

Vet. Med Degree 1984, Oslo Vet. Hoyskole and NASAS Cand. Scient.

Molecular Toxicology and Informatics 1987, University of Oslo.

Team leader at EMBL-EBI since 2001.

The External Services Team

Previous and current research

The External Services team is in charge of the development, management and maintenance of web and FTP resources at EMBL-EBI. These are in high demand as more and newer types of data are archived and distributed from the EBI's databases. The group's responsibilities include:

- maintaining the main EBI web portal as well as several Wellcome Trust, BBSRC and EU-funded projects websites (amongst others), including: 1000 Genomes Project, 2can, BioCatalogue, BioSapiens, European Genotype Archive (EGA), ELIXIR, EMBRACE, ENFIN, FELICS, IMPACT, INSDC, MIBBI and SYMBIOmatics;
- monitoring and supporting use of EBI services, such as the global search and retrieval services that include the EB-eye search engine and SRS (SRS especially for the European Nucleotide Archive – ENA), and core sequence search and analysis services such as BLAST, FASTA, InterProScan, as well as several mainstream multiple sequence alignment (MSA) services such as ClustalW2, MAFFT, T-Coffee, etc. Today all these services have SOAP/REST web services interfaces, which allows programmatic and systematic use;
- user support for EBI developers as well as general users. In this context, various members of the team have been heavily involved in training and conference activities and participation at local and international workshops.

Future projects and goals

The adoption of web services and XML technologies, which started during 2002, translates to better and more robust services today. Users now have systematic access to our analytical tools. Our web services present a uniform API for a wide variety of bioinformatic applications that count more than 250 individual algorithms. Access to these represents slightly over 900,000 compute jobs per month.

In the context of facilitating querying by our users, the group is involved in the development of browsable catalogues that will enable users to find web services, tools and databases. These projects are run in collaboration with members of the EMBRACE project (the EMBRACE service registry) and the staff at the University of Manchester (BioCatalogue). The projects are financed by the EU and the BBSRC respectively. The aims of these catalogues are to produce a much-needed 'yellow pages' for life sciences resources and their outcome will be linked with mainstream search engines such as EB-eye and Google.

Work in relation to sequence patents is continuing as part of the team's collaboration with the EPO. The latest developments in this context include non-redundant sets of data, level 1 and level 2 equivalencies, and additional patent sequence databases from both Korea and China. This work leads to a much clearer understanding of when and where a patent was first filed, but importantly, what is actually covered by a patent or by its application. This work is of particular importance to patent examiners currently using EBI services, which include the EPO (based in The Hague, Netherlands and Munich, Germany), Spain, UK, USPTO, KIPO and JPO.

Selected references

Chica, C. *et al.* (2008). A tree-based conservation scoring method for short linear motifs in multiple alignments of protein sequences. *BMC Bioinformatics*, 9, Article 229

Cochrane, G. *et al.* (2008). Priorities for nucleotide trace, sequence and annotation data capture at the Ensembl Trace Archive and the EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, 36, D5-D12

Robinson, J. *et al.* (2008). The IMGT/HLA database. *Nucleic Acids Res.*, 37, D1013-D1017

Larkin, M.A. *et al.* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23, 2947-2948

Chemogenomics and drug discovery

John Overington

PhD 1991, Birkbeck College, University of London.

Postdoctoral work at Imperial Cancer Research Fund.

Manager, Molecular Informatics Structure and Design, Pfizer, Sandwich.

Senior Director, Molecular Informatics, Inpharmatica.

Team leader at EMBL-EBI since 2008.

Previous and current research

The human genome project has delivered huge potential for the discovery of novel therapeutics. In sharp contrast to this potential and promise is the small actual translation of functional genomics discoveries into clinically useful agents. The goal of our group is to understand the signalling, regulatory and physicochemical characteristics of historically successful drug targets, and to provide predictive methods to prioritise future potential drug targets.

Our lab relies of a series of public domain databases to perform its research, many of which have been built and developed within our own group. These databases cover medicinal chemistry, clinical development and launched drugs, and are known by the name ChEMBL. Additionally, all our work is computational although we have collaborations with key experimental groups. Finally, we apply a broad range of knowledge discovery from data (KDD), data visualisation, and predictive methods in our work.

Key to the assessment of future drug targets is an integrative approach which can consider biochemical binding data, alongside functional and systemic effects and also include protein structure and binding site data. As such we apply a uniquely broad set of approaches from cheminformatics, bioinformatics, homology modelling, docking and machine learning.

We are a newly established group at EMBL-EBI, following the award of a grant from the Wellcome Trust to transfer our previous databases and research from the private to public domain.

Future projects and goals

We are interested in the informatics-based prioritisation of drug targets for application in the area of neglected diseases, in particular those diseases caused by pathogenic organisms. We are also interested in data-mining approaches to allow semi-automated design and optimisation of hit and lead chemical series using generalisation of systematic rules discovered in our databases. Finally we have a strong interest in the application of these techniques for newer classes of therapeutics such as monoclonal antibodies (mAbs) and non-human secreted proteins (for example, helminth proteins that naturally suppress immune response in the host organism).

Selected references

Agüero, F. *et al.* (2008). Genomic-scale prioritisation of drug targets: TDRtargets.org. *Nat. Rev. Drug. Discov.*, 7, 900-907

Hopkins, A.L., Mason, J.S. & Overington, J.P. (2006) Can we rationally design promiscuous drugs? *Curr. Opin. Struct. Biol.*, 16, 127-136

Overington, J.P., Al-Lazikani, B. & Hopkins, A.L. (2006) How many drug targets are there? *Nat. Rev. Drug Discov.*, 5, 993-996

Danilewicz, J.C. *et al.* (2002). Design and synthesis of thrombin inhibitors based on the (R)-Phe-Pro-Arg sequence. *J. Med. Chem.*, 45, 2432-2453

Mizuguchi, K. *et al.* (1998). HOMSTRAD: a database of protein structure alignments and homologous families. *Protein Science*, 7, 2469-2471

Peter Rice

BSc 1976, University of Liverpool.

Previously at EMBL Heidelberg (1987–1994), the Sanger Centre (1994–2000) and LION Bioscience (2000–2002).

Team leader at EMBL-EBI since 2003.

Grid and e-Science research and development

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 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 e-Science project) and for various projects (including EMBRACE and ComparaGrid) that integrate access to bioinformatics tools and data content.

To date, Grid development has focussed on the basic issues of storage, computation and resource management needed to make a global scientific community's information and tools accessible in a high-performance environment. However, from the e-Science point of view, the purpose of the Grid is to deliver a collaborative and supportive environment that enables geographically distributed scientists to achieve research goals more effectively, while allowing their results to be used in developments elsewhere.

Our group has been the biological specialist participant in the UK-funded myGrid project and this collaboration is continuing with the Open Middleware Infrastructure Institute (OMII-UK). This project was aimed at developing and maintaining open source high-level service-based middleware to support the construction, management and sharing of data-intensive *in silico* experiments in biology. EMBL-EBI's role is through the Taverna workbench and as an application and data service developer and provider which continues through the EMBRACE and EMBOSS projects.

A key factor in the success of EMBOSS, and in particular its selection as the application platform for the EMBRACE and myGrid projects, has been its development and implementation of the AJAX Command Definition standard or ACD files. These define the interface of each EMBOSS application and are directly used by the application on startup for all processing of the command line and interaction with the user.

The EMBRACE project, an EU-funded Network of Excellence, is now in its second year, with the aim of defining and implementing a consistent standard interface to integrate data content and analysis tools across all EMBL-EBI core databases and those provided by our partners. The early focus of this five-year project has been on the sequence and structure data resources at EBI and the EMBOSS applications. Our group is also active in defining the core technologies to be used by EMBRACE, including BioMart data federation methods, web services provided by the EBI External Services team, and the Taverna workbench as an end-user client.

Future projects and goals

The services provided by the group remain largely SOAP-based web services. These have proved 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 e-Science 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

Belhajjame, K. *et al.* (2008). Metadata management in the taverna workflow system. In 'Proceedings CCGRID 2008 – 8th IEEE International Symposium on Cluster Computing and the Grid', 651-656

Lanzen, A. & Oinn, T. (2008). The Taverna Interaction Service: Enabling manual interaction in workflows. *Bioinformatics*, 24, 1118-1120

Li, P. *et al*. (2008). Automated manipulation of systems biology models using libSBML within Taverna workflows. *Bioinformatics*, 24, 287-289

Li, P. *et al*. (2008). Performing statistical analyses on quantitative data in Taverna workflows: An example using R and maxdBrowse to identify differentially-expressed genes from microarray data. *BMC Bioinformatics*, 9, Article 334

The Microarray Software Development Team

Ugis Sarkans

PhD in Computer Science, University of Latvia, 1998.

Postdoctoral research at the University of Wales, Aberystwyth.

Technical team leader at EMBL–EBI since 2006.

Previous and current research

Our team has been developing software for ArrayExpress since 2001 (Sarkans *et al.*, 2005). As of October 2008 ArrayExpress holds data from almost 200,000 microarray hybridisations and is one of the major data resources of EMBL-EBI. The software development team has built the following components of the ArrayExpress infrastructure:

- Repository – the archival MIAME-compliant database for the data that support publications;
- Data Warehouse – a query-oriented database of gene expression profiles;
- MIAMExpress – a data annotation and submission system;
- Expression Profiler – a web-based data analysis toolset;
- components used internally by the ArrayExpress production team.

In 2008 we continued our efforts to rebuild the ArrayExpress infrastructure, with an emphasis on simplification of data management tasks and software components, tighter integration of data submission and management, and use of automated workflows for data processing.

The team is also working on data management and integration solutions for domains beyond microarray data. We participate in the MolPAGE project, an integrated EU effort that aims to find biomarkers for genetic diseases, and type II diabetes in particular. Our team's main achievements in this project are: the basic principles and architecture of a framework for managing human sample information and output of high-throughput analytical platforms; and a generic data reannotation, integration and warehousing solution.

We also participated in data standardisation efforts, in particular MAGE (Spellman *et al.*, 2002) and MAGE-TAB (Rayner *et al.*, 2006) and regard this work as crucial for the successful development and support of a high-throughput data management infrastructure.

Future projects and goals

The main goal for 2009 is gradually rolling out the next-generation infrastructure for ArrayExpress. The main tasks of this project include:

- replacing the central databases of ArrayExpress with a single unified database;
- organising a system for synchronising the old and new databases, enabling testing and gradual switchover to the new infrastructure;
- porting the existing workflows to the new system;
- porting ArrayExpress interfaces to work with the new back-end.

There are significant changes planned also from the perspective of external users. There is a clear conceptual separation between ArrayExpress Repository and ArrayExpress Warehouse/Atlas. The former serves the need to provide an archive of scientific record, working in tandem with journal publications, while the latter provides researchers with biologically meaningful information about expression of genes in various conditions – either in a more 'raw' form (as in the Warehouse), or already processed (as in the Atlas). However, there is no need to maintain a distinction between these resources from the web access point of view. We plan to build a unified interface that, in response to user queries, will provide information on all levels of granularity: whole experiments (Repository), sets of expression profiles (Warehouse), and interesting facts about gene expression (Atlas), providing users with a choice for further investigation.

Selected references

Haudry, Y. *et al.* (2008). 4DXpress: a database for cross-species expression pattern comparisons. *Nucleic Acids Res.*, 36, D847-D853

Jones, A.R. *et al.* (2007). The Functional Genomics Experiment model (FuGE): An extensible framework for standards in functional genomics. *Nat. Biotechnol.*, 25, 1127-1133

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, 1495-1501

Spellman, P.T. *et al.*, (2002). Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol.*, 3, RESEARCH0046

Christoph Steinbeck

PhD 1995, Rheinische Friedrich-Wilhelm-Universität, Bonn.

Postdoctoral research at Tufts University, Boston and the Max-Planck-Institute of Chemical Ecology, Jena, Germany, 1997-2002.

Habilitation, 2003, Organic Chemistry, Friedrich-Schiller-Universität, Jena, Germany, 2003.

Head of Research Group for Molecular Informatics, Cologne University Bioinformatics Center (CUBIC), 2002-2007.

Lecturer in Chemoinformatics, University of Tübingen, 2007.

Team leader at EMBL-EBI since 2008.

Chemoinformatics and metabolism

Previous and current research

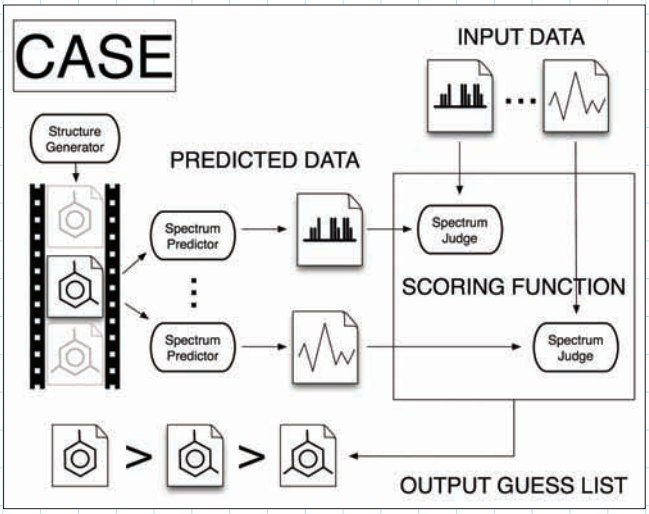
The Chemoinformatics and Metabolism team aims to provide the biomedical community with information on small molecules and their interplay with biological systems. The group develops methods to decipher, organise and publish the small molecule metabolic content of organisms. We develop tools to quickly determine the structure of metabolites by stochastic screening of large candidate spaces and enable the identification of molecules with desired properties. This requires algorithms for the prediction of spectroscopic and other physicochemical properties of chemical graphs based on machine learning and other statistical methods.

We are further investigating the extraction of chemical knowledge from the printed literature by text and graph mining methods, improved dissemination of information in life science publications, as well as open chemoinformatics workflow systems. Together with an international group of collaborators we develop the Chemistry Development Kit (CDK), the leading open source library for structural chemoinformatics as well as the chemoinformatics subsystem of Bioclipse, an award-winning rich client for chemo- and bioinformatics.

Future projects and goals

The recently acquired resource of large-scale drug activity data at the EBI creates exciting new opportunities both on the research and service side (www.ebi.ac.uk/Information/News/pdf/Press23July08.pdf). Our team has started to create an open source chemical search engine for the new resource, which will be the first open source chemistry search engine for the widely used OracleTM Database system. A combination of the new chemogenomics data and the Chemistry Development Kit will allow us to create open structure-activity models and to assist efforts in wet lab screening in areas such as library design.

On the service side, ensuring a sustainable growth for the ChEBI database will be the focus of our attention. The number of marketed and developed drugs in the world drug index alone currently amounts to more than 80,000 compounds. Assuming only a handful of metabolites are produced by organisms upon application of these drugs, the task ahead takes shape. Not only does this task require a larger team for data collection and curation but also research into the automated assembly and validation of ChEBI datasets to aid the human curators. Last but not least, 2009 will reveal the EBI's solution on how to integrate the chemogenomic data with existing chemical resources at the institute.



Computer-Assisted Structure Elucidation uses a structure generation engine to produce chemical spaces based on boundary conditions such as the gross formula of the unknown compound, determined for instance by mass spectrometry. These chemical spaces are then crawled and candidate structures in them inspected for fitness by comparing predicted and measured properties such as NMR spectra. Based on calculated fitness values, a ranking is presented to the user.

Selected references

Kuhn, S. *et al.* (2008). Building blocks for automated elucidation of metabolites: Machine learning methods for NMR prediction. *BMC Bioinformatics,* 9, 400

Willighagen, E.L. *et al.* (2007). Userscripts for the life sciences. *BMC Bioinformatics,* 8, 487

Han, Y.Q. & Steinbeck, C. (2004). Evolutionary-algorithm-based strategy for computer-assisted structure elucidation. *J. Chem. Inf. Com. Sci.,* 44, 489-498

Steinbeck, C. (2001). SENECA: A platform-independent, distributed, and parallel system for computer-assisted structure elucidation in organic chemistry. *J. Chem. Inf. Com. Sci.,* 41, 1500-1507

Literature resource development

Peter Stoehr

MSc Applied Genetics 1978, Birmingham University.

Statistician in Agriculture Faculty, Newcastle University.

Analyst/Programmer with AFRC, Cambridge and Harpenden.

Team leader at EMBL-EBI since 1996.

Previous and current research

The major goal of literature resource development is to integrate the scientific literature with the data in biological databases and provide public services to exploit this. We achieve this by implementing a state-of-the-art search engine, flexible web access, novel biomedical text mining methods and ontologies such as Gene Ontology (GO) and Unified Medical Language System (UMLS).

We develop literature resources for use in-house and in public services. A local copy of PubMed is maintained under lease from the US National Library of Medicine (NLM), supplemented by bibliographic data from other sources such as AGRICOLA (USDA-NAL) and Chinese Biology Abstracts (CAS-SICLS). Biological patent abstracts are captured from the European Patent Office (EPO) and the US Patent Office (USPTO).

CiteXplore has been developed as a tool for querying the scientific literature, showcasing text mining methods, and linking to biological databases. UMLS, GO, the NCBI taxonomy and gene synonyms from UniProt are used as thesauri. Text mining methods from the research community, such as the 'Whatizit' methods of the Rebholz-Schuhmann research group at EMBL-EBI (www.ebi.ac.uk/Rebholz-srv/whatizit/form.jsp) provide several 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.

Future projects and goals

Text mining functionality and enhanced CiteXplore: We plan to accelerate exploitation of recent methodology into CiteXplore in areas such as GO and MeSH annotations, related articles, inclusion of semantic information in the indexing, and sentence/paragraph retrieval in addition to whole documents. We will explore the use of BioLexicon, a terminological resource generated from EBI databases, to facilitate interoperability of literature with the databases.

Additional bibliographic data: In collaboration with the British Library we will identify additional content for UKPMC, for example UK National Health Service publications and NICE guidelines, among others, and the bibliographic data for these will be exposed via CiteXplore.

Citation networks: We will begin to process UKPMC articles that are available only as scanned images, using the results of OCR and citation extraction from NCBI. An extensive task is to complete the pipeline of harvesting relevant scientific articles from the web, and accurately extract citations and their context from these. As the citation network fills, we will include citation counts in the CiteXplore indexing and ranking function, in the same fashion as the Google PageRank method, to enable highly-cited articles to appear more prominently in search results.

Screenshot of a PubMed record in CiteXplore, showing mark-up of proteins and organisms in the text, links to the complete article (a free PDF at UKPMC), article citations and cross-references to EMBL nucleotide sequence databases.

Web services: Further web services interfaces to all of the CiteXplore functionality will be developed to make the bibliographic data aggregated at the EBI available to third party information systems, workflows and research tools.

From June 2009, the Literature Services team will be led by Jo McIntyre.

Selected references

Nikolov, N. & Stoehr, P. (2008). Integrating biomedical publications with existing metadata. In 'Proceedings - IEEE International Symposium on Computer-Based Medical Systems', 653-655

Nikolov, N. *et al.* (2007). CiteXtract: Extracting citation data from biomedical literature. In 'Proceedings - IEEE International Symposium on Computer-Based Medical Systems', 12-14

Rebholz-Schuhmann, D. *et al.* (2007). EBIMed - text crunching to gather facts for proteins from Medline. *Bioinformatics*, 23, e237-244

Weimin Zhu

MSc, 1993, University of Toronto.

Project Manager, GDB Project, Toronto, until 2000.

Head of Bioinformatics, Synax Pharmar, Toronto, Canada, until 2002.

Team leader at EMBL-EBI since 2002.

Database Research and Development Group activities

Previous and current research

In February 2008, the Database Applications team was reorganised to form the Database Research and Development group due to the creation of the PANDA group.

As the Database Application team, the team's main functions were to provide software/tool development and maintenance support for the EMBL Nucleotide Sequence Database, and Oracle database support for all the Oracle databases in the PANDA group. The team also coordinated hardware resources with the EBI's Systems and Networking team to ensure the hardware requirements for the projects within PANDA were met.

The group's new mandate is to conduct research and development to find new technologies and solutions to meet challenges related to very large databases (VLDB), which includes data distribution problems when network speed is a bottleneck, and the solutions required to manage and query VLDBs efficiently.

The size of bioinformatics databases has been increasing exponentially over the last ten years. Some core resources are approaching, or have already reached, multi-terabytes in size. This trend of growth has accelerated in recent years by the introduction of new data types and high-throughput data producing technologies. Today, we are facing all the challenges a VLDB brings, such as those in data operational management, data access performance, and data mirroring and distribution. Our current infrastructure in these areas thus require upgrading in order to realise the full potential of data-rich resources, and optimise the usage of our human and hardware resources.

This year, our main focus has been to analyse the uniqueness of bioinformatics datasets, and conduct research into possible solutions to large dataset distribution problems, especially over slow networks.

Future projects and goals

The development of a new delta compression algorithm will be continued. The goal is to develop prototype software, along with a data model and repository to store the pre-computed metadata for the algorithm, to test the key features of this new distribution system. The later stage of this work will be undertaken in collaboration with the Systems and Networking team and database projects requiring the distribution of large data files, as well as with external partners in countries with slow network connections to the EBI. This will allow us to benchmark the network saving for distributing large datasets using delta compression.

The data distribution problem is not unique to the bioinformatics community. It is also a problem for other domains, such as particle physics and earth science. A workshop is planned for 2009 to share knowledge among scientists across specialisations. Another focus of this workshop is to bring together Chinese bioinformaticians and network backbone administrators to explore the network optimisation possibilities between China and the EBI.

We will continue to seek external funding for the project's long term goal – to have a stable and well-maintained distribution system for the EBI's large databases.

Selected references

Cochrane, G. *et al.* (2008). Priorities for nucleotide trace, sequence and annotation data capture at the Ensembl Trace Archive and the EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, 36, D5-D12

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 ski slopes. The outstation, a laboratory of about 85 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 high-precision diffractometers and frozen crystal sample changers. More recently, a new X-ray small-angle scattering instrument has been built by ESRF and EMBL and is now operational.

High-throughput methods have also been introduced in other steps of the structure determination process. This development was closely connected to the outstation's involvement in the EU-funded SPINE project (Structural Proteomics in Europe) and its successor in FP6 (SPINE2-Complexes). A very successful robotic system for nanovolume crystallisation has been implemented, and a novel, high-throughput selection method, ESPRIT, has been developed for finding soluble protein fragments from otherwise badly expressed or insoluble proteins. More recently, a Eukaryotic Expression Facility (EEF) has been established to specialise in expression of multi-subunit complexes in insect cells. These platforms form part of the Partnership for Structural Biology (PSB, www.psb-grenoble.eu), which has been established with the neighbouring ESRF, ILL and the French national Institut de Biologie Structurale (IBS). The PSB is partly housed in a new building adjacent to the outstation, together with the CNRS-Grenoble University-EMBL Unit of Virus Host Cell Interactions (UVHCI).

As a result of these local developments, outstation scientists have access to a wide range of techniques including molecular biology, several biophysical techniques, cryo-electron microscopy, isotope labelling, NMR, neutron scattering, X-ray crystallography and small angle scattering. In 2008 a confocal microscope with facilities for cross-correlation spectroscopy was installed for the study of complex formation in cells, and during 2009 a new top-end electron microscope will become available with cryo-tomography capability. 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 cell 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 and has recently focussed on elucidation of the mode of action of a novel boron-containing antibiotic which targets leucyl-tRNA synthetase. Projects involving protein-RNA interactions also include cryo-EM studies of the signal recognition particle and its interaction with its receptor and the ribosome and other proteins and complexes involved in RNA processing, transport and degradation, such as the nonsense-mediated decay (NMD) pathway. The analysis of protein-DNA interactions and mechanisms of transcriptional regulation is another important topic. Structural analysis of eukaryotic transcription factor DNA complexes is continuing with new groups working on TFIID, complex enhanceosomes and the dosage compensation complex. Recently a group working on the molecular biology of microRNAs has been established which is trying to elucidate the role of piRNAs in the germ line.

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. Recently the first crystal structures of domains of the influenza virus polymerase have been determined, which depended on the prior identification of soluble fragments using the ESPRIT method mentioned above.

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, Senior Scientist and Head of Outstation since 1989. Joint appointment with Gene Expression Programme.

Director of EMBL/Grenoble University/CNRS Unit for Virus Host Cell Interactions.

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. A current major focus is on influenza virus polymerase and innate immune system receptors.

In eukaryotic cells, nascent Pol II RNA transcripts (e.g. mRNA or snRNA) are rapidly given an m^7Gppp 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 specifically charging 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.

Previously we studied 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 determined the structure of the adenovirus penton base, which is at the twelve 5-fold vertices of the icosahedral viral capsid, and showed how it interacts with the N-terminal tails of the trimeric fibre. This led to a quasi-atomic model of the whole virus capsid based on a 10Å cryo-EM reconstruction.

Future projects and goals

We have several ongoing projects related to RNA metabolism, our goal being to obtain and use structures of the complexes involved to understand function. These include studies on the vaccinia virus mRNA capping complex (we solved the structure of the complete tri-functional enzyme) and continued work on proteins involved in NMD (we recently determined the structure of the UPF1-UPF2 complex). Work is continuing on the leucyl- and prolyl-tRNA synthetase systems, both of which have editing activities. We collaborated in the elucidation of the mechanism of action of a new anti-fungal compound that targets the editing site of leucyl-tRNA synthetase and are now extending this by using structure-based approaches to design new anti-bacterials. A major ongoing project is structure determination of the trimeric influenza virus RNA-dependent polymerase, the viral replication machinery. We have collaborated in the structure determination of four distinct domains from the polymerase, the C-terminal domain of the PB2 subunit involved in nuclear import, the 627-domain of PB2 (which contains important host determinant, residue 627) and domains containing the two key active sites involved in the 'cap-snatching' process of viral mRNA transcription: the cap-binding site in PB2 and the endonuclease in PA. These results give some insight into the polymerase mutations required to adapt an avian virus to be able infect humans and also give a boost to structure-based antiviral drug design. We are also engaged in fluorescence studies of the transport and assembly of the influenza polymerase in living cells. Finally we are 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 RIG-I like helicases, which signal interferon production upon detection of viral RNA.

Selected references

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R. *et al.* (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.*, 15(5), 500-6.

De la Pena, M., Kyrieleis, O.J. & Cusack, S. (2007). Structural insights into the mechanism and evolution of the vaccinia virus mRNA cap N7 methyl-transferase. *EMBO J.*, 26, 4913-4925

Rock, F.L., Mao, W., Yaremchuk, A., Tukalo, M., Crepin, T. *et al.* (2007). An antifungal agent inhibits an aminoacyl-tRNA synthetase by trapping tRNA in the editing site. *Science*, 316, 1759-1761

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

Structural biology of eukaryotic complexes in gene expression

Imre Berger

PhD 1995, MIT Cambridge and Leibniz University, Hannover.

Postdoctoral research at MIT and the Institute of Molecular Biology and Biophysics (IMB), ETH Zürich.

Habilitation 2005, ETH. Group leader at IMB from 2005.

Group leader at EMBL since 2007.

Previous and current research

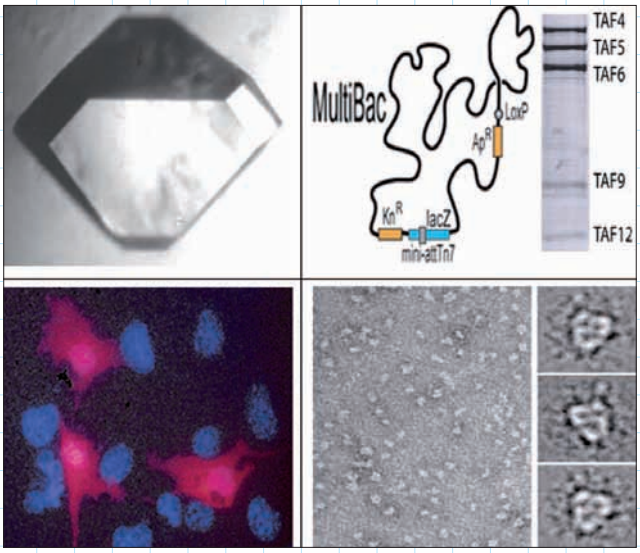
An intense focus of current biological research efforts is the elucidation of protein interaction networks (interactome). Many large multiprotein complexes are discovered. This poses considerable challenges for molecular level studies, in particular for eukaryotic multiprotein complexes with intracellular quantities refractory to large-scale extraction from source. Our research is focussed on developing new technologies to obtain, within a reasonable time-frame, well-defined and homogeneous samples of human multiprotein assemblies in transcription and hereditary disease, which we then use for high-resolution structural and functional analyses.

Our major methodologies include molecular biology and X-ray crystallography. However, we readily apply techniques from other fields that are required for our research, both by ourselves and through collaborative efforts. A prerequisite for successful structural study of many complexes, both by electron microscopy and X-ray crystallography, is production of homogeneous, stable specimens. Present recombinant expression methods often require considerable investment in both labour and materials before multiprotein expression, and after expression and biochemical analysis do not provide flexibility for expressing an altered multiprotein complex. To meet these demands, we have introduced MultiBac, a modular, baculovirus-based system specifically designed for eukaryotic multiprotein expression (Berger *et al.*, 2004, *Nat. Biotechnol.*). Recently, we have harnessed homologous and site-specific recombination methods in tandem for all steps involved in multigene assembly, thus providing a flexible, automatable platform for generation of multiprotein expression vectors and their rapid regeneration for revised expression studies. We have successfully implemented all steps involved in a robotics setup. By using our technology, we produced numerous large multiprotein assemblies in sufficient quantity and quality for structural studies, including the presumed ~700 kDa heterodecamer scaffold of human TFIID general transcription factor.

Future projects and goals

At EMBL Grenoble, we will continue to advance our expression technologies to entirely automate and standardise the process of production for eukaryotic gene regulatory multiprotein complexes including the entire human TFIID holoenzyme and its various isoforms. In collaboration with the Schaffitzel group (page 96), we will subject the complex specimens produced to electron microscopic analyses. We will use homogeneous complexes thus identified for X-ray crystallography.

By enlisting state-of-the-art mass spectrometric methods from systems biology, we will address a further bottleneck in complex crystallography, namely the challenge of defining crystallisable core assemblies of multiprotein complexes in a reasonable time frame (in collaboration with ETH Zürich). Also, we will expand our multiprotein expression strategies to prokaryotic and mammalian hosts.



We develop and utilise advanced, automated technologies to produce eukaryotic multiprotein complexes for structural and functional analysis by a variety of methods including X-ray crystallography.

Selected references

Bieniossek, C. & Berger, I. (2009). Towards eukaryotic structural complexomics. *J. Struct. Funct. Genomics*, 10, 37-46.

Bieniossek, C., Richmond, T.J. & Berger, I. (2008). MultiBac: multigene baculovirus-based eukaryotic protein complex production. *Curr Protoc Protein Sci.*, Chapter 5, Unit 5.20

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J. & Berger, I. (2006). Protein complex expression by using multigene baculoviral vectors. *Nature Methods*, 3, 1021-1032.

Schaffitzel, C., Oswald, M., Berger, I., Abrahams, J.P., Koerten, H., Koening, R. & Ban, N. (2006). Structure of the *E. coli* signal recognition particle bound to a translating ribosome. *Nature*, 444, 503-506.

Florent Cipriani

BSc 1974, Physics, University of Grenoble.

Senior engineer in nuclear and medical industries.

At EMBL Grenoble since 1991. Team leader since 2003.

Diffraction Instrumentation Team

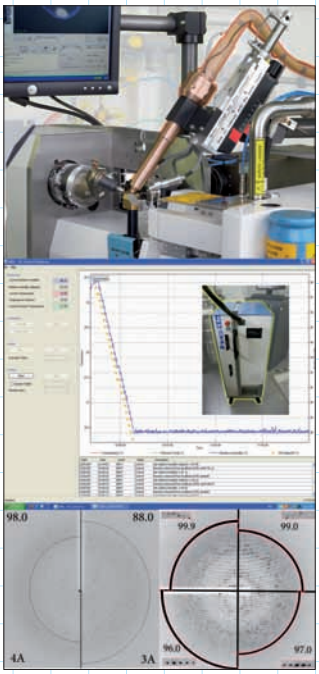
Previous and current research

Our main activity is developing instruments and methods for X-ray scattering experiments. Several collaborative projects are ongoing with the ESRF MX group, EMBL Hamburg and the MRC France (ESRF-BM14) to provide European scientists with state-of-the-art automated beamlines.

The instruments developed over the last five years, such as the MD2x diffractometers, the SC3 sample changer and the MK3 Kappa goniometer, have reached maturity. They are commercially available and equip a number of synchrotrons worldwide. The ESRF MX beamlines rely on this hardware and on our C3D crystal centring software to automatically process several hundred crystals per day. All these technologies are continuously being improved. A customised version of the MD2/MK3 equipment has been defined for the future EMBL@PETRA3 MX2 beamline (2010).

Despite continuous advances in MX, crystals diffracting quality is often a limiting factor to obtain the desired structural data. A new crystal dehydration device is being developed. The first version (HC1b) has been successfully tested in collaboration with the MRC-BM14 beamline and the ESRF (ID14-2, ID14-1; see figure 1). The HC1b will be available to users at the ESRF in the second quarter of 2009. The air dispensing nozzle installed at the position of the usual cryo-cooling head makes the new machine fully compatible with standard MX beamlines and provides a room temperature airflow with adjustable relative humidity (50-99%). The control software of the machine is connected to the host beamline to run dehydration scenarios while taking X-ray diffraction images automatically. The HC1b performance is similar to other dehydration systems but much more ergonomic and faster, and more controlled experiments than ever before are now possible. Several internal and external crystals have been tested and positive results observed in a number of them. These effects range from improved resolution, reduced mosaicity, reduced background, new cryo conditions, more homogenous cryo-cooling and space group changes. A new version (HC2), with temperature control down to 5°C, is under development.

Small Angle X-ray Scattering (SAXS) is a method to study native biological macromolecules in solution, ranging from individual proteins to large complexes. Based on the background acquired at EMBL Hamburg (X33 beamline), we have established a trilateral collaboration with the ESRF and EMBL@PETRA3 to develop an automated sample environment for the third generation synchrotron BioSAXS beamlines. The specification of a new sample changer for micro-volume of protein solutions has been defined and innovative liquid handling processes evaluated. Volumes down to 5µl were transferred automatically from 96 wells SBS microplates into a 2mm diameter quartz exposure capillary with less than 10% losses. An evaluation sample changer setup will be tested in early 2009 at the newly constructed ESRF ID14-3 BioSAXS beamline.



Future projects and goals

In addition to the support we provide for the ESRF-MRC and ESRFMX beamlines, two BioSAXS sample changers will be constructed to equip the ESRF ID14-3 beamline (2009) and an EMBL@PETRA3 beamline (2010). The development of the HC2 dehydration device will be continued in collaboration with Max-lab (I911-2) and Diamond (I02). Our team will also be actively involved in the ESRF MASSIF upgrade project. The goal is to develop a new generation of fast and flexible sample changers, and associated fast imaging/crystal centring systems based on 'on the flight' imaging and zoom-less technologies. Finally, micro-crystallography is a domain we would like continuing to develop. Recent tests with data collection in vertical Omega geometry were very promising; they've shown that sub-micrometer precision can be obtained with Kappa goniometer geometry.

Top to bottom: head of the HC1b at BM14; control software showing a humidity gradient with insert showing the device outside ID14-2; diffraction images showing the improvement observed in Chromatin modification complex (8 to 3Å) and F1-ATPase (3.5 to 2.3 Å).

Selected references

McGeehan, J., Ravelli, R.B., Murray, J.W., Owen, R.L., Cipriani, F., McSweeney, S., Weik, M. & Garman, E.F. (2009). Colouring cryo-cooled crystals: online microspectrophotometry. *J. Synchrotron Radiat.*, 16, 163-72

Cipriani, F., Felisaz, F., Launer, L., Aksoy, J.S. *et al.* (2006). Automation of sample mounting for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.*, 62, 1251-1259

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

Perrakis, A., Cipriani, F., Castagna, J.C., Claustre, L. *et al.* (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

High-throughput protein technologies

Darren Hart

PhD 1996, Oxford University.

Postdoctoral research at Cambridge University.

Group leader at Sense Proteomic Ltd., Cambridge.

Team leader at EMBL Grenoble since 2003.

Previous and current research

Our group develops new molecular biology methods and uses them to work on difficult biological problems. Combinatorial methods (e.g. directed evolution, phage display) are used to address problems that are too complex for rational design approaches. Large random libraries of variants are constructed and screened to identify rare hits with the desired property. In our ESPRIT process, for example, all truncations of a target protein are generated and screened using advanced picking and arraying robotics. With such technologies in hand, we are able to study certain biological questions with advantages over classical approaches. The proteins we study are generally enzymes of biological and medicinal interest:

Influenza RNA polymerase: There is worldwide concern that currently circulating avian influenza viruses will cross the species barrier and become highly pathogenic, human transmissible strains with pandemic potential. This could result from residue changes in several influenza proteins, either by point mutations or through shuffling of the segmented avian and mammalian viral genomes. We are now characterising the interactions of these mutants with host cell factors using both structural and biophysical methods with the aim of understanding mechanisms of influenza host specificity.

Human Kinases: Cells have intricate mechanisms of sensing and responding to environmental changes. Upon a stimulus detected by a cellular receptor the complex system of signal transduction is activated that results in changes in gene expression. Protein kinases play a crucial role in cellular stress responses as mediators between the upstream receptor and downstream gene regulation and are key components in coping with changes in the intra-/extracellular environment. When these mechanisms malfunction, diseases such as excessive inflammation, autoimmune disorders and cancer can occur. Kinases therefore represent important pharmaceutical targets for drug design. The multidomain nature of many kinases reflects the need to regulate the activity of the catalytic activity. We are screening for stable constructs that extend beyond the conserved regions of the catalytic domain, and well-expressed internal domains presumably implicated in complex formation or regulation.

Histone Deacetylases (HDACs): Using our construct screening technology, we have identified well-expressing, catalytically active constructs of an HDAC involved in cholesterol homeostasis. Using these proteins, we are investigating how new inhibitors bind using X-ray crystallography and enzymatic inhibition assays. Secondly, using a library-format protein interaction screen, we are trying to identify HDAC-interacting domains of cellular proteins. If identified, disruption of such protein-protein interactions suggests a new route towards specific HDAC inhibition.

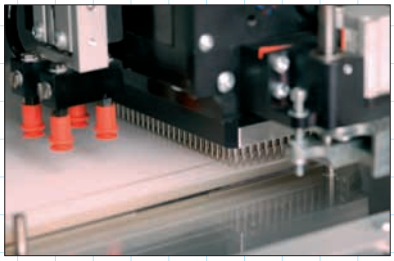


Figure 1: Screening tens of thousands of expression constructs of a target gene. Constructs are made as a random library and printed on membranes for soluble expression analysis by hybridisation of fluorescent antibodies.

Future projects and goals

Difficult biological projects require advanced new tools. We will continue to develop expression methods to handle protein complexes, targets that require eukaryotic expression for correct folding, and possibly aspects of membrane proteins. Each project uses 'real' targets of interest and the aim is use method advancements to yield previously unobtainable biological knowledge. For example, we are testing permutations of influenza-influenza and influenza-host proteins with the aim of defining expressible, crystallisable protein complexes that should provide insights into virus host cell interactions.

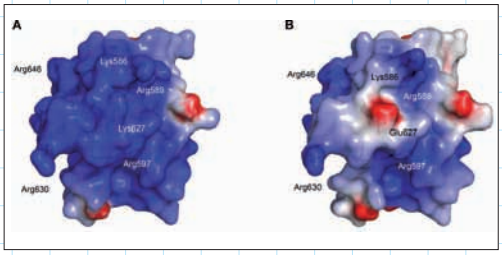


Figure 2: A previously unsuspected domain from influenza polymerase, identified by high throughput expression screening of tens of thousands of random DNA constructs, and structurally characterised by X-ray crystallography. A single mutation to lysine at residue 627 (A) can be responsible for the evolution of human influenza viruses from wild-type avian viruses that have a glutamic acid at this position (B). The mutation of residue 627 reinforces or disrupts a striking basic surface patch and we are seeking to understand how this affects polymerase function.

Selected references

Angelini, A., Tosi, T., Mas, P., Acajjaoui, S., Zanotti, G., Terradot, L. & Hart, D.J. (2009). Expression of *Helicobacter pylori* CagA domains by library-based construct screening. *FEBS J.*, 276, 816-824

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J., Ruigrok, R.W., Ortin, J., Hart, D.J. & Cusack, S. (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.*, 15, 500-506

Tarendeau, F., Crepin, T., Guilligay, D., Ruigrok, R.W., Cusack, S. & Hart, D.J. (2008). Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog.*, 4, e1000136

Tarendeau, F., Boudet, J., Guilligay, D., Mas, P.J., Bougault, C.M., Boulo, S., Baudin, F., Ruigrok, R.W., 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, 229-233

José A. Márquez

PhD 1997, University of Valencia.

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

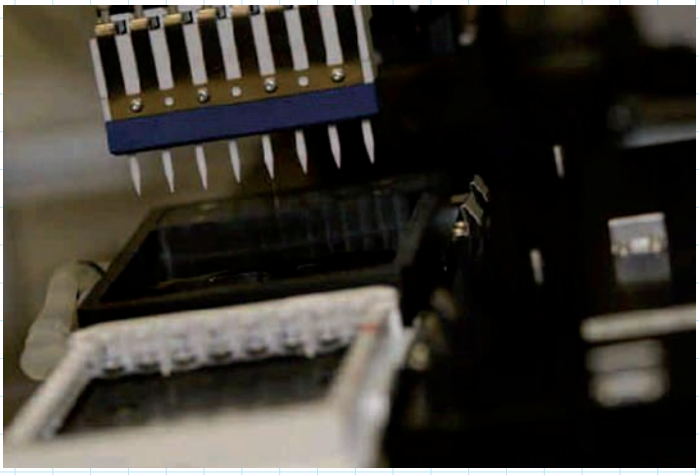
Finding conditions in which biological macromolecules form crystals is recognised as one of the major bottlenecks in structural biology. Once macromolecules are purified, they need to be assayed for crystallisation with a collection of precipitants under different chemical environments. This leads to 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, the HTX Lab, 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. Starting in April 2009 the HTX Lab will offer automated crystallisation screening services to European research through the EC-funded PCUBE project.

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 three hundred registered users, and over two 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. Access is also granted to European researchers through the PCUBE project.

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 coordination 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, EMBL Heidelberg 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 (page 90) 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.



High-throughput crystallisation robot at the HTX Lab.

Selected references

Dimasi, N., Flot, D., Dupeux, F. & Marquez, J.A. (2007). Expression, crystallization and X-ray data collection from microcrystals of the extracellular domain of the human inhibitory receptor expressed on myeloid cells IREM-1. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, 63, 204-208

Marquez, J.A., Galfre, E., Dupeux, F., Flot, D., Moran, O. & Dimasi, N. (2007). The crystal structure of the extracellular domain of the inhibitor receptor expressed on myeloid cells IREM-1. *J. Mol. Biol.*, 367, 310-318

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

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

Synchrotron Crystallography Team

Andrew McCarthy

PhD 1997, National University of Ireland, Galway.

Research associate, Utrecht University.

Postdoctoral research at Massey University and Auckland University.

Staff scientist at EMBL Grenoble.

Team leader at EMBL Grenoble since 2007.

Previous and current research

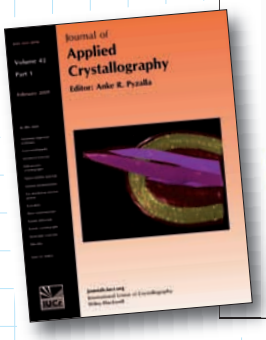
The Synchrotron Crystallography Team, in close collaboration with the MX group of the European Synchrotron Radiation Facility (ESRF), is involved in the design, construction and operation of macromolecular crystallography (MX) beamlines. We are currently responsible for two macromolecular beamlines (ID14-4 and ID23-2) and the newly constructed BioSAXS beamline at ID14-3. The MAD beamline ID14-4 was the first undulator MX beamline to celebrate a decade of user service and has been instrumental in the structural determination of biologically important molecules as well as significantly contributing to pioneering radiation damage studies on biological samples. The highly successful microfocus beamline ID23-2 was the first such beamline in the world, and the newly converted BioSAXS beamline ID14-3 commenced user operation in November 2008. The team also provides scientific and technical support for the CRG beamline BM14 at the ESRF. This partnership with the UK MRC and India enables access to BM14 for EMBL member state scientists. We also work in close collaboration with the Diffraction Instrumentation Team (page 90) to develop hardware, software and novel methodologies for sample handling and data collection possibilities. Recent examples (funded by BIOXHIT) include the mini-kappa goniometer head (MK3) and associated software for optimal crystal reorientation strategies as well as the use of X-ray tomography in MX (figure 1).

The team also studies proteins involved in neuronal development. We are particularly interested in the Slit-Robo signalling complex that is essential for the normal development of the central nervous system. This signalling system has also been implicated in heart morphogenesis, angiogenesis and tumour metastasis. With part funding by SPINE and SPINE2Complexes we have determined a number of structures (figure 2) from this system that maybe important for the development of novel cancer therapeutics. We are also interested in understanding the molecular mechanism of proteins involved in the biosynthesis of plant secondary metabolites, and recently published the structures of two enzymes involved in caffeine biosynthesis. These studies suggest it may be possible to generate a single protein capable of producing caffeine in plants. Such a possibility, when coupled with caffeine's ability to act as a natural pesticide, could enable new ecologically friendly and pest-resistant plants to be created.

Future projects and goals

On ID14-4 the Synchrotron Crystallography Team will continue to develop novel data collection schemes using the MK3 for challenging structural biology projects and the integration of X-ray tomography methods in MX. On ID23-2 we plan to develop specialised methods for the handling and collecting of optimal data from ever smaller crystals. On ID14-3 our team, in collaboration with the Instrumentation team, the ESRF, and the EMBL Hamburg, will be actively involved in the provision of a highly automated BioSAXS beamline. On BM14 we will form a new partnership with India and the ESRF for running the beamline. We hope that all our combined efforts will push the boundaries of protein crystallography currently available to better understand the biological functions of more complex biological systems.

In the laboratory we will continue our research on the Slit-Robo complex by trying to decipher how exactly Slit activates Robo on the cell surface. We plan to tackle this by studying larger fragments of Robo and Slit and using complementary methods to MX where necessary. In collaboration with the ESRF MX group and Néstle Research, France, we plan to expand our current research on secondary metabolic pathways in coffee.



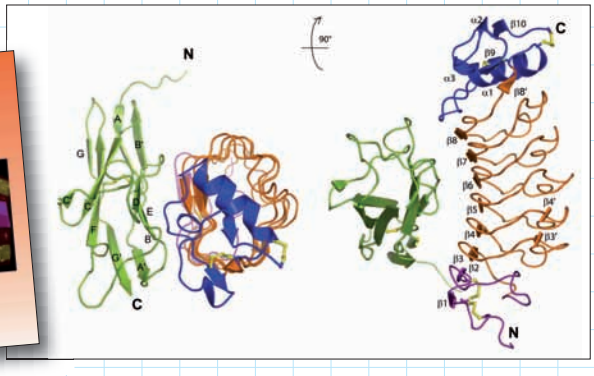


Figure 1: Cover illustration for the 2009 J. Appl. Cryst.*; see Brockhauser* et al.
Figure 2: Structure of Slit2 D2 bound to Robo1 Ig1.

Selected references

Brockhauser, S., Di Michiel, M. McGeehan, J.E., McCarthy, A.A. & Ravelli, R.B.G. (2008). X-ray tomographic reconstruction of macromolecular samples. *J. Appl. Cryst.*, 41, 1057-1066

Martinez Molina, D., Wetterholm, A., Kohl, A., McCarthy, A.A., Niegowski, D. *et al.* (2007). Structural basis for synthesis of inflammatory mediators by human leukotriene C4 synthase. *Nature*, 448, 613-616

McCarthy, A.A. & McCarthy, J.G. (2007). The structure of two N-methyltransferases from the caffeine biosynthetic pathway. *Plant Physiol.*, 144, 879-889

Morlot, C., Thielens, N.M., Ravelli, R.B., Hemrika, W., Romijn, R.A., Gros, P., Cusack, S. & McCarthy, A.A. (2007). Structural insights into the Slit-Robo complex. *Proc. Natl Acad. Sci. USA*, 104, 14923-14928

Daniel Panne

PhD 1999, University of Basel.

Postdoctoral research at Harvard University, Boston.

Group leader at EMBL Grenoble since 2007.

Integrating signals through complex assembly

Previous and current research

Most cellular processes depend on the action of large multi-subunit complexes, many of which are assembled transiently and change their shape and composition during their functional cycle. The modular nature of the components, as well as their combinatorial assembly, can generate a large repertoire of regulatory complexes and signalling circuits. The characterisation and visualisation of such cellular structures is one of the most important challenges in molecular biology today. Characterisation of multicomponent systems requires expertise in a number of techniques including molecular biology, biochemistry, biophysics, structural biology and bioinformatics. We visualise cellular entities using low-resolution imaging techniques such as electron microscopy (EM) and small angle X-ray scattering (SAXS) or high-resolution techniques such as NMR and macromolecular X-ray crystallography (figure 1).

The systems we have been studying are involved in transcriptional regulation. Transcriptional regulation is mediated by transcription factors which bind to their cognate sites on DNA, and through their interaction with the general transcriptional machinery, and/or through modification of chromatin structure, activate or repress the expression of a nearby gene. The so-called 'cis-regulatory code', the array of transcription factor binding sites, is thought to allow read-out and signal processing of cellular signal transduction cascades. Transcriptional networks are central regulatory systems within cells and in establishing and maintaining specific patterns of gene expression. One of the best-characterised systems is that of the interferon-β promoter. Three different virus-inducible signalling pathways are integrated on the 60 base pair enhancer through coassembly of eight 'generic' transcription factors to form the so-called 'enhanceosome', which is thought to act as a logic AND gate. The signal transducing properties are thought to reside in the cooperative nature of enhanceosome complex assembly.

To understand the signal transducing properties of the enhanceosome, we have determined co-crystal structures that give a complete view of the assembled enhanceosome structure on DNA (figure 2). The structure shows that association of the eight proteins on DNA creates a continuous surface for the recognition of the enhancer sequence. Our structural analysis gives us, for the first time, detailed insights into the structure of an enhanceosome and yields important insight into the design and architecture of such higher-order signalling assemblies.

Future projects and goals

We are particular interested in understanding the signal processing through higher order assemblies. As such, the enhanceosome has served as a paradigm for understanding signal integration on higher eukaryotic enhancers. The interferon (IFN) system is an extremely powerful anti-viral response and central to innate immunity in humans. Most serious viral human pathogens have evolved tools and tricks to inhibit the IFN response. Many viruses do so by producing proteins that interfere with different parts of the IFN system. Therefore, our studies are of fundamental interest to understand important signal processing pathways in the cell and may also point to better methods of controlling virus infections; for example, novel anti-viral drugs might be developed which prevent viruses from circumventing the IFN response. Misregulation of IFN signalling pathways is also involved in inflammation and cancer and is therefore of fundamental importance for human health. We will also expand our multiprotein crystallisation strategies to complexes involved in modification of chromatin structure.

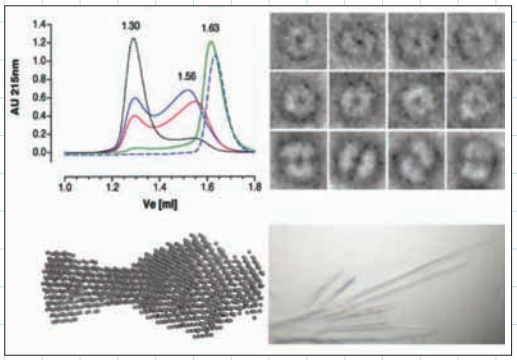


Figure 1 (left): We employ a number of different resolution techniques to visualise cellular structures.

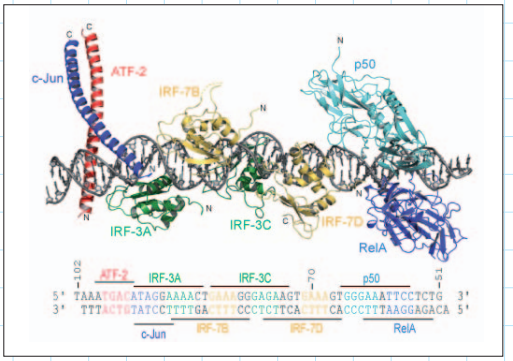


Figure 2 (right): Atomic model of the INF-b enhanceosome.

Selected references

Panne, D. (2008). The enhanceosome. *Curr. Opin. Struct. Biol.*, 18, 236-42

Panne, D., Maniatis, T. & Harrison, S.C. (2007). An atomic model of the interferon-beta enhanceosome. *Cell*, 129, 1111-1123

Panne, D., McWhirter, S.M., Maniatis, T. & Harrison, S.C. (2007). Interferon regulatory factor 3 is regulated by a dual phosphorylation-dependent switch. *J. Biol. Chem.*, 282, 22816-22822

Panne, D., Maniatis, T. & Harrison, S.C. (2004). Crystal structure of ATF-2/c-Jun and IRF-3 bound to the interferon-β enhancer. *EMBO J.*, 23, 4384-4393

Regulation of gene expression by non-coding RNAs

Ramesh Pillai

PhD 2002, University of Bern.

Postdoctoral research at the Friedrich Miescher Institute, Basel.

Group leader at EMBL Grenoble since 2006.

Previous and current research

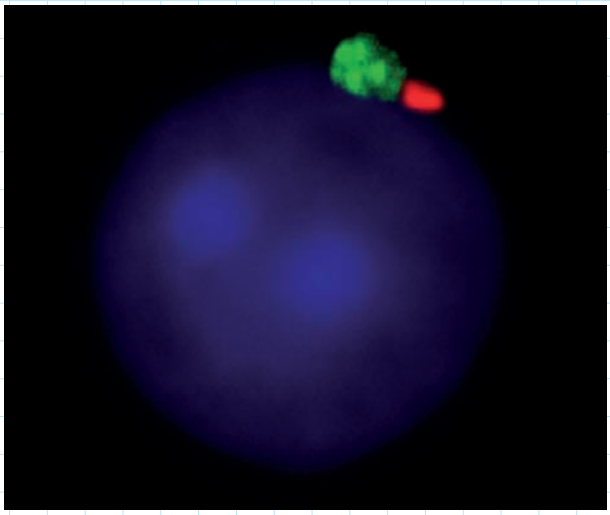
The aim of our research is to understand the molecular mechanisms by which non-coding RNAs, specifically microRNAs (miRNAs), regulate gene expression. MicroRNAs are an abundant class of small non-protein-coding RNAs that function as negative gene regulators at the post-transcriptional level. They are 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 processing bodies (P-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 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. We are also investigating the biogenesis and function of a class of germline-specific small RNAs called the piwi-interacting RNAs (piRNAs) using mouse genetics and biochemical approaches.

Our goal is to understand the mechanism and biology of regulation of gene expression by noncoding RNAs.



Localisation of the piRNA methyltransferase mHEN1 (red) in a cytoplasmic body distinct from the chromatoid body (green) in murine germ cells. Haploid round spermatids from mouse testis were stained with a monoclonal antibody to the piwi protein MILI (green) and rabbit anti-mHEN1 antibody (red). DNA is stained with DAPI (blue).

Selected references

Mathonnet, G., Fabian, M.R., Svitkin, Y.V., Parsyan, A., Huck, L., Murata, T., Biffo, S., Merrick, W.C., Darzynkiewicz, E., Pillai, R.S., Filipowicz, W., Duchaine, T.F. & Sonenberg, N. (2007). MicroRNA inhibition of translation initiation *in vitro* by targeting the cap-binding complex eIF4F. *Science*, 317, 1764-1767

Pillai, R.S., Bhattacharyya, S.N. & Filipowicz, W. (2007). Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol.*, 17, 118-126

Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E. & Filipowicz, W. (2005). Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science*, 309, 1573-157

Pillai, R.S., Artus, C.G. & Filipowicz, W. (2004). Tethering of human Ago proteins to mRNA mimics the miRNA-mediated repression of protein synthesis. *RNA*, 10, 1518-1525

Christiane Schaffitzel

PhD 2001, University of Zürich.

Postdoctoral research at the Institute of Molecular Biology and Biophysics, ETH Zürich.

Team leader at EMBL Grenoble since 2007. Habilitation, ETH Zürich, 2008.

Cryo-EM of ribosomal complexes

Previous and current research

How exactly the folding of a nascent polypeptide chain into its native structure is achieved is one of the fundamental questions in biology. Synthesis and folding of proteins require concerted interactions of the translating ribosome with translation factors, processing enzymes, molecular chaperones and factors involved in the export of proteins. The functional and structural characterisation of such complexes is the aim of our team.

We use cryo-EM and single particle analysis to study nascent chain co-translational folding, targeting and translocation. Structures of translating ribosomes in complex with the factors involved provide important mechanistic insight into the molecular mechanism of protein sorting.

A prerequisite for any structural studies on translating ribosomes is to produce large amounts of homogenous, stable complexes. To this end, we adapted and decisively improved existing *E. coli in vitro* translation protocols which are optimised for protein expression for the production of ribosome nascent chain complexes. We have solved the structures of a translating ribosome-signal recognition particle (SRP) complex and of a non-translating ribosome without nascent chain in complex with SRP. Conformational changes of SRP which occur when SRP recognises a signal sequence on the nascent chain emerging from the ribosomal exit tunnel explain previous biochemical findings such as experimental crosslinks, salt-sensitivity of the 70S-SRP complex and a more than 100-fold improved affinity of SRP towards ribosomes displaying a nascent chain with a signal sequence. Furthermore, the structure of a translating ribosome in complex with the *E. coli* translocation machinery at 15 Å resolution reveals that both active and inactive translocon are composed of dimers of SecYEG. Cryo-EM analysis of a ribosome in complex with the molecular chaperone Trigger Factor suggests that Trigger Factor forms a hydrophobic arch over the ribosomal exit tunnel, thereby providing a folding interface for the nascent polypeptide and protection against other cytosolic proteins to prevent aggregation and degradation of nascent proteins. Our findings in aggregate support a view of the ribosomal exit tunnel as an anchor point for a multitude of factors involved in co-translational processes. A concerted action of these factors ensures correct processing, folding and targeting of nascent proteins.

Future projects and goals

In the future, we will continue analysing ribosomal complexes involved in co-translational processes to gain structural insight into all steps along the pathway of co-translational targeting and translocation.

In collaboration with the Berger group (page 89), we study human macromolecular complexes involved in class II gene expression. We use EM to analyse the structure of these molecular machines, which we produce by means of advanced recombinant technologies.

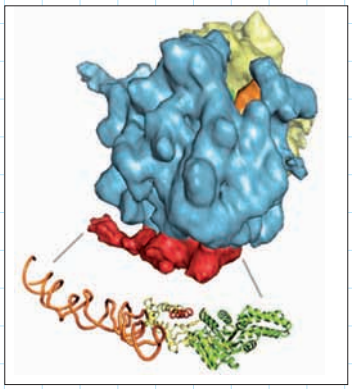


Figure 1: Cryo-EM structure of the E. coli *signal recognition particle bound to a translating ribosome. Ribosomal subunits are shown in yellow (30S) and blue (50S), tRNAs in green and orange, the SRP in red. Below the atomic model of* E. coli *SRP as fitted into the density is depicted.*

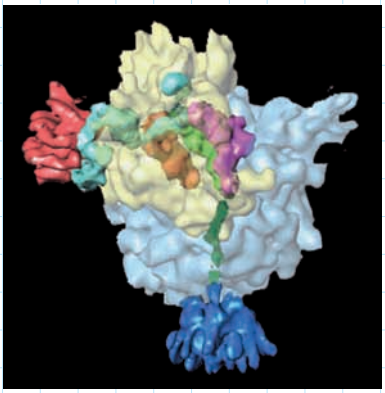


Figure 2: EM reconstruction of the ribosome nascent chain complex and the translocon. The 30S ribosomal subunit is shown in yellow, the 50S subunit in light blue, the A, P and E site tRNAs are in magenta, green and orange respectively. The mRNA is depicted in cyan and the nascent chain in the polypeptide exit tunnel in green. The active translocon in dark blue is located at the exit of the ribosomal tunnel. A second, non-translocating translocon (red) is bound to the mRNA.

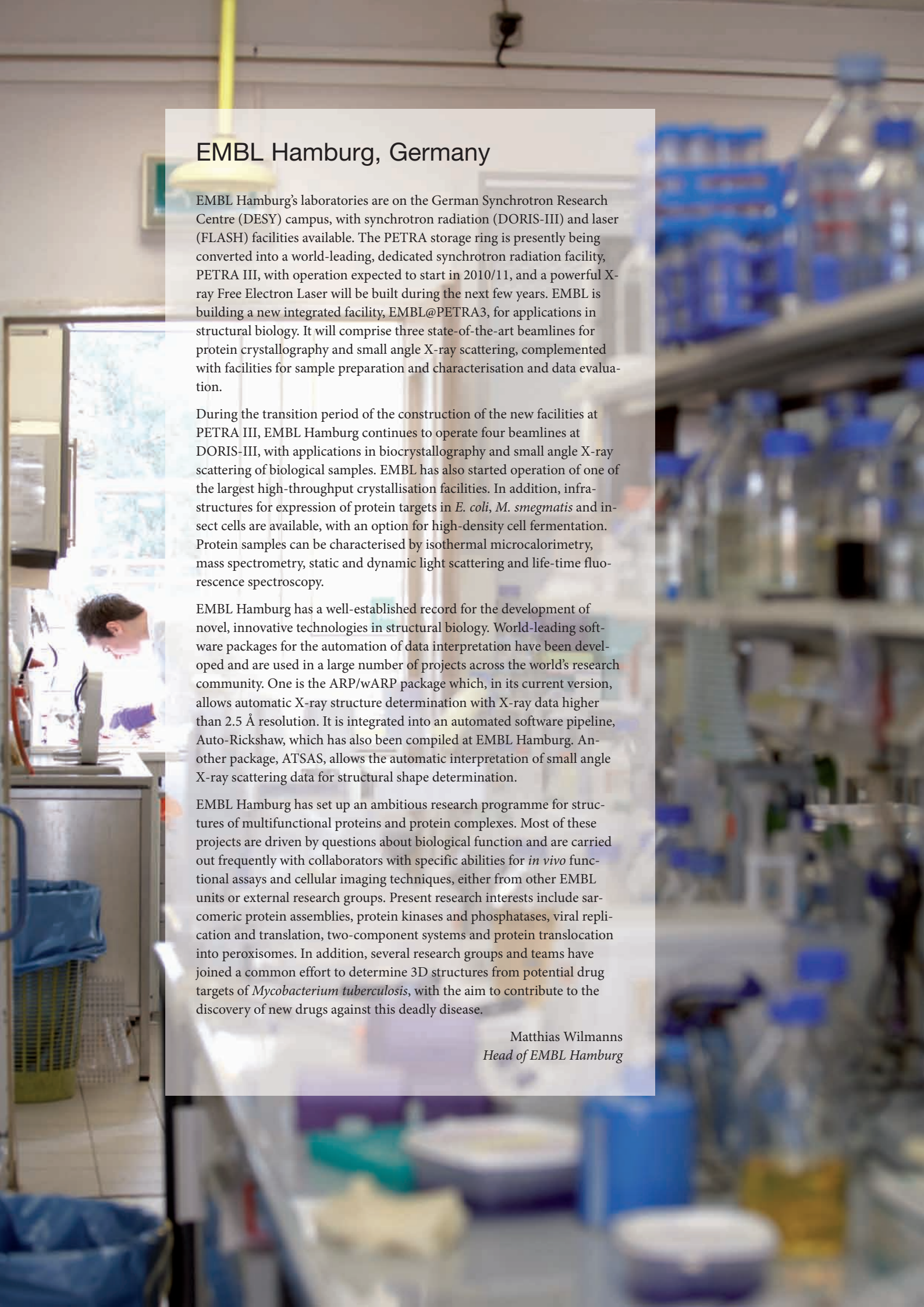
Selected references

Zhang, X., Schaffitzel, C., Ban, N. & Shan, S.O. (2009). Multiple conformational switches in a GTPase complex control co-translational protein targeting. *Proc Natl Acad Sci. USA*, 106, 1754-9

Bingel-Erlenmeyer, R., Kohler, R., Kramer, G., Sandikci, A., Antolic, S., Maier, T., Schaffitzel, C., Wiedmann, B., Bukau, B. & Ban, N. (2008). A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. *Nature*, 452, 108-11

Schaffitzel, C., Oswald, M., Berger, I., Ishikawa, T., Abrahams, J.P., Koerten, H.K., Koning, R.I. & Ban, N. (2006). Structure of the *E. coli* signal recognition particle bound to a translating ribosome. *Nature*, 444, 503-506

Mitra, K., Schaffitzel, C., Shaikh, T., Tama, F., Jenni, S., Brooks CL, 3rd, Ban, N. & Frank, J. (2005). Structure of the *E. coli* protein-conducting channel bound to a translating ribosome. *Nature*, 438, 318-324



EMBL Hamburg, Germany

EMBL Hamburg's laboratories are on the German Synchrotron Research Centre (DESY) campus, with synchrotron radiation (DORIS-III) and laser (FLASH) facilities available. The PETRA storage ring is presently being converted into a world-leading, dedicated synchrotron radiation facility, PETRA III, with operation expected to start in 2010/11, and a powerful X-ray Free Electron Laser will be built during the next few years. EMBL is building a new integrated facility, EMBL@PETRA3, for applications in structural biology. It will comprise three state-of-the-art beamlines for protein crystallography and small angle X-ray scattering, complemented with facilities for sample preparation and characterisation and data evaluation.

During the transition period of the construction of the new facilities at PETRA III, EMBL Hamburg continues to operate four beamlines at DORIS-III, with applications in biocrystallography and small angle X-ray scattering of biological samples. EMBL has also started operation of 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.

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.

EMBL Hamburg 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 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.

Postdoctoral research at the University of California, Los Angeles.

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 for distance measurements and *in vivo* imaging methods. We are interested in exploiting present and future opportunities, provided by synchrotron and laser facilities at DESY (DORIS-III, PETRA III, FLASH, X-FEL), to develop new methods and to apply them to biology-driven projects. Some of our specific interests are highlighted below (more information can be found at www.embl-hamburg.de/~wilmanns/home.html).

Protein-protein complexes of the muscle sarcomere, including titin, myomesin and binding partners. Titin is the largest gene product of the human genome, and it comprises 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. Titin is inter-connected to other long filament proteins, such as nebulin, myomesin, M-protein and obscurin. Recently, we have been able to determine the structures of the N-terminal assembly complex of titin in the Z-disk (Zou *et al.*, 2006) and the C-terminal assembly complex of myomesin in the M-band (Pinotsis *et al.*, 2008). In the latter complex, we have identified a novel type of helical linker, connecting neighbouring Ig domains. By making use of a series of additional high resolution structures of different parts of the C-terminal myomesin filament and small angle X-ray scattering data, we have been able to build a model of the complete filament (figure 1). Our future work will concentrate on protein-ligand complexes from sarcomeric filament proteins.

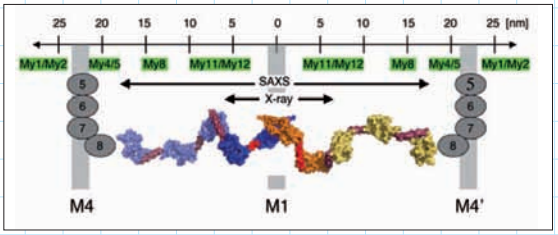


Figure 1: C-terminal myomesin model as M-band cross-linker. The shape of the C-terminal myomesin filament been fit into previous immuno-EM data. The approximate locations of the M1, M4, and M4' bands and some of the N-terminal myomesin domains are indicated schematically.

The architecture of the translocon of peroxisomes. Peroxisome are cell organelles that allow sequestered metabolic processes that would interfere with other processes that, for instance, take place in the cytosol. Those proteins that are involved in these processes are generally translocated as active and folded targets. We have been able, 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.

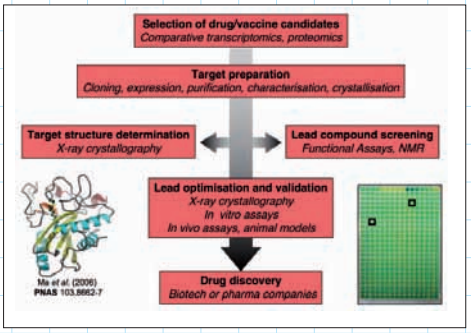


Figure 2: Work flow of the X-MTB structural proteomics project (www.xmtb.org).

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. For instance, we have 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.*, unpublished) 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.*, unpublished). 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. The workflow of our projects on *M. tuberculosis* targets is outlined in figure 2.

Selected references

Lamber, E.P., Vanhille, L., Textor, L.C., Kachalova, G.S., Sieweke, M.H. & Wilmanns, M. (2008). Regulation of the transcription factor Ets-1 by DNA-mediated homo-dimerization. *EMBO J.*, 27, 2006-2017

Pinotsis, N., Lange, S., Perriard, J.C., Svergun, D.I. & Wilmanns, M. (2008). Molecular basis of the C-terminal tail-to-tail assembly of the sarcomeric filament protein myomesin. *EMBO J.*, 27, 253-264

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

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

Instrumentation for synchrotron beamlines for structural biology at PETRA III

Stefan Fiedler

PhD 1997, Johann-Wolfgang-Goethe-Universität, Frankfurt.

Postdoctoral fellow then Staff Scientist at ESRF, Grenoble.

At EMBL Hamburg since 2004. Team leader since 2006.

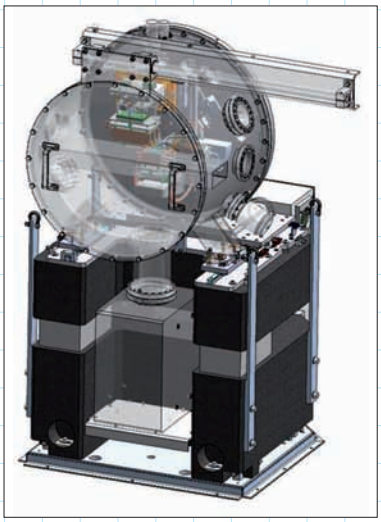
Previous and current research

EMBL will design, build and operate three beamlines for structural biology at the PETRA III synchrotron radiation source, currently under construction on the DESY campus. The EMBL facilities will be dedicated to the leading techniques for X-ray-based structural research of biological samples: small angle X-ray scattering (BioSAXS beamline) and macromolecular crystallography (MX1 and MX2 beamlines). The construction of these facilities includes the provision of X-ray optical elements, experimental endstations, vacuum system, control system, data acquisition system, technical infrastructure and parts of the civil engineering, and our team provides expertise in X-ray optics, precision mechanical engineering, robotics, control software and electronics.

In 2008, several projects for the principal optical elements at the new PETRA III beamlines have been launched. In particular, the design phase of the high-heatload monochromators and cooling systems has started, in collaboration with an external company (FMB OXFORD) and HASYLAB. In this and most of the other projects, our work focusses on the selection, customisation and integration of mechanics, control electronics and control software. In addition, our activity spectrum ranges from the development of instruments from scratch to the performance of relevant validation experiments and commissioning of instrumentation for user experiments. An example of this is the construction of a focussing double multilayer monochromator (MLM) for the existing BW7A beamline (at the DORIS storage ring) which serves as test platform for developments for the PETRA III beamlines (see also Hermes group, page 100). Very high resolution structure refinements and first *ab initio* structure solution have been possible with the data collected on the new instrument which is now in user operation.

The main development project in 2008 was the construction of a completely updated version of a robotic sample mounting system for macromolecular crystals, MARVIN (see figure), characterised by increased capacity, higher sample mounting speed, improved geometry and maintainability with respect to the previous design.

As a side project, a new high precision goniometer axis with sub-micrometer precision has been developed. These elements are now integrated into a software-based modern control system which allows for a heterogeneous control environment and provides distributed access and monitoring. This is a prerequisite for the planned remote-controlled user experiments at PETRA III.



Schematical design of experimental endstation for MX with sample mounting system MARVIN.

Future projects and goals

In the next year our efforts will concentrate on:

- installation and commissioning of the new beamline elements;
- integration of control hardware and software into a generic control architecture;
- further automation of alignment, sample handling and data acquisition;
- rapid feedback on positional and intensity variations of the incident beam;
- follow-up or start of a limited number of selected instrumental development projects (e.g. robotic sample changer for MX, beam separation optics, beam conditioning unit);
- continuation of our investigations in MX with multilayer radiation.

In the longer term, time-resolved structural biology studies down to the microsecond will be performed on the new beamlines. We plan to develop the instruments and the level of synchronisation that are necessary for these kinds of experiments. Furthermore, we will also explore the possibilities and experimental requirements for biological imaging research at PETRA III.

Selected references

Fernandez, M., Suhonen, H., Keyrilainen, J., Bravin, A., Fiedler, S., Karjalainen-Lindsberg, M.L., Leidenius, M., von Smitten, K. & Suortti, P. (2008). USAXS and SAXS from cancer-bearing breast tissue samples. *Eur. J. Radiol*., 68, S89-94

Keyrilainen, J., Fernandez, M., Karjalainen-Lindsberg, M.L., Virkkunen, P., Leidenius, M., von Smitten, K., Sipila, P., Fiedler, S., Suhonen, H., Suortti, P. & Bravin, A. (2008). Toward high-contrast breast CT at low radiation dose. *Radiology*, 249, 321-7

Kelly, M.E.B., Schültke, E., Fiedler, S., Nemoz, C., Guzmann, R. et al. (2007). Synchrotron-based intravenous cerebral angiography in a small animal model. *Physics in Medicine and Biology*, 52, 1001-1012

Roessle, M.W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D. et al. (2007). Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *Journal of Applied Crystallography*, 40, S190-S194

Christoph Hermes

PhD 1981, Technical University, Munich.

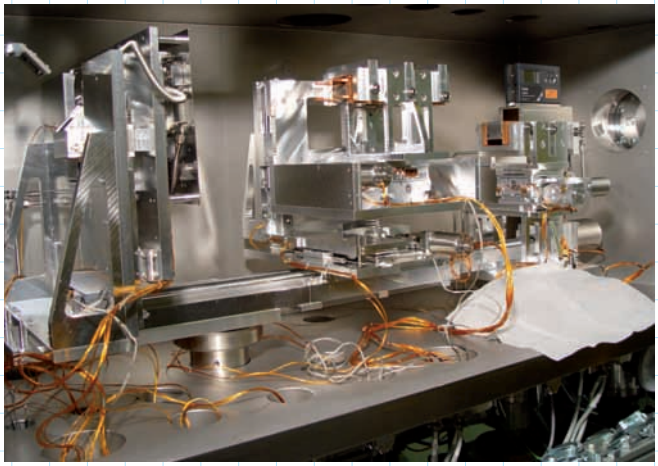
At EMBL Hamburg since 1981. Group leader since 1988.

Instrumentation for structural research at EMBL Hamburg

Previous and current research

Research carried out at the Hamburg outstation since 1974 has covered a broad spectrum of X-ray methods used for structural investigations in biology. Small angle solution scattering (SAXS) allows the study of 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, data acquisition and control electronics. During 2007 a Multilayer Monochromator (ML) system was designed, built, installed and commissioned on wiggler beamline BW7A which can be used alternatively to the standard optical set-up of this branch of the BW7 wiggler comprising a focussing Si(111) double crystal monochromator (DCM) for MAD data collection on protein crystals. The ML mode of operation was used very successfully in 2008 for PX experiments. This was mainly due to the considerable gain in intensity, allowing very rapid data collection and hence collecting data from a large number of crystals per shift. In these measurements the newly developed beamline control system which is based on economic industrial electronics and improved software proved to be reliable and user-friendly. In preparation and as test cases for the new beamlines on the worldwide unique radiation source PETRA III, we are in the process of optimising the end-stations of the 'old' beamlines at the storage ring. We plan to install for example improved rotation axes, centring devices and automatic sample changers to obtain results which will give us confidence for future applications at PETRA.



The multilayer monochromator in its vacuum vessel during installation at beamline BW7A. From right to left (beam direction), first and second multilayer located on their respective positioning devices and the slit system with integrated intensity monitor are shown.

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 high-throughput data collection.

EMBL is building and will operate three beamlines on PETRA III. In this context, major challenges and opportunities in the field of beamline instrumentation, sample handling, control electronics and software will have to be mastered (see the Fiedler group, page 99).

The opening of the European X-Ray Laser (XFEL) in Hamburg is scheduled for 2014, 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.

Selected references

Roessle, M.W., Klaering, R., Ristau, U., Robrahn, B., Jahn, D., Gehrmann, T., Konarev, P., Round, A., Fiedler, S., Hermes, C. & Svergun, D. (2007). Upgrade of the small-angle X-ray scattering beamline X33 at the European Molecular Biology Laboratory, Hamburg. *Journal of Applied Crystallography*, 40, S190-S194

Hermes, C. Gehrmann, T., Jahn, D., Ristau U., Robrahn, B. & Siambanis, T. (2004). (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

Pohl, E., Gonzalez, A., Hermes, C. & van Silfhout, R.G. (2001). Overview of the tunable beamlines for protein crystallography at the EMBL Hamburg Outstation; an analysis of current and future usage and developments. *J. Synchrotron Radiat.*, 8, 1113-1120

Macromolecular crystallography

Victor S. Lamzin

PhD 1987, Moscow State University.

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 since 1997.

Accurate modelling of biological processes depends on an intimate knowledge of the macromolecules involved. The availability of models of the interacting components allows a deep understanding of the basic processes that underlie human disease and all other cellular events. One recent example is a basis for neurodegenerative pathologies provided by the structure of a stable serpin dimer (Yamasaki *et al.*, 2008, *Nature*).

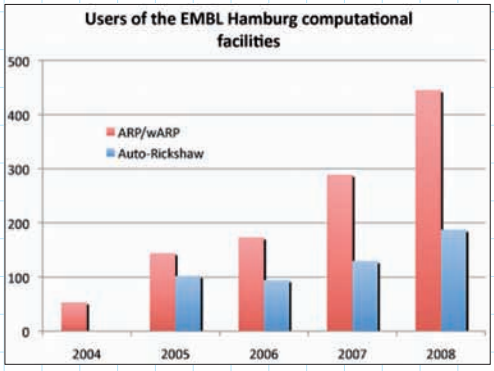
Macromolecular structures are produced by a large number of independent scientists, all of whom approach specific problems in their own manner. Methods are needed to establish objective structure determination and quality assessment, which will impact further system-wide analysis. The group's major activity is development of novel approaches and software tools for high-throughput 3D structure determination and detailed interpretation of biological macromolecules and their complexes. Specific attention is given to extension of the current limits of applicability of MX and its use with complementary techniques.

Previous and current research

In MX, availability of comprehensive software packages has a major impact on structural biology. Traditionally crystallographic model building is done by expert users with the aid of specialised graphics software. The automation of this process, first exemplified in the ARP/wARP package (Langer *et al.*, 2008, *Nature Protocols*), was promptly followed by developments worldwide. ARP/wARP has been used extensively for thousands of structure determinations and is often used as a benchmark to evaluate the quality of electron density obtained by new methods; it has been integrated into many crystallographic software pipelines as the default model-building engine. A wide spectrum of ARP/wARP functionalities (e.g. Hattne *et al.*, 2008, *Acta Cryst.*) makes the software particularly attractive to users.

We continue the developments of the automated structure determination pipeline AutoRickshaw (led by S. Panjikar in collaboration with the Tucker group and Weiss team (pages 106 and 107)), which uses an artificial intelligence decision-making system (Manjasetty *et al.*, 2008, *Proteomics*). With the help of ARP/wARP and other software, AutoRickshaw produces an interpretable electron density map and a partial structure shortly after data acquisition. At our MX beamlines the users receive immediate feedback whether the measured data are of sufficient quality for successful structure determination.

Our group has initiated the installation of the EMBL Hamburg advanced computational facilities, where the in-house, beamline and external users can carry out 3D structure determination remotely. These services for AutoRickshaw and particularly ARP/wARP have become increasingly popular (see figure).



User numbers of the ARP/wARP and AutoRickshaw remote computation facilities are increasing rapidly. In 2008 the facilities were used by more than 600 people, and over 1,000 macromolecular structures were solved.

At the other end of the scale, we continue pushing the limits of model interpretation at ultra-high resolution. Using quantum chemical calculations we were able to decipher subtle changes in the active site residues upon substrate binding in Hydroxynitrile lyase (Schmidt *et al.*, 2008, *J. Biol. Chem.*). We showed that His235 of the catalytic triad must be protonated in order for catalysis to proceed. We also detected considerable pKa shifts that had been hypothesised earlier.

Future projects and goals

In the future we will witness the further integration of MX with complementary techniques, and the underlying methodology will face a number of challenges. Crystalline samples are already obtained from increasingly complex biological systems and are becoming smaller and more difficult to handle, with data harder to extract. The group's activities will cast within the anticipated trend and aim at furthering the understanding of molecular biological processes by increasing the amount of 3D information that can be obtained from biological macromolecules.

Selected references

Yamasaki, M., Li, W., Johnson, D.J. & Huntington, J.A. (2008). Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature,* 455, 1255-8

Hattne, J. & Lamzin, V.S. (2008). Pattern-recognition-based detection of planar objects in three-dimensional electron-density maps. *Acta Crystallogr. D Biol. Crystallogr.*, 64, 834-842

Langer, G., Cohen, S.X., Lamzin, V.S. & Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.*, 3, 1171-1179

Manjasetty, B.A., Turnbull, A.P., Panjikar, S., Bussow, K. & Chance, M.R. (2008). Automated technologies and novel techniques to accelerate protein crystallography for structural genomics. *Proteomics*, 8, 612-625

Schmidt, A., Gruber, K., Kratky, C. & Lamzin, V.S. (2008). Atomic resolution crystal structures and quantum chemistry meet to reveal subtleties of hydroxynitrile lyase catalysis. *J. Biol. Chem.*, 283, 21827-21836

Wolfram Meyer-Klaucke

PhD 1996, Hamburg University.

Postdoctoral research at the Medical University of Lübeck.

At EMBL Hamburg since 1998. Team leader since 2000.

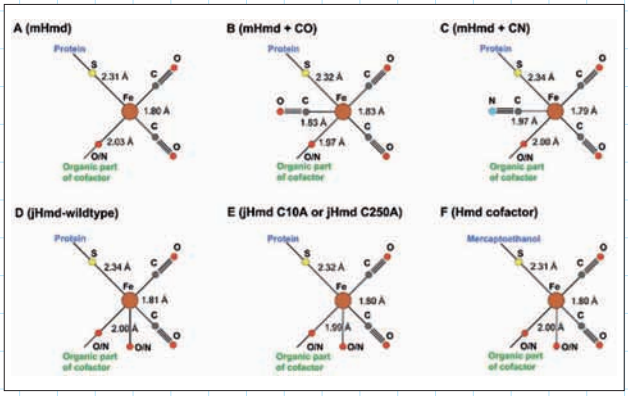
Trace elements in biological systems

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; Wellenreuther, 2007), 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 (Kostelecky 2006). Their physiological importance varies from putative association with cancer and antibiotic resistance to different roles in cellular detoxification.

Metal regulation. Members of the ferric/zinc uptake regulator (Fur/Zur) family are the central metal-dependent regulator proteins in many Gram-negative and -positive bacteria. They are responsible for the control of a wide variety of basic physiological processes and the expression of important virulence factors in human pathogens. Therefore, Fur has gathered significant interest as a potential target for novel antibiotics. Recently, we solved the crystal structure of FurB from *Mycobacterium tuberculosis*, which together with biochemical and spectroscopic data allowed us to propose the functional role of this protein. Although the overall fold of FurB with an N-terminal DNA binding domain and a C-terminal dimerisation domain is conserved among the Zur/Fur family, large differences in the spatial arrangement of the two domains with respect to each other can be observed. The biochemical and spectroscopic analysis revealed that *M. tuberculosis* FurB is Zn(II)-dependent and is likely to control genes involved in the bacterial zinc uptake. The combination of the structural, spectroscopic and biochemical results enabled us to determine the structural basis for functional differences in this important family of bacterial regulators (Lucarelli, 2007).



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

New metal binding motifs. Hydrogenases are enzymes that catalyse 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 H2 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 sulphur, and two nitrogen/oxygen atoms coordinated to the iron, the sulphur 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 H2 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), but spatial resolved spectroscopy and advanced spectroscopic techniques will play an increasing role (Wellenreuther *et al.*, 2009).

Selected references

Wellenreuther, G., Cianci, M., Tucoulou, R., Meyer-Klaucke, W. & Haase, H. (2009). The ligand environment of zinc stored in vesicles. *Biochem. Biophys. Res. Commun.*, 380, 198-203

Küpper, F.C., Carpenter, L.J., McFiggans, G.B., Palmer, C.J., Waite, T.J., Boneberg, E.M. *et al.* (2008). Iodide accumulation provides kelp with an inorganic antioxidant impacting atmospheric chemistry. *Proc. Natl Acad. Sci. USA*, 105, 6954-8

Shima, S., Pilak, O., Vogt, S., Schick, M., Stagni, M.S., Meyer-Klaucke, W., Warkentin, E., Thauer, R.K. & Ermler, U. (2008). The crystal structure of [Fe]-hydrogenase reveals the geometry of the active site. *Science*, 321, 572-575

Lucarelli, D., Russo, S., Garman, E., Milano, A., Meyer-Klaucke, W. & Pohl, E. (2007). Crystal structure and function of the zinc uptake regulator FurB from *Mycobacterium tuberculosis*. *J. Biol. Chem.*, 282, 9914-9922

Disease-related proteins and the high-throughput crystallisation facility

Jochen Müller-Dieckmann

PhD 1994, Albert-Ludwigs-Universität, Freiburg.

Postdoctoral research at the University of California, Berkeley.

Associate Director, SGX, San Diego, until 2004.

Team leader at EMBL Hamburg since 2004.

Previous and current research

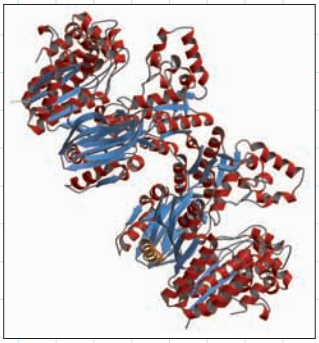
Disease-related proteins from *Mycobacterium tuberculosis* and *Schistosoma mansoni*: Tuberculosis remains the single most infectious disease, killing over 2 million individuals every year. The disease is curable; treatment, however, is prolonged and requires the combination of several drugs. At the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

We have focussed on targets 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. The picture below shows the structure of iso-citrate dehydrogenase II (Rv0066c). This rare dehydrogenase displays unexpected oligomeric characteristics and we are currently in the process to determine its functional significance.

The trematode *S. mansoni*, a parasitic worm, is the causing agent of schistosomiasis. Also known as bilharzia or snail fever, it is second only to malaria in socio-economic and public health importance. Infection with *S. mansoni* triggers a specific reaction by the immune system and we have recently determined the structure of this novel immune modulator.

High-throughput crystallisation: A complete understanding of the function of biological macromolecules requires knowledge of their 3D structures. X-ray crystallography has emerged as the method of choice in structural biology due to its ability to resolve structures at atomic resolution without size limitation. Current bottlenecks of this technique are the production of pure and soluble sample material and the necessity to crystallise it. As a consequence, both aspects of macromolecular crystallography have been automated.

The high-throughput crystallisation facility in Hamburg, which is open to the general user community, has been in operation since 2005. The automated set up of initial screens with vapour diffusion or free interface diffusion experiments has resulted in over 1,000,000 individual tests over the last years. In 2008, we have also developed and implemented procedures for the automated set-up of optimisation experiments. More information is available at www.embl-hamburg.de/services/crystallisation.



Ribbon representation of Rv0066c (Icd2) of M. tuberculosis. *Shown is a dimer of Icd2 with a two-fold axis in the image plane. α-helices are in red and β-strands are in blue. Each monomer consists of 745 amino acids. Substrate and NADPH are bound in the large cleft between domains.*

Ethylene perception in plants: Ethylene is gaseous hormone in plants which regulates a multitude of processes, ranging from seed germination and fruit ripening to leaf abscission and organ senescence. Signal transduction involving ethylene is initiated by five receptors in *Arabidopsis thaliana*. Their domain structure is related to bacterial two-component systems. In plants, however, the immediate downstream target of this system is CTR1, a Raf-like Ser/Thr protein kinase which initiates a MAP kinase-like cascade.

The mechanism of signal transduction of eukaryotic two-component systems remains unclear. We have produced a variety of constructs of the different signal receptors as well as of the cognate protein kinase. These domains and their complexes are analysed by X-ray crystallography and X-ray small angle scattering with the goal to functionally characterise the early steps of ethylene signalling through inter- and cross-domain activation of its components.

Future projects and goals

The importance of construct design on protein expression, solubility and crystallisability has been widely accepted. In collaboration with Darren Hart in Grenoble (page 91) we will evaluate the possibility to rapidly screen a large number (~40) of protein constructs and to identify those which are more amenable to crystallisation using the Fluidigm technology. For a proof of principle study we used 7 different constructs of NFκB which were identified with ESPRIT. All 7 constructs were expressed and purified in a one step procedure in Hamburg and subject to crystallisation with Fluidigm's Topaz chips. Small crystals were obtained under conditions that are very similar to published conditions for one construct. We are now planning to demonstrate the general applicability of this process by using *de novo* protein samples. This project is funded under the EU's FP7 initiative.

Selected references

Hatzopoulos, G.N., Kefala, G. & Mueller-Dieckmann, J. (2008). Cloning, expression, purification, crystallization and preliminary X-ray crystallographic analysis of isocitrate dehydrogenase 2 (Rv0066c) from *Mycobacterium tuberculosis*. *Acta Crystallogr. Sect F Struct. Biol. Cryst. Commun.*, 64, 1139-1142

Groves, M.R., Muller, I.B., Kreplin, X. & Muller-Dieckmann, J. (2007). A method for the general identification of protein crystals in crystallization experiments using a noncovalent fluorescent dye. *Acta Crystallogr. D Biol. Crystallogr.*, 63, 526-535

Thomas Schneider

PhD 1996, Technical University of Munich and EMBL.

Postdoctoral research at the Max-Planck-Institute for Molecular Physiology, Dortmund, and the University of Göttingen.

Group leader at the FIRC Institute of Molecular Oncology, Milan.

Group leader at EMBL since 2007. Coordinator of the EMBL@PETRA3 project.

Tools for structure determination and analysis

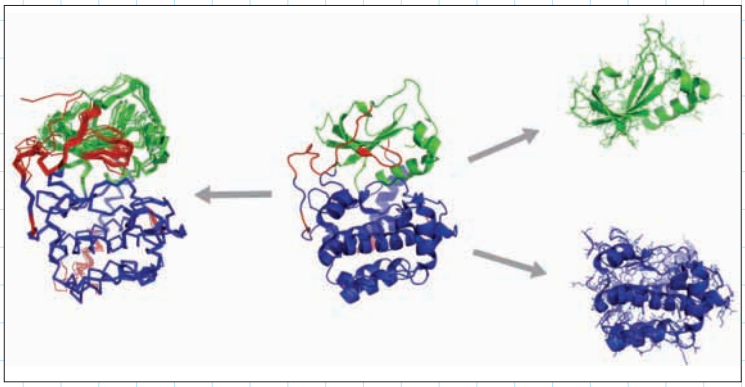
Previous and current research

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 structural data.

The three beamlines we are constructing 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 construction of the beamlines is done in close collaboration with Stefan Fiedler's team (page 99).

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 developing 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.



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).

Future projects and goals

For the integrated facility for structural biology, our goal is to provide beamlines that are ready for user experiments by 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-focussing 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.

Selected references

Mosca, R. & Schneider, T.R. (2008). RAPIDO: a web server for the alignment of protein structures in the presence of conformational changes. *Nucleic Acids Res.*, 36, W42-W46

Schneider, T.R. (2008). Synchrotron radiation: micrometer-sized X-ray beams as fine tools for macromolecular crystallography. *HFSP Journal*, Vol. 2, No. 6, pp. 302–306.

Penengo, L., Mapelli, M., Murachelli, A.G., Confalonieri, S., Magri, L., Musacchio, A., Di Fiore, P.P., Polo, S. & Schneider, T.R. (2006). Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. *Cell*, 124, 1183-1195

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

SAXS studies of biological macromolecules in solution

Dmitri Svergun

PhD 1982, Institute of Crystallography, Moscow.

Dr. of Science 1997, Institute of Crystallography, Moscow.

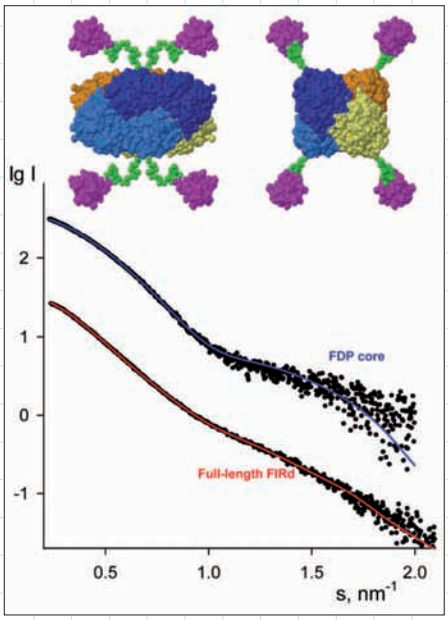
At EMBL since 1991. Group leader since 2003.

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, it elucidates structural changes in response to variations in external conditions, protein-protein and protein-ligand interactions, and time-resolved studies permit one to characterise structural kinetics of assembly/dissociation or folding/unfolding.

Our group runs beamline X33, dedicated to biological solution SAXS, at DESY's storage ring, DORIS-III. We develop novel methods to construct 3D models of individual macromolecules and their complexes from the X-ray and neutron scattering data with advanced mathematical and bioinformatical approaches. These methods are extensively used to interpret the experimental data in collaborative user projects at X33. The rapidly-growing demand for SAXS in the biological community has led to a dramatic increase in the user turnover at X33 (over five times more than in 2000). The beamline was completely refurbished in 2004-2008, including automation of the experiment and data analysis procedures. The novel analysis methods and improved experimental facilities allow us to solve exciting biological problems in collaboration with the users of X33. Here, SAXS is used to study quaternary and domain structure of individual proteins, nucleic acids and their complexes, oligomeric mixtures, conformational transitions upon ligand binding, flexible systems and intrinsically unfolded proteins, processes of amyloid fibrillation and many other objects of high biological and medical importance (see figure for a recent example).



Structural organisation of Flavorubredoxin tetramer reconstructed from the SAXS patterns (intensity versus scattering angle; dots: experimental data, solid lines: computed from the models). Solution scattering study showed that: (i) the tetrameric flavodiiron (FDP) core (depicted in blue and yellow/orange) is more anisometric than the crystallographic models of the homologous FDPs; (ii) Rubredoxin (Rd) domains (magenta) are located on the periphery, being loosely connected to the core by extended linkers (green) and are freely available to participate in redox reactions with protein partners.

Future projects and goals

The present and future work of the group includes:

- participation in numerous collaborative projects at X33 beamline employing SAXS to study the structure of a wide range of biological systems in solution;
- development of novel methods and approaches for the reconstruction of tertiary and quaternary structure of macromolecules and their complexes from X-ray and neutron scattering data;
- joint use of SAXS with other structural, biophysical and computational methods including neutron scattering, crystallography, NMR, electron microscopy, FRET, bioinformatics, etc;
- maintenance and upgrade of the X33 beamline including automation of SAXS experiments and their analysis;
- collaboration with the PETRA III group at EMBL Hamburg (opposite) in designing a new high-brilliance biological SAXS beamline at the third-generation PETRA storage ring at DESY.

Selected references

Petoukhov, M. V., Vicente, J. B., Crowley, P. B., Carrondo, M. A., Teixeira, M., and Svergun, D. I. (2008). Quaternary structure of flavorubredoxin as revealed by synchrotron radiation small-angle X-ray scattering. *Structure,* 16, 1428-1436

Shukla, A., Mylonas, E., Di Cola, E., Finet, S., Timmins, P., Narayanan, T. & Svergun, D.I. (2008). Absence of equilibrium cluster phase in concentrated lysozyme solutions. *Proc. Natl Acad. Sci. USA*, 105, 5075-5080

She, M., Decker, C.J., Svergun, D.I., Round, A., Chen, N., Muhlrad, D., Parker, R. & Song, H. (2008). Structural basis of dcp2 recognition and activation by dcp1. *Mol. Cell*, 29, 337-349

Xu, X. F., Reinle, W. G., Hannemann, F., Konarev, P. V., Svergun, D. I., Bernhardt, R. & Ubbink, M. (2008). Dynamics in a pure encounter complex of two proteins studied by solution scattering and paramagnetic NMR spectroscopy. *J. Am. Chem. Soc*, 130, 6395-6403

Paul Tucker

PhD 1972, University of Essex, Colchester.

Postdoctoral research at the universities of Groningen, Leicester, the Australian National University in Canberra, and Exeter.

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) continues to pose a global health threat and is our organism of choice because understanding the way in which the bacteria responds to externally generated stress is a possible way to the discovery of novel antibiotics. In the last years we have extended our work to investigate the structure of proteins involved in regulating the production of the stress response sigma factor, σ^F. The most interesting results have been on a protein that contains four regulatory domains, a sensor PAS domain, a kinase domain (the anti-sigma factor), a phosphatase domain and an anti-sigma factor antagonist domain. In this system we are beginning to understand the structural basis of the regulation mechanism, and, from the structural work, have some idea as to what signal is used to switch on σ^F dependent genes. Two structures are shown in figure 1.

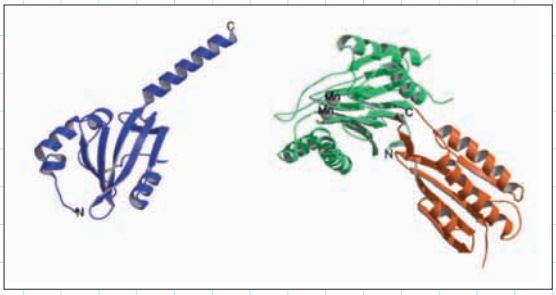


Figure 1: The structures of the PAS (sensor) domain (blue), the kinase (red) and phosphatase (green) domains of the Mycobacterium tuberculosis *regulatory protein Rv1364c.*

A second area of study has been on proteins that are involved in replication of the genome of a variety of RNA viruses. Our goal is still to obtain structural information on the RNA dependent RNA polymerase domains of the caliciviridae, the flaviviridae and, more importantly, 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 (figure 2).

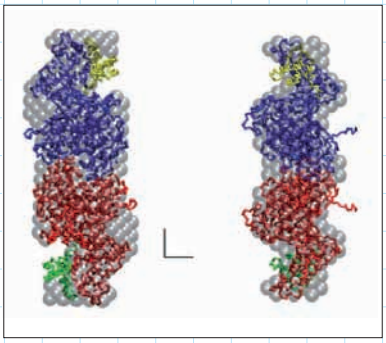


Figure 2: The solution structure (represented by grey spheres) from small-angle X-ray scattering of a dimer of the Epstein-Barr virus single stranded DNA binding protein BalF2. Superimposed is a homology model derived from the equivalent Herpes Simplex Virus 1 protein, ICP8.

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

We will continue our work on the unusual actinobacterial two-component system (PdtaR/PdtaS) we discovered a few years ago, by determining the structure of the sensor domain of the (cytosolic) histidine kinase. We will continue our structural work on domains of the coronavirus NSP3 protein, as part of a comparative analysis of these domains and their organisation in the different classes of the coronaviridae.

We are expecting to push forward our structural work on lipid binding proteins from nematodes. There is a complete lack of structural information on this class of proteins and, for parasitic nematodes, they are essential for the viability of the organism in the host.

Selected references

King-Scott, J., Nowak, E., Mylonas, E., Panjikar, S., Roessle, M., Svergun, D.I. & Tucker, P.A. (2007). The structure of a full-length response regulator from *Mycobacterium tuberculosis* in a stabilised 3D domain-swapped, activated state. *J. Biol. Chem.*, 282, 37717-29

Stockigt, J. & Panjikar, S. (2007). Structural biology in plant natural product biosynthesis – architecture of enzymes from monoterpenoid indole and tropane alkaloid biosynthesis. *Natural Product Reports*, 24, 1382-1400

Tucker, P.A., Jens, E.N. & Morth, P. (2007). Two-component systems of *M. tuberculosis*: structure-based approaches. *Methods Enzymol.*, 423, 479-501

Tucker, P.A. & Sallai, L. (2007). The AAA+ superfamily-a myriad of motions. *Curr. Opin. Struct. Biol.*, 17, 641-652

X-ray crystallography of biological macromolecules

Manfred S. Weiss

PhD 1992, Albert-Ludwigs-Universität, Freiburg.

Postdoctoral research at the University of California, Los Angeles.

Senior Research Assistant at the IMB, Jena.

Team leader at EMBL Hamburg since 2001.

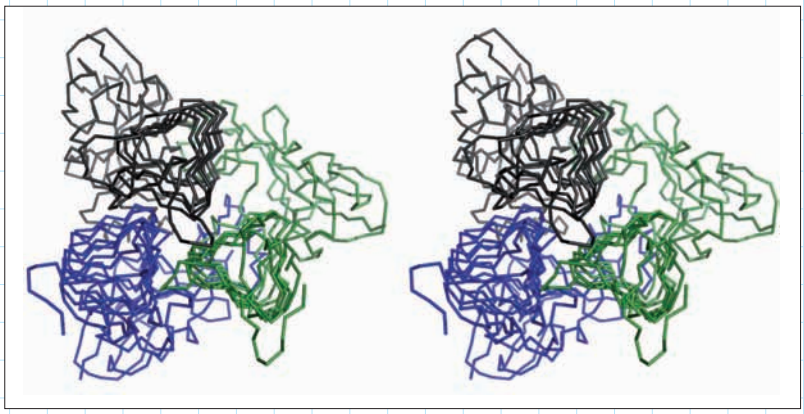
Previous and current research

Structural analysis of proteins from *Mycobacterium tuberculosis*. Tuberculosis, which is typically caused by *Mycobacterium tuberculosis* (Mtb) is a major human health threat. In order to identify and characterise potential new targets for intervention strategies, the focus of the work in our team 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 eight enzymes from Mtb: two from the leucine biosynthesis and six from the lysine biosynthesis pathway. As an example, the trimeric structure of tetrahydrodipicolinate N-succinyltransferase (Rv1201c; DapD) 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 Mtb. These structures will also provide the basis for the design of new drugs. In addition, we have determined the structure of two hypothetical proteins from *M. tuberculosis*, the function of which are currently under investigation.

Development of new methods for macromolecular structure determination. In collaboration with Paul Tucker's group (opposite) and Christoph Müller-Dieckmann (ESRF, France), 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. Furthermore, the use of anomalous differences originating from light atoms in proteins in molecular replacement, refinement and substructure definition is studied in our group.

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.



Stereo Cα-trace of trimeric DapD (Rv1201c) from Mtb. The three subunits are coloured in dark grey, blue and green, respectively. The most striking feature of the structure is a three stranded, left handed, parallel β-helix, which is responsible for binding of succinyl-CoA.

Selected references

Kefala, G., Evans, G.L., Griffin, M.D., Devenish, S.R., Pearce, F.G., Perugini, M.A., Gerrard, J.A., Weiss, M.S. & Dobson, R.C. (2008). Crystal structure and kinetic study of dihydrodipicolinate synthase from *Mycobacterium tuberculosis*. *Biochem. J.*, 411, 351-360

Mueller-Dieckmann, C., Panjikar, S., Schmidt, A., Mueller, S., Kuper, J., Geerlof, A., Wilmanns, M., Singh, R.K., Tucker, P.A. & Weiss, M.S. (2007). On the routine use of soft X-rays in macromolecular crystallography. Part IV. Efficient determination of anomalous substructures in biomacromolecules using longer X-ray wavelengths. *Acta Crystallogr. D Biol. Crystallogr.*, 63, 366-380

Weyand, S., Kefala, G. & Weiss, M.S. (2007). The three-dimensional structure of N-succinyldiaminopimelate aminotransferase from *Mycobacterium tuberculosis*. *J. Mol. Biol.*, 367, 825-838

Mueller-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

EMBL Monterotondo, Italy

The EMBL Monterotondo outstation is 20km north of Rome, in a green riverside campus that it shares with Italian national research groups (IBC-CNR) and the headquarters of the European Mouse Mutant Archive (EMMA). Central Rome is a short train ride away, and the surrounding Lazio countryside features medieval hill towns, spectacular mountain terrain and lakes for hiking, biking, skiing and swimming in the delightful Mediterranean climate.

EMBL Monterotondo houses the Mouse Biology Unit, where research groups use the powerful tools of mouse functional genomics and advanced genetic manipulation to investigate wide ranging aspects of mammalian biology including development and differentiation, cancer and regeneration, behaviour and memory. The collective work of PhD students, postdoctoral fellows and staff in these groups has established EMBL Monterotondo as a hub for international mouse research. Dynamic partnerships and exchanges with other international academic research and clinical centres, and participation in multiple EU-wide mouse research and informatics initiatives, serve to disseminate our discoveries in genetics/genomics, cell biology and pathophysiology. Amongst the groups, collaborative studies focus on developmental neurobiology, genetic and pharmacologic manipulations of adult behaviour, inflammation and regenerative processes, stem cell biology, microRNA-mediated control of gene expression, mechanisms of pain perception and biomedical applications.

Excellent core facilities are at the heart of EMBL research culture. A state-of-the-art animal facility provides a full range of gene recombineering, mouse transgenic and gene knockout production, embryo rederivation and cryopreservation services, and a fully-equipped phenotyping suite. The continued refinement of genetic and epigenetic perturbations through the conditional and inducible mouse mutations at the outstation is generating ever more accurate models of human disease and multigenic disorders, providing a wide array of technologies to the scientific community. Other centralised core facilities include histology, confocal microscopy, flow cytometry and monoclonal antibody production. Interaction with groups at other EMBL campuses through the Interdisciplinary Postdoc (EIPOD) fellowships provides exciting new prospects for joint projects. Partnerships established with Imperial College London and its associated hospitals and with the Centre for Genomic Regulation (CRG) in Barcelona add additional opportunities for translational research.

Training activities at EMBL Monterotondo focus on mouse genetic manipulation in collaboration with local faculty at CNR, EMMA and the Jackson Laboratory. A popular initiative organised by the European Learning Laboratory for the Life Sciences (ELLS) to train international and Italian secondary teachers in current biological themes reflects EMBL's longstanding tradition of transmitting new information and expertise to local society. An international 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

Regenerative mechanisms in heart and skeletal muscle

Nadia Rosenthal

PhD 1981, Harvard Medical School.

Postdoctoral research at the National Cancer Institute.

Assistant Professor, Boston University Medical Center. Associate Professor, Mass. General Hospital, Harvard Medical School.

Group leader and Head of EMBL Monterotondo since 2001.

Previous and current research

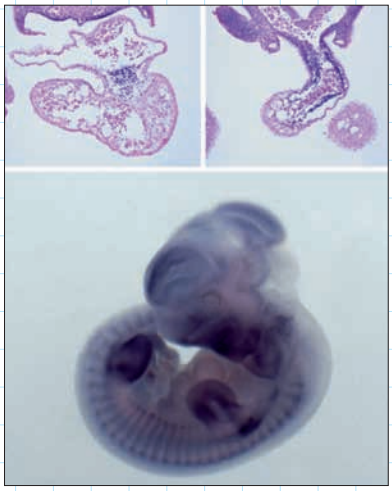
Our laboratory focuses on regenerative biology, which explores the processes that restore the architecture of damaged or degenerating tissues, often by recapitulating original embryonic development. We aim to reduce the impediments to effective regeneration by recapturing the remarkable regenerative capacity of lower vertebrates. Using the mouse to define the mechanisms involved in the mammalian response to injury, disease and ageing, we are identifying and modulating key signalling pathways that induce the recruitment of progenitor cells to sites of tissue damage and augment local repair mechanisms.

We found that insulin-like growth factors attenuate muscle atrophy and improve repair in ageing, muscular dystrophy and cardiomyopathies. Delivery of an unprocessed IGF-1 isoform (mIGF-1) to various neuromuscular pathologies implicate this growth factor as a powerful enhancer of the regeneration response. Selective muscle fibre loss and fibrosis in ageing and diseased skeletal muscle can be blocked by transgenic or viral delivery of mIGF-1, which augments local repair mechanisms and promotes recruitment of stem cells to sites of injury. Supplemental mIGF-1 expression reduces specific inflammatory cytokines, suggesting that improvement both skeletal and cardiac regeneration operates in part by modulation of the inflammatory response. In collaboration with the Nerlov lab (page 113) we have explored the role of the innate immune system in the regeneration process, linking the pathways leading to macrophage polarisation and effective tissue repair.

In the heart, supplemental mIGF-1 expression increases progenitor cell pools, induced new signalling pathways and results in complete cardiac repair after myocardial infarction with minimal scar formation. We are currently exploring the signals in the epicardium, the outer cell layer of the heart, which may contribute to improved regenerative response. More recently, we have extended our studies of regeneration to the skin, where supplemental mIGF-1 expression improves wound healing and accelerates hair follicle formation and cycling.

Expression of IGF-1 isoforms *in vivo* has allowed us to assign specific functions of different peptide domains in muscle hypertrophy and regeneration. The different responses evoked by various IGF-1 isoforms suggest specific mechanisms through which combinations of supplemental growth factors can improve regeneration, providing new targets for clinical intervention. Further studies in skeletal and cardiac muscle have implicated NFκB, calcineurin and Notch-mediated signalling pathways in the intervention of tissue damage and disease.

In a new project we have extended our studies of cell signalling in development to address the role of a Fibroblast Growth Factor decoy receptor FGFRL1 in embryonic patterning. FGFRL1 null mice present multiple dysmorphologies reminiscent of human Wolf-Hirschhorn syndrome, implicating the decoy receptor in the etiology of this congenital disease.



Expression of Fibroblast Growth Factor Receptor-Like 1 (Fgfrl1) in the 11.5 day mouse embryo. Expression is prominent in the brain, cranial placodes, pharyngeal arches, somites and heart. Sections through the heart (top) show expression in the endocardial cushions of the atrioventricular canal and outflow tract.

Future projects and goals

In our future research, we will harness conditional and inducible mouse genetics to characterise key mechanisms implicated in the regenerative response. We will characterise the molecular action of growth factors and their intracellular intermediates to identify further candidates for therapeutic application. Our studies are designed to define the common nodal points of signalling in mammalian regenerative processes as they relate to embryonic development. At the cellular level, we are particularly interested in the role played by myeloid cell lineages in controlling inflammation and promoting tissue repair. We hope to use this knowledge for developing clinically relevant interventions in ageing, injury and degenerative disease.

Selected references

Klimanskaya, I., Rosenthal, N. & Lanza, R. (2008). Derive and conquer: sourcing and differentiating stem cells for therapeutic applications. *Nat. Rev. Drug Discov.*, 7, 131-142

Lara-Pezzi, E., Felkin, L.E., Sarathchandra, P., George, R., Hall, J.L., Yacoub, M.H., Rosenthal, N., Birks, E.J. & Barton, P.J. (2008). Follistatin gene expression in heart failure and myocardial recovery following LVAD combination therapy. *Journal of Heart and Lung Transplantation*, 27, S220

Mourkioti, F. & Rosenthal, N. (2008). NFκB signaling in skeletal muscle: prospects for intervention in muscle diseases. *J. Mol. Med.*, 86, 747-59

Semenova, E., Koegel, H., Hasse, S., Klatte, J.E., Slonimsky, E., Bilbao, D., Paus, R., Werner, S. & Rosenthal, N. (2008). Overexpression of mIGF-1 in keratinocytes improves wound healing and accelerates hair follicle formation and cycling in mice. *Am. J. Pathol.*, 173, 1295-310

Cornelius Gross

PhD 1995, Yale University.

Postdoctoral research at Columbia University.

Group leader at EMBL Monterotondo since 2003. Senior Scientist since 2009.

Developmental programming of anxiety

Previous and current research

Anxiety disorders are debilitating mental illnesses characterised by excessive worry and exaggerated responses to threatening stimuli. Epidemiological studies suggest that both genetic and environmental factors contribute to the prevalence of these disorders. For example, 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 to study the neural circuits underlying anxiety behaviour in mice. Several ongoing projects in the lab are addressing this question from different angles.

Early gene-by-environment risk factors: We are particularly interested in understanding how exposure to early adverse experiences can program anxiety 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. We are also examining changes in gene expression and epigenetic marks associated with altered early environmental exposure. Finally, we are collaborating with psychiatrists to examine whether gene-by-environment risk factors identified in the mouse are also predisposing factors for behavioural disorders in humans.

Cellular substrates of anxiety: To help identify the cellular substrates of anxiety, we are using pharmaco-genetic transgenic tools for the rapid modulation of electrical activity in selected cell-types in the brain. We have used a pharmaco-genetic inhibition strategy to examine the contributions of hippocampal and amygdala cell-types to anxiety and fear behaviour. We are further developing these tools and combining them with electrophysiological recordings in awake behaving mice to identify the cell-types and circuits involved.

Future projects and goals

- identification of molecular mechanisms that mediate the long-term programming of behaviour by early environmental experiences in mice and humans (genetic, epigenetic, hormonal, electrophysiological, morphological and signalling mechanisms);
- creation of mouse models of specific human genetic variations that have been associated with behavioural disorders;
- development and application of pharmaco-genetic transgenic technologies for the tissue and cell-type specific suppression of neural activity in behaving mice;
- identification and manipulation of the neurophysiological correlates of anxiety in awake behaving mice;
- study of copy number variations as predisposing factors for disease in mice.

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

Audero, E., Coppi, E., Mlinar, B., Rossetti, T., Caprioli, A., Banchaabouchi, M.A., Corradetti, R. & Gross, C. (2008). Sporadic autonomic dysregulation and death associated with excessive serotonin autoinhibition. *Science*, 321, 130-133

Carola, V., Frazzetto, G., Pascucci, T., Audero, E., Puglisi-Allegra, S., Cabib, S., Lesch, K.P. & Gross, C. (2007). Identifying molecular substrates in a mouse model of the serotonin transporter x environment risk factor for anxiety and depression. *Biol. Psychiatry*, 63, 840-6

Tsetsenis, T., Ma, X.H., Lo Iacono, L., Beck, S.G. & Gross, C. (2007). Suppression of conditioning to ambiguous cues by pharmacogenetic inhibition of the dentate gyrus. *Nat. Neurosci.*, 10, 896-902

Gross, C. & Hen, R. (2004). The developmental origins of anxiety. *Nat. Rev. Neurosci.*, 5, 545-552

Molecular physiology of somatosensation

Paul Heppenstall

PhD 1997, University of Edinburgh.

Postdoctoral work at the Max Delbrück Centrum, Berlin.

Junior Professor at the Charité, Berlin.

Group leader at EMBL Monterotondo since 2008.

Previous and current research

Somatosensation is the process by which we sense touch and pain. It is dependent upon specialised sensory neurons which extend from the skin to the spinal cord and are tuned to detect mechanical, thermal and chemical stimuli. Surprisingly, the mechanisms that transduce these forces into electrical signals at the peripheral endings of sensory neurons are not well understood. Our research focuses on identifying and characterising these transduction components and exploring how they are altered during chronic pain states.

In order to study somatosensation we use a combination of electrophysiological, molecular and cellular techniques. With the help of intact electrophysiology preparations we are able to examine functional properties of sensory neurons at their peripheral and central terminals *in vitro*. For example, using a hemisected spinal cord preparation, we investigated the role of the neurotrophic factor BDNF in synaptic plasticity in the spinal cord. We demonstrated that BDNF is released from nociceptors onto spinal neurons and modulates spinal reflex activity. Furthermore, we were able to show that this occurs via an acute mechanism, supporting the idea that BDNF acts as a synaptic modulator. Thus, BDNF has a direct role in pain-related neurotransmission and might mediate the central sensitisation associated with chronic pain.

In a second project we are interested in mechanisms of touch sensitivity of peripheral sensory neurons. Normal mechanical sensitivity is dependent on interactions between stomatin-like proteins and a family of ion channels called ASICs. We are currently using biochemical, electrophysiological and molecular imaging techniques to probe the nature of these interactions and to characterise the mechanotransduction complex in detail.

More recently we have been investigating the ion channel TRPA1, a member of the Transient Receptor Potential (TRP) family of channels. TRPA1 is an excitatory ion channel that is expressed by nociceptors and has a key role in detecting noxious chemicals. We have demonstrated that intracellular Ca^{2+} directly activates TRPA1 and that this occurs via an EF-hand domain in the N-terminus of the protein. Furthermore, we have shown that this domain is essential for the normal function of TRPA1 and that in the absence of Ca^{2+}, the sensitivity of TRPA1 to noxious chemicals is almost abolished. This data could serve as a starting point for the development of TRPA1 antagonists with prospective clinical applications as new types of analgesics.

Future projects and goals

A major focus of the laboratory is to correlate cellular studies on somatosensation with observation made at the whole animal level. To this end we are employing genetic approaches in mice combined with electrophysiological and molecular imaging techniques. Future goals include:

- point mutagenesis of ion channels and associated proteins to determine their role in sensory transduction;
- identification of novel genes involved in touch and pain;
- tissue-specific and conditional mutagenesis of pain-related genes in defined subpopulations of sensory neurons;
- development of new techniques to measure functional properties of sensory neurons at their terminals.

Selected references

Wetzel, C., Hu, J., Riethmacher, D., Benckendorff, A., Harder, L., Eilers, A., Moshourab, R., Kozlenkov, A., Labuz, D., Caspani, O., Erdmann, B., Machelska, H., Heppenstall, P.A. & Lewin, G.R. (2007). A stomatin-domain protein essential for touch sensation in the mouse. *Nature*, 445, 206-9

Zurborg, S., Yurgionas, B., Jira, J.A., Caspani, O. & Heppenstall, P.A. (2007). Direct activation of the ion channel TRPA1 by Ca^{2+}. *Nat. Neurosci.,* 10, 277-9

Heppenstall, P.A. & Lewin, G.R. (2006). A role for T-type Ca^{2+} channels in mechanosensation. *Cell Calcium*, 40, 165-74

Heppenstall, P.A. & Lewin, G.R. (2001). BDNF but not NT-4 is required for normal flexion reflex plasticity and function. *Proc. Natl. Acad. Sci. USA*, 98, 8107-12

Liliana Minichiello

PhD, University of Naples, Italy, and The National Cancer Institute, NIH, Bethesda, (MD), USA.

Postdoctoral work at EMBL Heidelberg 1993/99.

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

Studying signal transduction in the mouse nervous system is the main focus of the group. In particular, one major question is to understand the molecular mechanisms of learning. To address this question we are using different approaches. To determine whether the molecular pathways required for learning are also those generating long-term potentiation (LTP, considered to be the mechanism for acquisition and storage of information by synapses in the brain) when measured directly on the relevant circuit of a learning animal, we have employed a novel combination of *in vivo* methods combined with highly defined genetic mouse models, which allow us to interfere with single phosphorylation sites on a large receptor protein. Thus, we have been able to show that signalling through the TrkB receptor and its PLCγ docking-site is important for associative learning and parallel LTP (see figure), indicating 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

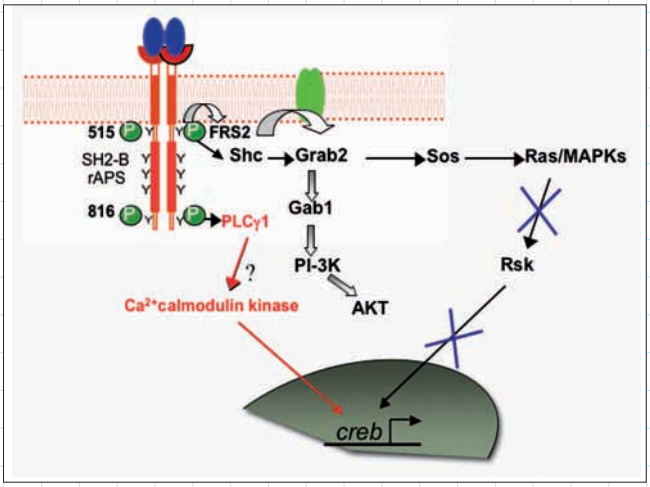
We have now begun to address molecular mechanisms by which TrkB receptor regulates synaptic function by using *in vivo* cell specific proteomics. We would like to underpin specific signalling complexes formed downstream of the TrkB receptor. The methods we are employing include the co-precipitation/mass spectrometry approach combined with 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.

Critical to brain function is the balance between inhibition and excitation. Thus, to determine functional and/or morphological feature of inhibitory interneurons, current areas of interest include selective knockouts of key molecules expressed in inhibitory interneurons. As the neurotrophins and their cognate receptors are expressed also in this 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 so far generated a few new 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:

- to define molecular mechanisms regulating synaptic plasticity;
- to understand the molecular basis of neurodegeneration;
- to understand mechanisms underling neuronal diversification.



The PLCγ site, and subsequent phosphorylation of calcium calmodulin kinase/s and CREB couples learning and parallel changes in hippocampal synaptic plasticity in vivo.

Selected references

Serguera, C., Triaca, V., Kelly-Barrett, J., Banchaabouchi, M.A. & Minichiello, L. (2008). Increased dopamine after mating impairs olfaction and prevents odor interference with pregnancy. *Nat. Neurosci.*, 11, 949-956

Gruart, A., Sciarretta, C., Valenzuela-Harrington, M., Delgado-Garcia, J.M. & Minichiello, L. (2007). Mutation at the TrkB PLCγ-docking site affects hippocampal LTP and associative learning in conscious mice. *Learn. Mem.*, 14, 54-62

Calella, A.M., Nerlov, C., Lopez, R.G., Sciarretta, C., von Bohlen Und Halbach, O., Bereshchenko, O. & Minichiello, L. (2007). Neurotrophin/Trk receptor signaling mediates C/EBPα, -β and NeuroD recruitment to immediate-early gene promoters in neuronal cells and requires C/EBPs to induce immediate-early gene transcription. *Neural Develop.*, 2, 4

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, 3803-3814

Transcription factor function in development, physiology and disease

Claus Nerlov

PhD 1995, University of Copenhagen.

Postdoctoral research at EMBL Heidelberg.

Head of Laboratory of Gene Therapy Research, Copenhagen Uni. Hospital.

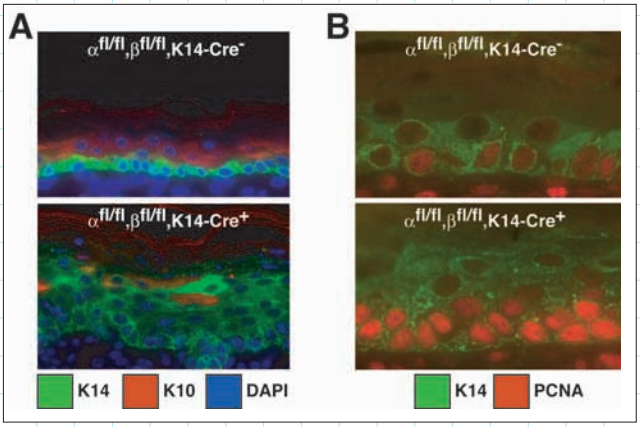
Group leader at EMBL since 2001.

Visiting Professor, Stem Cell Center, University of Lund, Sweden from 2005.

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 region leucine 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.



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.

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.

Selected references

Kirstetter, P., Schuster, M.B., Bereshchenko, O., Moore, S., Dvinge, H., Kurz, E., Theilgaard-Monch, K., Mansson, R., Pedersen, T.A., Pabst, T., Schrock, E., Porse, B.T., Jacobsen, S.E., Bertone, P., Tenen, D.G. & Nerlov, C. (2008). Modeling of C/EBPα mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells. *Cancer Cell*, 13, 299-310

Nerlov, C. (2007). The C/EBP family of transcription factors: a paradigm for interaction between gene expression and proliferation control. *Trends Cell Biol.*, 17, 318-324

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., 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

Dónal O'Carroll

PhD 1999, Research Institute of Molecular Pathology, Vienna.

Postdoctoral research at The Rockefeller University, New York.

Group leader at EMBL Monterotondo since 2007.

Small non-coding RNA function in development and physiology

Previous and current research

MicroRNAs (miRNAs) are small non-coding RNA molecules that have been identified as potent negative regulators of gene expression. miRNA-mediated gene silencing is executed by the multi-protein RNA–induced silencing complex (RISC). At the core of RISC are Dicer and an Argonaute (Ago) protein that interact to generate miRNA and execute their function. Dicer cleaves miRNA from its precursors whereas Ago proteins (Ago1-4) bind miRNA and mediate gene silencing. Using hematopoiesis in mice as a model system to study the physiological function of RISC components and the mechanism of miRNA-mediated gene silencing *in vivo*, we inactivated components of this complex using conditional alleles of Dicer (Dcr) and Argonaute (Ago1-4) genes. While Dcr function is absolutely required for early hematopoiesis, we found that Ago2 selectively controls early development of B lymphoid and erythroid cells. We showed that the unique and defining feature of Ago2, the Slicer endonuclease activity, is dispensable for hematopoiesis. Instead, we have identified Ago2 as a key regulator of miRNA homeostasis, with deficiency in Ago2 impairing miRNA biogenesis from precursor-miRNAs.

The major foci of the laboratory are to understand how miRNAs regulate gene expression *in vivo* and to contribute to the development and homeostasis of hematopoiesis and spermatogenesis. The analysis of Ago2 deficiency during hematopoietic development has uncovered the processes of early erythroid and B lymphocyte development as being sensitive to miRNA dosage. The identity and mechanism of the miRNAs that control these developmental transitions are currently being investigated. Similar approaches are being applied to determine the dependence of spermatogenesis on miRNA function. By identifying the miRNAs that control these various developmental systems and their network of targets, we strive to understand the physiological merits of this gene-silencing pathway. In the pursuit of these goals we currently employ genetic, biochemical and miRNA profiling/sequencing, as well as pharmacological approaches.

Future projects and goals

- To determine the physiological mechanism of Ago2 function in miRNA biogenesis and the execution of miRNA function.
- To identify the miRNAs and their respective targets that control early erythropoiesis and B cell development.
- To explore the function of small RNAs during spermatogenesis.

Selected references

O'Carroll, D., Mecklenbrauker, I., Das, P.P., Santana, A., Koenig, U., Enright, A.J., Miska, E.A. & Tarakhovsky, A. (2007). A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway. *Genes Dev.*, 21, 1999-2004

Schaefer, A., O'Carroll, D., Tan, C.L., Hillman, D., Sugimori, M., Llinas, R. & Greengard, P. (2007). Cerebellar neurodegeneration in the absence of microRNAs. *J. Exp. Med.*, 204, 1553-8

Tang, F., Kaneda, M., O'Carroll, D., Hajkova, P., Barton, S.C., Sun, Y.A., Lee, C., Tarakhovsky, A., Lao, K. & Surani, M.A. (2007). Maternal microRNAs are essential for mouse zygotic development. *Genes Dev.*, 21, 644-8

Yi, R., O'Carroll, D., Pasolli, H.A., Zhang, Z., Dietrich, F.S., Tarakhovsky, A. & Fuchs, E. (2006). Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nat. Genet.*, 38, 356-62

Notes

Notes

Notes

Index of group leaders

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

Texts:
EMBL group and team leaders and Unit Coordinators

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

Cover Design:
Vienna Leigh

Printed by ColorDruck, Leimen

European Molecular Biology Laboratory

Heidelberg • Hinxton • Grenoble • Hamburg • Monterotondo