European Molecular Biology Laboratory

Research at a Glance 2010

EMBL

Contents

EMBL – Europe's leading 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, engineering, 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.

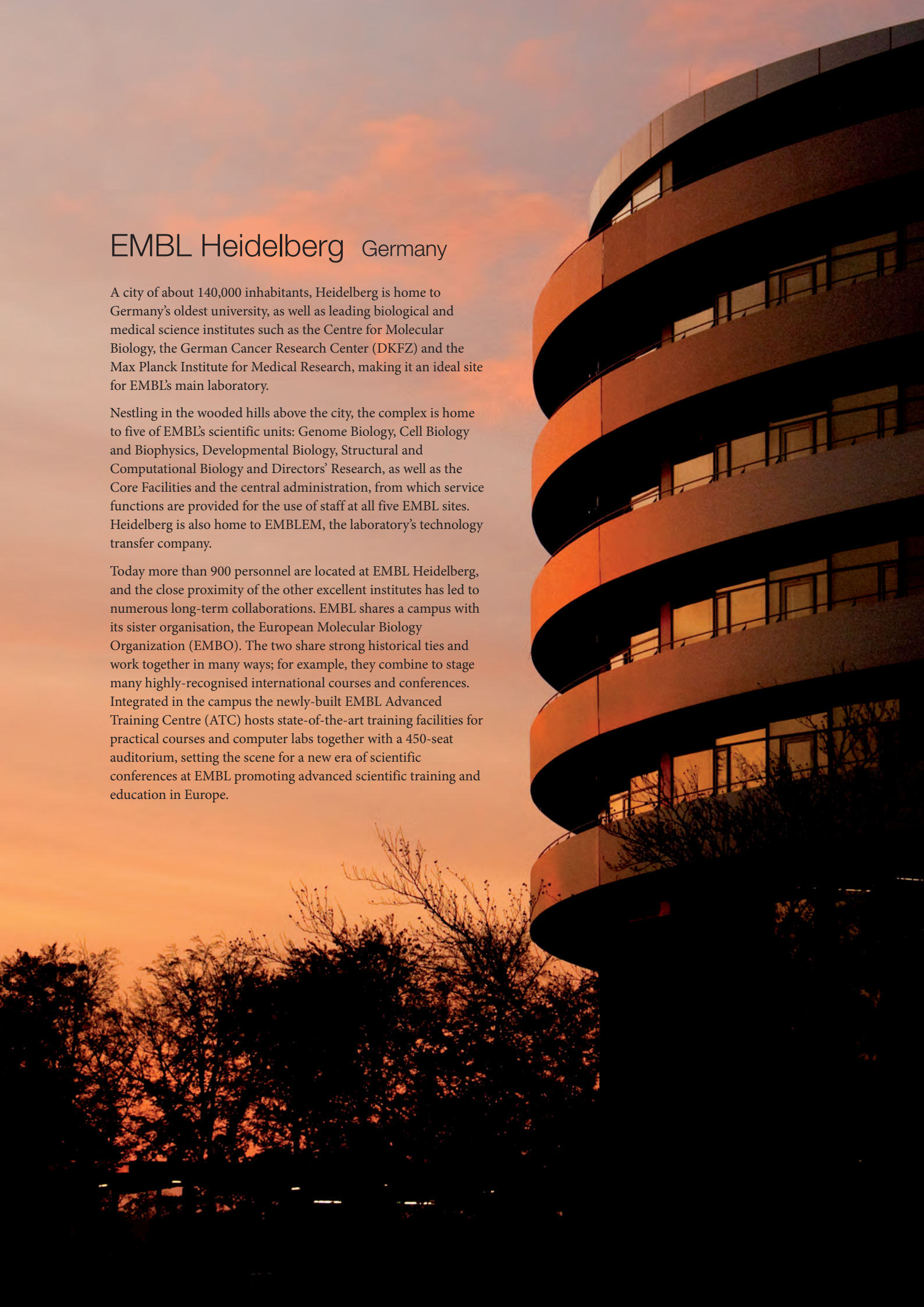
Iain Mattaj
EMBL Director General

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: Genome Biology, 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 (EMBO). 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 campus the newly-built EMBL Advanced Training Centre (ATC) hosts state-of-the-art training facilities for practical courses and computer labs together with a 450-seat auditorium, setting the scene for a new era 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
Head of the 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 Head of Cell Biology and Biophysics Unit since 2009; Head of Unit since 2010.

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 ER and but perforated by nuclear pore complexes (NPCs) to mediate selective traffic between nucleus and cytoplasm. In M-phase, most metazoan cells reversibly dismantle the highly ordered structure of the NE. Chromosomes that are surrounded by nuclear membranes in interphase are taken over by cytoplasmic spindle microtubules in order to segregate them. After chromosome segregation, the roles are reversed again and 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. Biogenesis of the nucleus and the formation of M-phase chromosomes are essential but poorly understood processes. To study them, we are using advanced fluorescence microscopy-based methods to understand the dynamics and function of structural and regulatory proteins. Quantitative imaging is coupled with computerised image processing and simulations to extract biophysical parameters and build mechanistic models. As biological models, we are using somatic mammalian cells for mitosis, as well as oocytes from starfish and mouse, in which we study the asymmetric meiosis they undergo to become a fertilizable egg.

In somatic mammalian cells, we could show that nuclear membrane formation originates from the endoplasmic reticulum. We found that nuclear breakdown is triggered by the stepwise disassembly of NPCs followed by microtubule-driven tearing of the nuclear lamina and that nuclear reformation initiates by the ordered reassembly of nucleoporin complexes directly on chromosomes. We could further show that mitotic chromosome positions are non-randomly transmitted through the cell cycle and that chromosomes surprisingly reach their most compact state in anaphase to facilitate the correct segregation of their arms. In animal oocytes, we could show that asymmetric transport of chromosomes to the cell surface is mediated by a contractile F-actin network (figure 1) rather than by microtubules.

Future projects and goals

Objective of our future work is to gain comprehensive mechanistic insight into cell cycle remodelling of the nucleus. In somatic cells, we are focusing on the mechanism of nuclear growth in interphase, as well as chromosome condensation and organisation in somatic cells. Here, we recently showed a fractal organisation of chromosomes in interphase nuclei (figure 2). To achieve a systems level understanding and assay all relevant proteins, we have automated and standardised many fluorescence imaging modalities. For example, high-throughput live imaging combined with RNAi screening enabled us to identify many novel genes with functions in mitosis. In oocytes we are pursuing the molecular mechanism of actin-mediated chromosome transport as well as homologous chromosome segregation.

Selected references

Bancaud, A., Huet, S., Daigle, N., Mozziconacci, J., Beaudouin, J. & Ellenberg, J. (2009). Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. *EMBO J.*, 28, 3785-3798

Schuh, M. & Ellenberg, J. (2008). A new model for asymmetric spindle positioning in mouse oocytes. *Curr Biol.*, 18, 1986-1992

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, 857-865

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

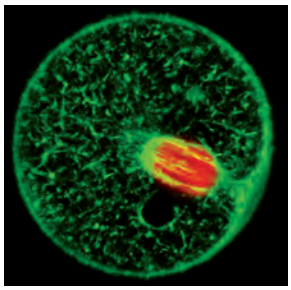
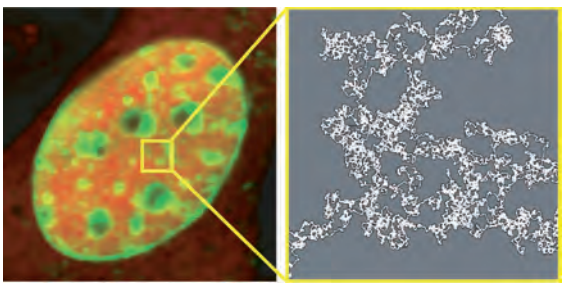


Figure 1: The meiotic microtubule spindle (red) uses a cytoplasmic F-actin network (green) for its asymmetric spindle positioning in live mouse oocytes.

Figure 2: Chromatin (green) in the nucleus of a live somatic mammalian cell leads to volume exclusion of other macromolecules (red, fluorescently labelled dextran) (left panel); and exhibits a fractal organisation (right panel).



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.

Selected references

Khmelinskii, A., Roostalu, J., Roque, H., Antony, C. & Schiebel, E. (2009). Phosphorylation-dependent protein interactions at the spindle midzone mediate cell cycle regulation of spindle elongation. *Dev. Cell*, 17, 244-256

Welsch, S., Miller, S., Romero-Brey, I., Merz, A., Bleck, C.K., Walther, P., Fuller, S.D., Antony, C., Krijnse-Locker, J. & Bartenschlager, R. (2009). Composition and three-dimensional architecture of the dengue virus replication and assembly sites. *Cell Host Microbe*, 5, 365-375

Kitamura, E., Tanaka, K., Komoto, S., Kitamura, Y., Antony, C. & Tanaka, T.U. (2010). Microtubules generated at kinetochores with distal plus ends: their roles and limited lifetime in mitosis. *Dev. Cell*, in press

Roque, H. & Antony, C. (2010). Electron microscopy of model systems: the fission yeast *Schizossacharomyces pombe*. *Methods in Cell Biology*, in press

Cellular electron tomography of cells

Previous and current research

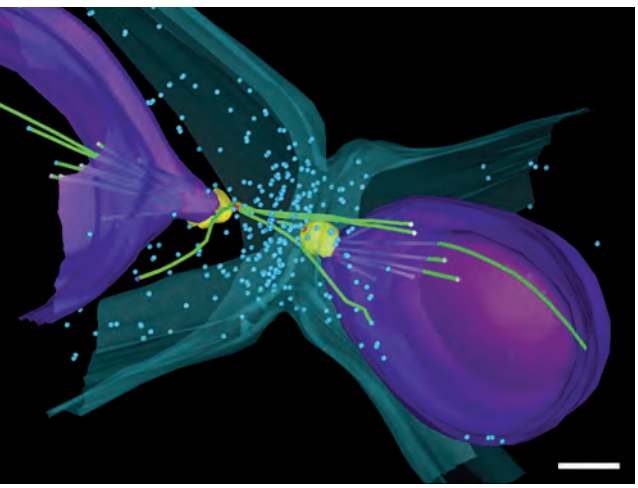
The installation in 2008 of a new Tecnai F30 tomography microscope has allowed us and other EMBL groups to develop projects aiming at resolving the 3D organisation of cells in various contexts. The focus of interest in the team is the organisation of microtubular cytoskeleton arrays in both fission yeast and budding yeast as well as in the *Xenopus* mitotic spindle. For this purpose we use electron tomography, which not only allows the reconstruction, modelling and quantification of subcellular elements, but also enables the visualisation of a number of fine structural features which would not be detectable by conventional EM.

Fission yeast microtubule dynamics: In fission yeast we have 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 by Johanna Höög in collaboration with the Brunner group (both formerly EMBL).

Microtubule bundling factors in fission yeast: In parallel we also investigated microtubule bundling factors in fission yeast. To understand inter-microtubule bonds, we reconstructed microtubule arrays in strains with deletions affecting 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. We observed that cells lacking ase1p and klp2p can still bundle microtubules but with an altered intermicrotubule spacing.We further identified Dis1p (XMAP215 homologue) as an alternative microtubule bundling factor which accumulates at microtubule overlap areas in cells that are deleted for both ase1 and klp2. Dis1p appears to be a critical factor for the maintenance of interphase microtubules. The project was carried out by Helio Roque in collaboration with the Brunner group (both formerly EMBL).

Spatial organisation of microtubules in the budding yeast mating pathway: A new project has been started by Romain Gibeaux (PhD student) in collaboration with the Knop group (page 13) concerning the morphological and molecular analysis of the karyogamy process in the budding yeast mating pathway. This process involves dynamic events with microtubule connections between the two nuclei and associated SPBs which are instrumental in bringing the two nuclei in close apposition until nuclear fusion. The morphological 3D analysis of these events are currently carried out by electron tomography performed on shmooing cells and fused cells in the process of nuclear congression. The molecular basis of the process and the role played by key molecules, either SPB or microtubule-associated, is studied in collaboration with the Knop group.

3D reconstruction of mating yeast cells. At this stage of the karyogamy process the plasma membranes of the two cells have already fused but the SPBs (yellow) are not joined yet. Microtubules (green) originating from both SPBs are invading the opposite cell respectively. Nuclear envelope (blue); plasma membrane (blue-green), transport vesicles (blue). The latter are accumulating in the fusion area showing an intense trafficking activity in this area. (Bar, 300 nm; picture by Romain Gibeaux).



Future projects and goals

In conjunction with in-house and external research groups (collaboration with Francois Nédélec (page 14) and Rebecca Heald (UCB)) we are now starting a major project aiming to reconstruct the *Xenopus laevis* mitotic spindle from egg extracts. Cryofixation and preparation of samples for tomography acquisition are the first steps. The large-scale reconstruction of such a huge structure, or parts of it, will be performed using extensive montaging and joining of tomograms. In the course of this project we intend to elucidate the spindle microtubule architecture at high resolution, and in doing so derive information about microtubule polarity, in particular in the midzone of the spindle. Xavier Heiligenstein (PhD student) and Erin Tranfield (postdoc) are working on this project.

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.

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.

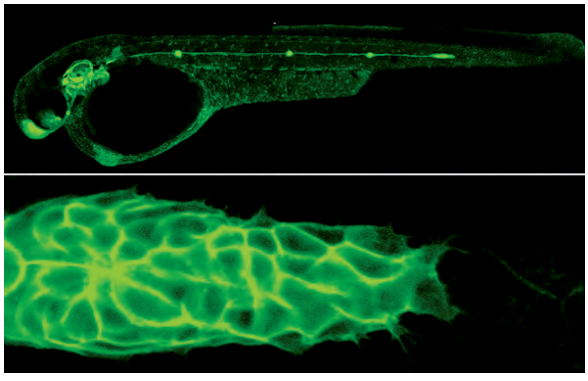


Figure 1: The zebrafish migrating lateral line organ allows collective migration to be easily studied in vivo.

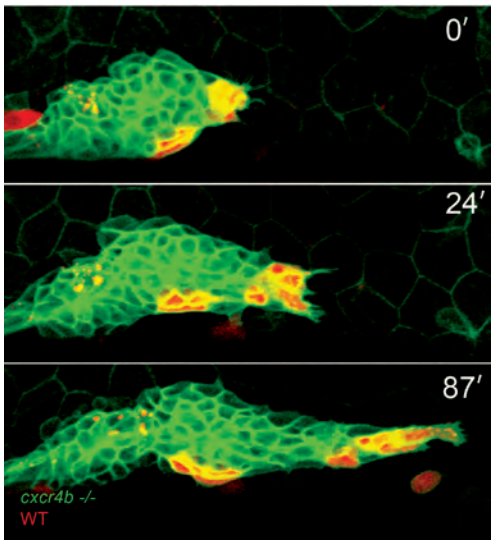


Figure 2: Transplanted wild-type cells (red) rescue the migration of cxcr4 mutant primordia (green).

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

PhD 2003, Institute of Molecular Pathology, Vienna.
Postdoctoral work at the University of Oxford.
Group leader at EMBL since 2007.

Christian Häring

Chromosome structure and dynamics

Previous and current research

Eukaryotic chromosomes undergo enormous changes in structure and organisation over the course of a cell cycle. The most fascinating of these changes was first observed by cell biologists more than 125 years ago; as cells prepare for cell division, sister chromatid pairs individualise into highly condensed rod-shaped structures, which attach to the mitotic spindle via their kinetochores. Once all sister kinetochore pairs have attached in a bipolar fashion, the connection between sister chromatids is released to trigger their segregation towards opposite cell poles. These ordered events ensure 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 in humans.

The overall aim of our research is to understand the action of molecular machines that organise chromosomes prior to and during cell divisions. Recent research has identified two highly conserved multi-subunit protein complexes called cohesin and condensin as central players for chromosome segregation. Both complexes are composed of heterodimers of SMC (structural maintenance of chromosomes) subunits whose ABC ATPase head domains are connected by so-called kleisin subunits to form gigantic tripartite ring structures.

A body of evidence supports the notion that the cohesin complex holds sister chromatids together by entrapping them within its ring structure (figure 1) until a protease named separase opens the ring by site-specific cleavage of cohesion's kleisin subunit and thereby initiates chromosome segregation. Our working hypothesis is that condensin uses a similar topological principle to stabilise loops of chromatin in order to give mitotic chromosomes their characteristic shape (figure 2). We use an interdisciplinary combination of biochemistry, molecular biology, cell biology and, in collaboration, chemical and structural biology to discover how these two protein complexes function at the molecular level in yeast and mammalian cells.

In an independent project, we are exploring novel approaches to identify additional players that direct the formation of mitotic and meiotic chromosomes using genetics and time-resolved light microscopy methods.

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 chromosome?

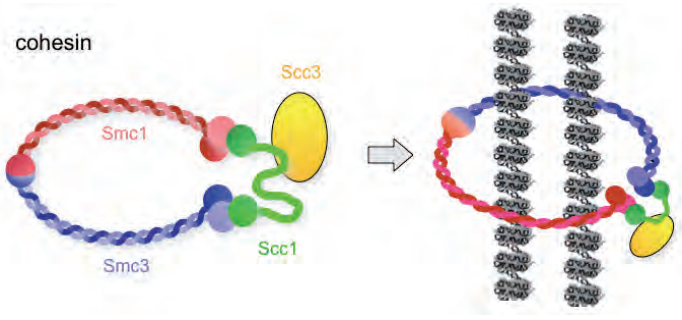


Figure 1 (top): Model of the cohesin ring holding sister chromatids together.

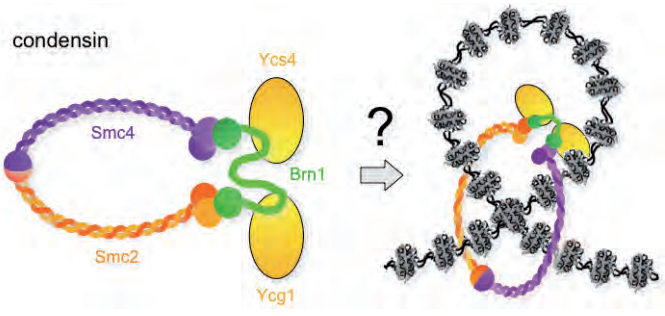


Figure 2 (bottom): Model of the condensin ring structuring chromosomes.

Selected references

Nasmyth, K. & Haering, C.H. (2009). Cohesin: its roles and mechanisms. *Annu. Rev. Genet.*, 43, 525-558

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

Arumugam, P., Nishino, T., Haering, C.H., Gruber, S. & Nasmyth, K. (2006). Cohesin's ATPase activity is stimulated by the C-terminal Winged-Helix domain of its kleisin subunit. *Curr. Biol.*, 16, 1998-2008

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

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.

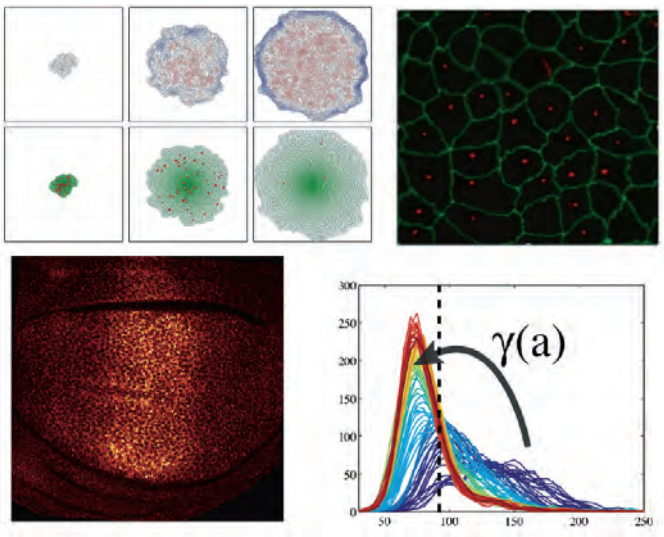
We have recently investigated the interplay between the growth of *Drosophila* wing imaginal discs and the formation of the Dpp morphogen gradient. Our results suggest a new scenario of size determination, where disc size is determined relative to the fixed morphogen distribution. Our model shows that 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.

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.



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

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.

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

Marko Kaksonen

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

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

Dynamics of membrane trafficking

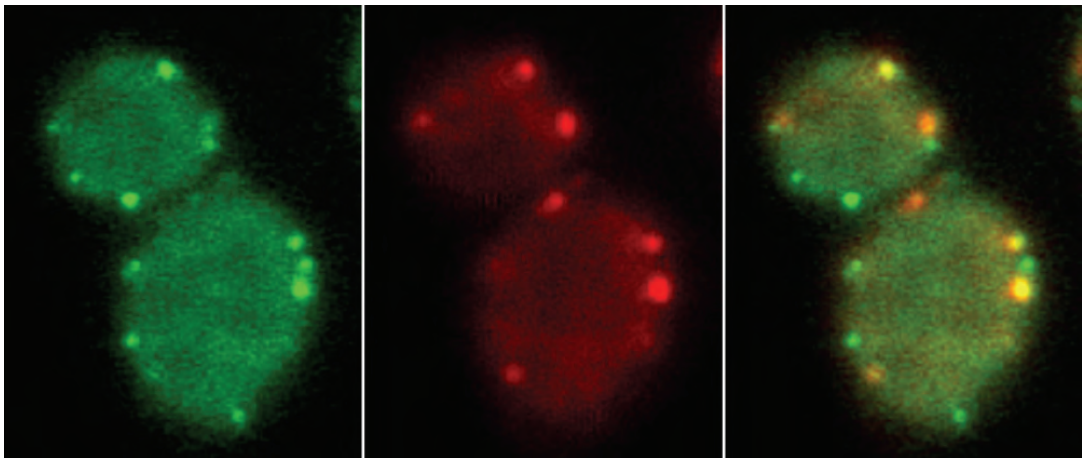
Previous and current research

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

More specifically, we are interested in the formation of cargo-loaded transport vesicles, such as endocytic vesicles. The formation of the endocytic vesicle is driven by a highly dynamic molecular machinery composed of more than 50 different protein species and several thousand individual protein molecules. Our main experimental organism is budding yeast, *Saccharomyces cerevisiae*. We combine powerful yeast genetics with quantitative live-cell imaging methods, 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.

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. We mainly focused 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, in the initiation of spore morphogenesis and on membrane shaping of the spore.

Mating is another important aspect of the life cycle of yeast. How do yeast cells faithfully 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 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 (figure 1). We consider both quiescent and signalling conditions. The goal is to enhance our understanding of the spatial and dynamic organisation of the signaling processes, which will help us derive and further develop quantitative models of the processes that regulate signalling through these pathways.

Our work on meiosis has gradually shifted to questions that relate to the role und function of genome recombination. As a model, we use computer simulations of populations of yeast-like genomes that undergo yeast-like live cycles. Here we address the role of meiosis and recombination and the impact of genome architecture on handling deleterious mutational load (figure 2). 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 life-cycle properties to *S. cerevisiae*, but which has one notable and most interesting difference: this species appears to not 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.

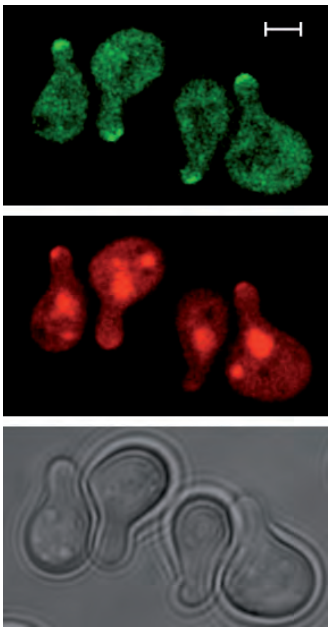


Figure 1: Cellular distribution of some of the components of the MAP kinase signal module, which transmits the signal in yeast cells stimulated with mating pheromone.

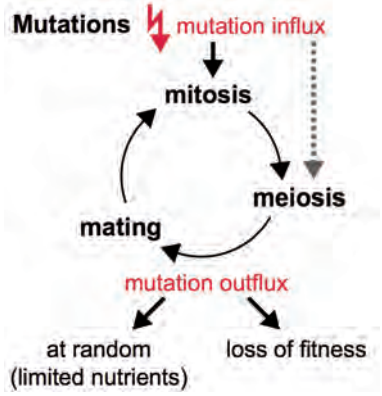


Figure 2: Evolution is fuelled by mutations. Sexual cycles (mating and meiosis) constitute processes that enable efficient handling of good (beneficial) and bad (deleterious) mutations.

Selected references

Keller, P.J. & Knop, M. (2009). Evolution of mutational robustness in the yeast genome: a link to essential genes and meiotic recombination hotspots. *PLoS Genet.*, 5, e1000533

Taxis, C., Stier, G., Spadaccini, R. & Knop, M. (2009). Efficient protein depletion by genetically controlled deprotection of a dormant N-degron. *Mol. Syst. Bio.*, 5, 267

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

Taxis, C., Keller, P., Kavagiou, Z., Jensen, L.J., Colombelli, J., Bork, P., Stelzer, E.H.K. & Knop, M. (2005). Spore number control and breeding in *Saccharomyces cerevisiae*: a key role for a self-organizing system. *J. Cell Biol.*, 171, 627-640

François Nédélec

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

Selected references

Brun, L., Rupp, B., Ward, J.J. & Nedelec, F. (2009). A theory of microtubule catastrophes and their regulation. *Proc. Natl. Acad. Sci. USA*, 106, 21173-8

Dinarina, A., Pugieux, C., Corral, M.M., Loose, M., Spatz, J., Karsenti, E. & Nedelec, F. (2009). Chromatin shapes the mitotic spindle. *Cell*, 138, 502-513

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

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

Cellular architecture

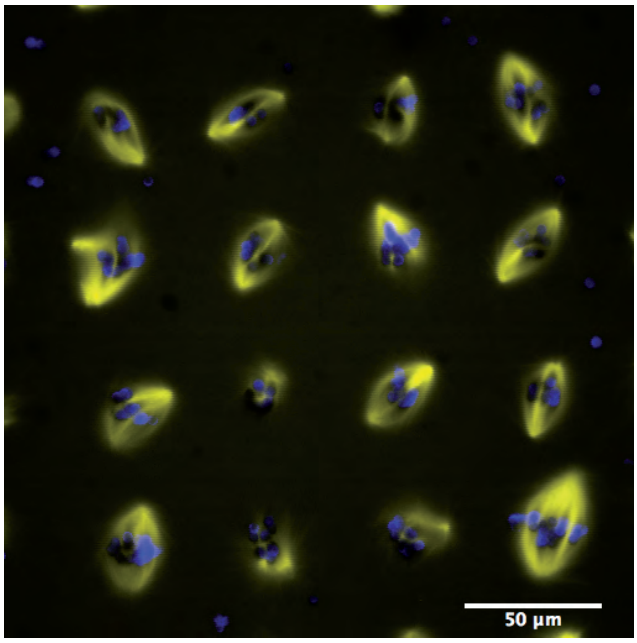
Previous and current research

Modern microscopy has demonstrated the dynamic nature of biological organisation. The mitotic spindle, for example, is a stable and solid cellular structure; in a given cell type, it has a precise symmetry and very reproducible dimensions. Yet its main components – polar filaments called microtubules – are in rapid turnover. They grow, shrink and disappear in a matter of minutes, within a spindle that may remain steady for hours. Chromosomes and microtubules are connected by proteins which continuously and stochastically bind and unbind. The resulting assembly is highly dynamic and yet stable and remarkably precise; it applies the balanced forces necessary to position and segregate the chromosomes exactly.

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

Obviously, understanding the collective behaviour is the challenge here, but it cannot be deduced from a simple statistical average. It is a challenging problem for several reasons: 1) the diversity of molecular players is often enormous; 2) their interactions are often dynamic and out-of-equilibrium; and 3) the properties of the proteins have been selected for the biological task by natural evolution. Understanding biological phenomena from their multiple biological components – systems biology – is a cutting-edge research topic.

We address this problem 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 even be specified at will. In the past, we developed innovative numerical methods to simulate the collective behaviour of multiple polar fibres and of their associated proteins. They are implemented in a simulation engine called cytosim, which is also made available to our community. Simulations are often used to validate or refute existing ideas, but we try to use them 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 new hypotheses.



An array of mitotic spindles obtained in vitro *with* Xenopus laevis *egg extracts.*

Future projects and goals

We will study systems in which experiments and theory can be synergistically combined. We currently focus on *Xenopus* egg extracts, an experimental system in which many aspects of mitosis can be recapitulated. We are also generally interested in modeling cellular processes in which the cytoskeleton is a major player, such as the different stages of mitosis, the generation of cell shape in *S. pombe*, or the generation of asymmetry during cell division.

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.

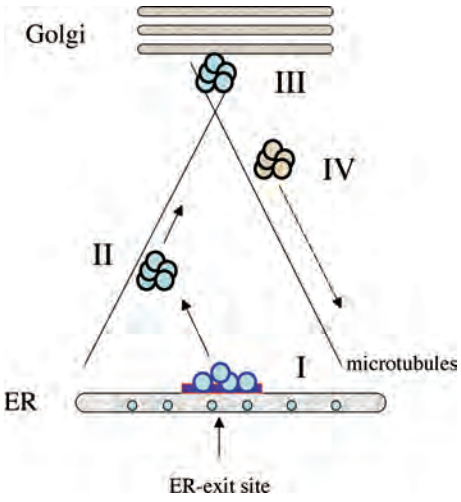
The specific questions we are presently addressing in this context are: 1) what are the mechanisms underlying the regulation of ER-exit sites biogenesis and function; 2) how are ER exit and microtubule mediated ER to Golgi transport coupled at the molecular level; 3) what are the mechanisms of Golgi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

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



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.

Selected references

Bartz, F., Kern, L., Erz, D., Zhu, M., Gilbert, D., Meinhof, T., Wirkner, U., Erfle, H., Muckenthaler, M., Pepperkok, R. & Runz, H. (2009). Identification of cholesterol-regulating genes by targeted RNAi screening. *Cell Metab.*, 10, 63-75

Erfle, H., Neumann, B., Rogers, P., Bulkescher, J., Ellenberg, J. & Pepperkok, R. (2008). Work flow for multiplexing siRNA assays by solid-phase reverse transfection in multiwell plates. *J. Biomol Screen.*, 13, 575-80

Simpson, J.C., Cetin, C., Erfle, H., Joggerst, B., Liebel, U. *et al.* (2007). An RNAi screening platform to identify secretion machinery in mammalian cells. *J. Biotechnol.*, 129, 352-65

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, 173-9

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.

Selected references

Laketa, V., Zarbakhsh, S., Morbier, E., Subramanian, D., Dinkel, C., Brumbaugh, J., Zimmermann, P., Pepperkok, R. & Schultz, C. (2009). Membrane-permeant phosphoinositide derivatives as modulators of growth factor signaling and neurite outgrowth. *Chem. Biol.*, 16, 1190-1196

Cobos-Correa, A., Trojanek, J.B., Diemer, S., Mall, M.A. & Schultz, C. (2009). Membrane-bound FRET probe visualizes MMP12 activity in pulmonary inflammation. *Nat. Chem. Biol.*, 5, 628-630

Neef, A.B. & Schultz, C. (2009). Selective fluorescence labeling of lipids in living cells. *Angew Chem. Int. Ed. Engl.*, 48, 1498-1500

Schultz, C., Neef, A., Goedhart, J. & Gadella, T.W.J. (2010). Imaging lipids in living cells. In 'Live Cell Imaging', Spector, D.L., Svedlow, J. & Jaron, B.F (eds), Cold Spring Harbor Press

Chemical Cell Biology

Previous and current research

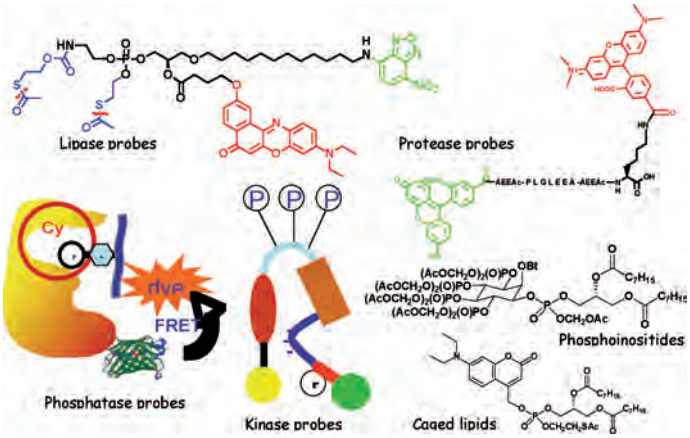
Before joining EMBL, our research focused on finding novel ways to stimulate chloride and water secretion of epithelial cells to help with the genetic disease cystic fibrosis (CF). Our compounds helped to investigate some of the underlying intracellular signalling pathways and provided drug candidates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

At EMBL, our interest has shifted to the basic signalling network underlying epithelial secretion as well as receptor endocytosis and recycling. We have 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 finding compounds that might be beneficial for unraveling basic principles in signal transduction and ultimately for CF patients. Currently, we use our prodrug approaches to dissect signalling networks by increasing the concentration of single lipids such as phosphoinositides in a non-invasive manner (Laketa *et al.*, 2009). In addition, we introduced a novel method to fluorescently label lipids inside fixed and living cells (Neef & Schultz, 2009). These efforts to specifically manipulate and detect small molecules and proteins in cells are supported by novels ways to model intracellular signalling networks. The imaging abilities within the group 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 (Cobos-Correa *et al.*, 2009).

Future projects and goals

In 2010, 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 (Subramanian *et al.*, accepted). Vesicle trafficking and endocytosis is investigated in collaboration with Rainer Pepperkok's group (page 15).



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.

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 8).

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 function.

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 physiological approaches to modern biology

Previous and current research

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

Although many modern technologies could operate in 3D, they are still 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, the mechanical properties, the media flux and the 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 applies and develops technologies for the observation of large and complex 3D biological specimens as a function of time. The technology of choice is the light sheet-based fluorescence microscopy (LSFM), which illuminates a specimen from the side and observes it at an angle of 90°. 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 of LSFM are the single plane illumination microscope (SPIM) and its more refined version (DSLM). LSFM take advantage of modern camera technology and are combinable with essentially every contrast and specimen manipulation tool found in modern light microscopes.

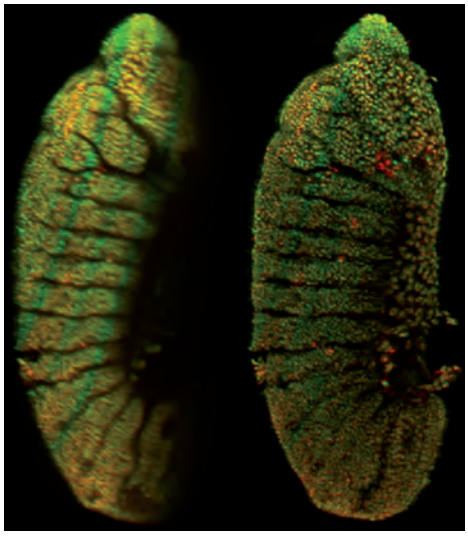
Future projects and goals

The optical path in LSFM is designed to allow high flexibility and modularity. We successfully integrated our nano-scalpel and devised a toolbox of photonic nano-tools. We plan to integrate them into our light sheet-based fluorescence microscopes and 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 temperature conditions provides model systems for studying organ morphogenesis.

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 sub-micrometre particles and the stimulation of selected compartments with optically trapped probes.

Amongst the biological specimens we currently use and intend to use in the future are *Drosophila*, zebra fish and amphioxus. A major effort is placed on the investigation of multi-cell structures such as spheroids and cysts. The group integrates the efforts of engineers as well as biologists, physicists and mathematicians.



Drosophila melanogaster *embryo: uncondensed DNA (green), condensed DNA (red). Left: Maximum-value projections of a single-view data set. Right: an 8-view MVD-Wiener fusion. From Swoger* et al., *2007,* Optics Express.

Selected references

Keller, P.J., Schmidt, A.D., Wittbrodt, J. & Stelzer, E.H.K. (2009). The zebrafish digital embryo: *in toto* reconstruction of zebrafish early embryonic development with Digital Scanned Light Sheet Fluorescence Microscopy. In 'Advanced Microscopy Techniques', Campagnola, P.J., Stelzer, E.H.K. & von Bally, G. (eds), *SPIE Conference Proceedings*, 7367-21, 1-10

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, 1065-1069

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

Huisken, J. *et al.* (2004). Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science*, 305, 1007-1009

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.

Physical systems biochemistry of cytoskeleton dynamics and function

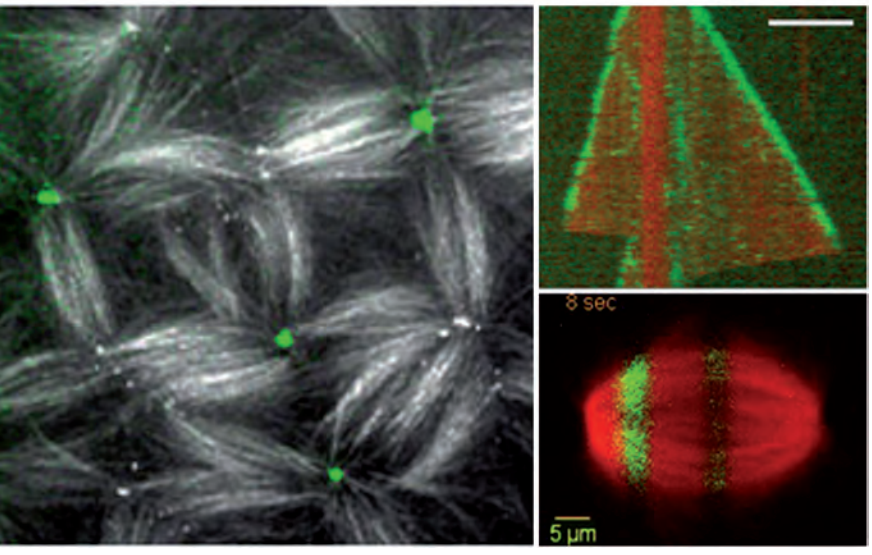
Previous and current research

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.



Far left: self-organised network of microtubules and plus- and minus-motors (Surrey et al., *2001,* Science*). Top left: time-space plot of microtubule end tracking (Bieling* et al., *2007,* Nature*). Bottom left: 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 precise coordination of all basic cellular and molecular processes in space and time. Groups in the Developmental Biology Unit seek to elucidate the principles and mechanisms underlying fundamental developmental processes. Using animal and plant models, our research integrates numerous, complementary approaches to understand how cellular processes are coordinated in living organisms.

Cell polarisation underlies many fundamental decisions in development in plants and animals. In many organisms, embryonic development begins before the onset of zygotic transcription, under the control of mRNAs and proteins localised asymmetrically in the egg. Mechanisms underlying cell polarisation, mRNA transport and translational control in *Drosophila* are under investigation in the unit. In plants, the polarised transport of auxin, which determines the positioning of lateral organs; how this molecule specifies different cell types is another topic of research.

During development, progenitor cells divide and differentiate into tissues of characteristic shape and function. Research in the unit aims to elucidate how cells in the early *Drosophila* embryo reorganise their content in response to the expression of key developmental transcription factors and, specifically, how tissue-specific gene expression controls protein and membrane trafficking, and how this trafficking regulates cell fate and behaviour.

Elucidating the temporal organisation of embryonic development is another area 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 signalling 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 in the context of a developing embryo.

The marine annelid *Platynereis* is an ideal model for exploring the evolution of cell types. Large-scale expression profiling at cellular resolution has revealed the evolutionary origin of the vertebrate hypothalamus. Using this model, research in the unit is aimed at solving 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. Combining live imaging and 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. Focusing on the regulatory architecture of key developmental loci, an aim of research in the unit is to understand the molecular mechanisms that control functional interactions between genes and remote cis-regulatory elements and to determine how they contribute to phenotypic variations during vertebrate evolution and in humans.

Anne Ephrussi
Head of the Developmental Biology Unit

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; Head of Unit since 2007.

Selected references

Trucco, A., Gaspar, I. & Ephrussi, A. (2009). High resolution analysis reveals the hierarchical assembly of endogenous *oskar* transport RNPs with their motors. *Cell*, 139, 983-998

Besse, F. & Ephrussi, A. (2008). Translational control of localized mRNAs: restricting protein synthesis in space and time. *Nat Rev Mol Cell Biol*, 9, 971-980

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*, 428, 959-963

Cell polarity and RNA localisation

Previous and current research

Polarity is a main feature of eukaryotic cells, underlying cell fate decisions, as well as many basic cellular functions and developmental processes. Cell polarisation involves the specific organisation of cytoskeletal structures and regulated targeting of organelles and molecules, including RNAs, to specific subcellular locations. Intracellular RNA transport coupled with localised translational control is a highly prevalent, conserved and powerful mechanism contributing to the functional polarisation of cells. We seek to understand the mechanisms regulating these basic cellular processes in a developmental context, in a living organism.

In *Drosophila*, asymmetrically localised cell fate determinants specify the body axes and patterning of the future embryo. The key determinants, *bicoid*, *gurken* and *oskar*, are transported as mRNAs to specific sites within the oocyte, where they are anchored and locally translated, ensuring spatial restriction of their activities. The cytoskeleton and specific motor proteins mediate mRNA transport and anchoring within the cell. We use these RNAs as models to understand how RNA localisation and translational control are regulated in space and time.

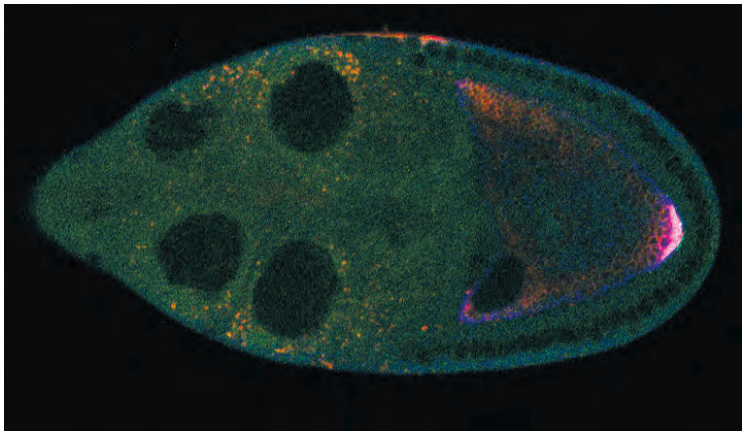
Drosophila is ideally suited for genetic, biochemical, and cell biological investigation of the processes of cell polarisation, mRNA localisation and translational control. We make use of this model system to study (1) cytoskeletal polarisation, (2) the assembly of the RNA transport complexes and their association with motors and the cytoskeleton mediating their movement, and (3) spatial control of translational within cells.

Future projects and goals

Combining genetics, 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 and associate with their motor proteins to form functional RNA transport complexes;
- the mechanisms coupling mRNA transport and translational control.

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.



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

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, postdoctoral 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.

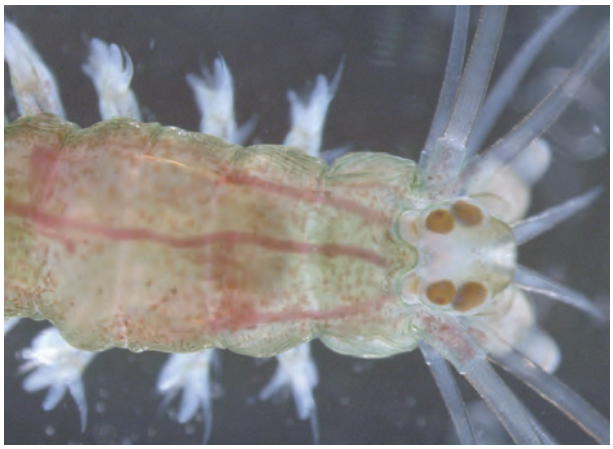
Finally, we have also established neurobiological assay systems for larval swimming and for adult learning, combined with computer modelling of these and of other complex behavioural traits, in order to investigate the functions of conserved cell type and to gain insight into the neurobiology of marine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

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



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

Alexander Aulehla

MD 2002, Albert-Ludwigs-University, Freiburg.

Research at the MD Anderson Cancer Center, Houston, USA and the MPI, Freiburg.

PhD 2008, Paris VI University.

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

Group leader at EMBL since 2009.

Timing of mammalian embryogenesis

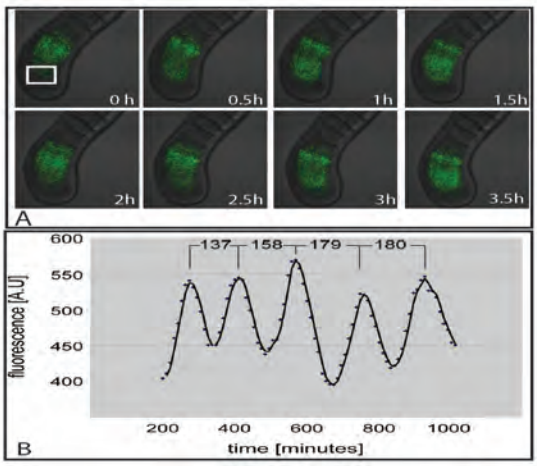
Previous and current research

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

During an embryo's journey from a single cell to a complex organism, countless patterning processes unfold with remarkable precision, both spatially and 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 time measured during embryonic development, and what extrinsic and intrinsic signals control this timing? How do embryonic clocks function? 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 such embryonic clock, the somite segmentation clock, is thought to control the formation of the pre-vertebrae that form periodically in a head-to tail sequence within the paraxial mesoderm, with a periodicity around two hours. In mouse embryos, this clock drives the oscillatory activity of several signalling 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. To address these questions, the ability to directly observe and quantify the temporal dynamics of signalling pathway activity is a key prerequisite.

We have been able to visualise oscillatory transcriptional activity in developing mouse embryos with high temporal and spatial resolution. We are now developing this further to establish a novel, versatile real-time reporter system that will allow us to visualise the dynamics of Wnt-signalling activity at various levels. This signalling pathway serves a multitude of evolutionary conserved functions during development and has been shown to play an essential role during somite formation. The reporter system is designed to reflect Wnt-signalling activity both at 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 and to functionally test their role in embryonic patterning. We are particularly interested in identifying the intrinsic and extrinsic factors responsible for controlling these oscillations within the segmentation process. The insight gained from studying this specific oscillation phenomenon will be combined with our efforts to address the mechanisms that control the overall timing of development.



A: Data from two-photon real-time imaging experiments performed in a transgenic reporter mouse embryo. The fluorescence reflects transcriptional activity of the lunatic fringe gene. Note dynamic changes of fluorescence in tail part of embryo (white box). B: Quantification of fluorescent signal within the tail region (white box in A) identifies striking oscillations.

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:

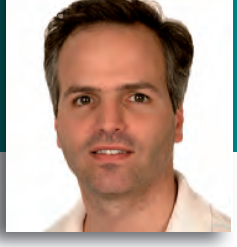
- identification of extrinsic and intrinsic signals controlling the timing of mouse development;
- generation of a real-time imaging reporter system for Wnt-signalling oscillations in mouse embryos using embryonic stem cell technology;
- analysis of the mechanisms underlying Wnt-signalling oscillations during embryogenesis;
- discovering novel oscillatory phenomena during embryogenesis.

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. & 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.

Developmental modulation of intracelluar trafficking during tissue morphogenesis

Previous and current research

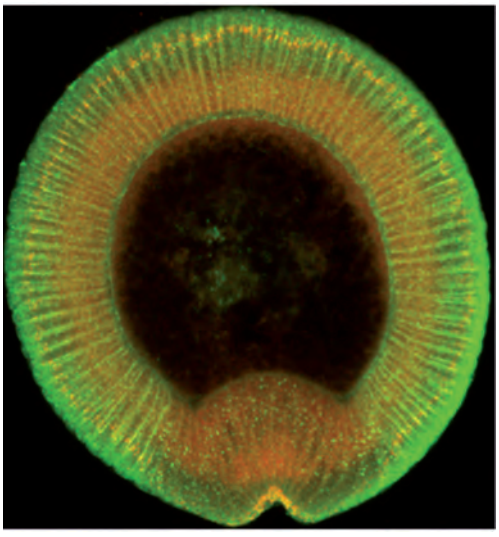
Our research focuses on how cells reorganise their cyto-architecture during tissue morphogenesis. Using a combination of imaging techniques and biochemistry, the group aims to understand how changes in protein and membrane dynamics regulate changes in cell morphology and tissue movement during development.

The early *Drosophila* embryo provides an excellent system for our studies. 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. 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 morphogenesis 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 imaging, genetics and biochemical approaches we wish to identify the cell biological basis underlying the pathways controlling changes in membrane dynamics in the early *Drosophila* embryo. Our long-term goal is to analyse the differentiation of intracellular pathways in other cell types and tissues as well. We wish to elucidate how machineries controlling intracellular trafficking reorient during differentiation and how this in turns impacts on global changes in tissue morphology.



Cross-section of a developing Drosophila *embryo showing polarised trafficking of Notch signaling components (ventral is down). The signaling receptor Notch is endocytosed (green dots) specifically in cells undergoing invagination (ventral furrow formation, 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

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.

Developmental patterning in plants

Previous and current research

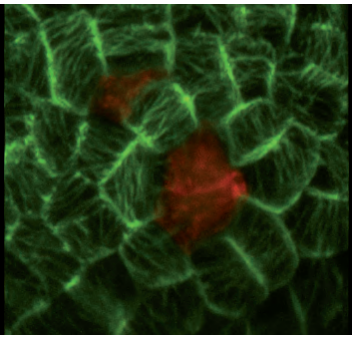
In addition to providing us with the air we breathe, the food we eat and much of the energy we use, plants exhibit a unique beauty associated with their strikingly symmetrical patterns of development. Multicellularity also evolved independently in plants giving us an opportunity to compare and contrast the developmental strategies used in different kingdoms. We investigate plant development by focusing 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.

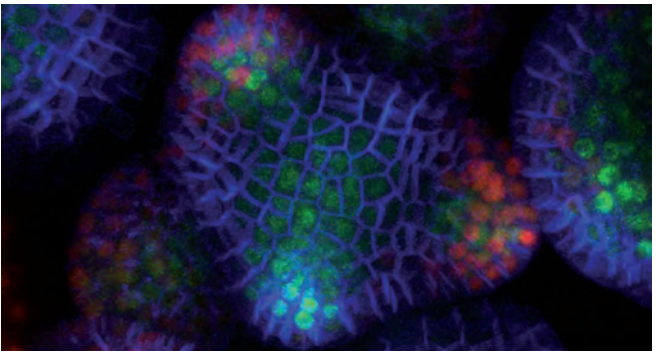
Future projects and goals

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.



Left: microtubules (green) form concentric alignments surrounding laser ablated plant cells (red).



Left: 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.*

Selected references

Hamant*, O., Heisler*, M.G., Jonsson*, H., Krupinski, P., Uyttewaal, M. *et al*. (2008). Developmental patterning by mechanical signals in *Arabidopsis*. *Science*, 322, 1650-1655

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, 3539-3548

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, 1633-1638

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, 1899-1911

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. Labeling 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.

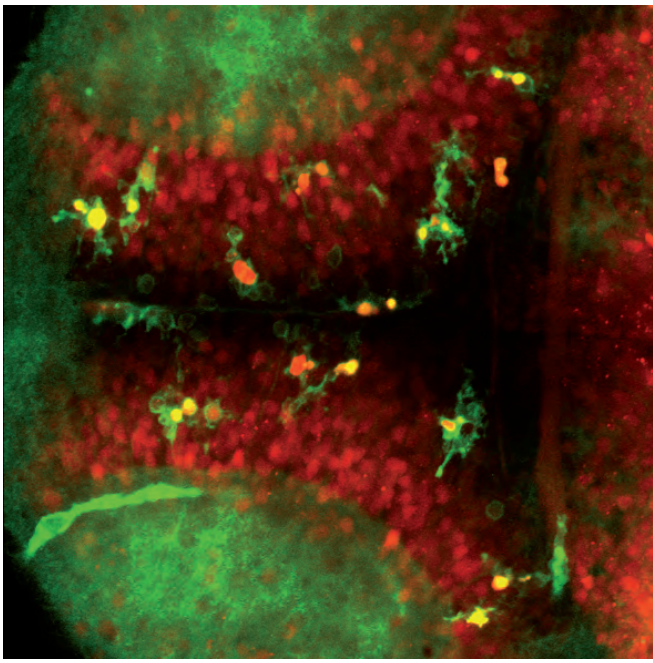
Future projects and goals

Despite the importance of microglia in several neuronal pathologies, many fundamental questions concerning microglialneuronal interactions remain unaddressed. How these cells collectively ensure that the entire brain is surveyed and how they react to damage with high precision is still entirely unknown.

Recent findings suggest that diffusible molecules such as lipids and nucleotides could attract microglia in response to neuronal apoptosis and injury, respectively. While these molecules can trigger dynamic changes in microglia motility *in vitro*, elucidating how their activity is controlled within the intact brain, both in space and time, remains the most important challenge in understanding this fascinating biological problem.

We aim to further exploit the massive imaging potential of the transparent zebrafish embryo for studying microglial biology *in vivo*. By combining forward and reverse genetic approaches with quantitative imaging technology, we will directly address the mechanisms underlying the attraction of microglia towards apoptotic, sick and injured neurons.

By applying cutting edge microscopy technology, such as the SPIM/DSLM (Selective Plane Illumination Microscopy) invented by Ernst Seltzer (page 17) and colleagues at EMBL Heidelberg, we will image all interactions between neurons and microglia and derive from this time-lapse analysis real quantitative data in a spatiotemporal manner.



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

François Spitz

PhD 1997, Institut Cochin de Génétique Moléculaire, Paris.

Postdoctoral research at the University of Geneva.

Group leader at EMBL since 2006.

Gene regulation and genome architecture in development and evolution

Previous and current research

Embryonic development is controlled by a large number of genes, whose expression levels and specificities are tightly regulated at the transcriptional level. In vertebrates, this is achieved for each gene through batteries of cis-regulatory elements, which act in modular, synergetic and complementary manners. These regulatory elements are often spread over large chromosomal domains that contain multiple genes with distinct functions and expression profiles. Hence, the mechanisms that translate such intermingled arrays of genes and cis-regulatory elements into coherent gene-specific expression programmes are playing a central role in controlling gene activity. The dramatic consequences of several human chromosomal rearrangements illustrate the importance of such mechanisms and the impact of the genome regulatory architecture on gene expression and function. Many recent genome-wide association studies and the extensive structural variations found in humans also underscore the importance for human phenotypic diversity of the mechanisms that control the interplay between genes and regulatory elements, and which are so far mostly elusive.

Our lab explores the nature of the vertebrate regulatory genome and aims to identify the underlying molecular mechanisms. As model systems, we focus on few megabase-large genomic loci whose organisation is extensively conserved within vertebrates, or where rearrangements have been associated with developmental genetic abnormalities. 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 to reshuffle their organisation in various ways, including modelling known human genetic disorders. By combining these approaches, we are starting to unravel the intricate regulatory organisation of these loci, as well as getting insights into the mechanisms that control and restrict the functional interactions between cis-regulatory elements and neighbouring genes. It is known that changes in chromatin structure, conformation and localisation within the nucleus are associated with different transcriptional activities, but the hierarchy of these events and how they are determined by the genomic sequences are both poorly understood. Our engineered mouse lines with altered chromosomal organisation allow us to investigate the relationships between genome sequence, chromatin structure and gene expression, for which we combine phenotypic and gene expression analysis with chromatin profiling and imaging. We are also interested in comparing the regulatory architecture of these regions in different species, to trace back its emergence during evolution and the associated evolutionary regulatory tinkering.

Future projects and goals

By combining mouse genetics with computational, biochemical and functional genomic approaches, we are aiming to:

- Examine the relationship between the genome sequence and its functional and structural conformation in the cell nucleus, as well as identify potential protein complexes involved in the associated mechanisms.
- Explore the functional and regulatory organisation of the mammalian genome using *in vivo* transposition and recombination approaches. In particular, we are interested in developing models of structural variations found in humans to unravel their phenotypic consequences.
- Identify the mechanisms associated with the emergence of novel cis-regulatory modules during evolution, using both computational and functional approaches. 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

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D.B. & Trono, D. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature*, 463, 237-240

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., 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

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

Genome Biology 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 Genome Biology 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 genome biology, 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 Heads of the Genome Biology Unit

Eileen Furlong

PhD 1996, University College Dublin.
Postdoctoral research at Stanford University.
Group leader at EMBL since September 2002.
Senior scientist and Joint Head of Genome Biology Unit since 2009.

Selected references

Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M. & Furlong, E.E. (2009). Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature*, 462, 65-70

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, 280-291

Jakobsen, J.S., Braun, M., Astorga, J., Gustafson, E.H., Sandmann, T., Karzynski, M. *et al.* (2007). Temporal ChIP-on-chip reveals Biniou as a universal regulator of the visceral muscle transcriptional network. *Genes Dev.*, 21, 2448-2460

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, 436-449

Regulatory networks driving cell fate decisions: 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 expression states, it is initially established through the integration of signalling and transcriptional networks converging on enhancer elements, or cis-regulatory modules (CRMs). 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 regulatory elements, how these modules are integrated to regulate more global cis-regulatory networks remains a key challenge. Even in the extensively studied model organism, *Drosophila*, 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 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.

Future projects and goals

Chromatin remodelling during cell fate decisions. We are currently developing a new system to investigate cell type specific changes in chromatin status using a number of genetic tools. This will allow changes in chromatin remodelling to be integrated with dynamic changes in transcription factor occupancy

Evolution of cis-regulatory networks. To gain a better understanding of the core functional features of the transcriptional network driving mesoderm specification we will take advantage of the fact that all of the key transcription factors involved are highly conserved at both a sequence and phenotypic level. We plan to extend the global transcription network generated in *Drosophila* melanogaster to other Drosophilds and non-arthropod species.

Predictive models of embryonic development. We have recently demonstrated that using only information on combinatorial occupancy of transcription factors is sufficient to predict spatio-temporal cis-regulatory activity. We plan to extend this analysis from predicting CRM activity to predicting a gene's expression. Ultimately our ultimate goal is to use this systems-level approach to make predictive models of embryonic development and the effect of genetic perturbations. Working in *Drosophila* allows us to readily test the predicted outcome of network perturbations on embryonic development.

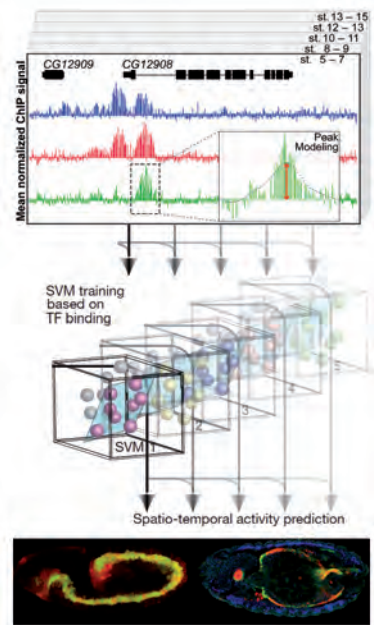


Figure 1 (left): Transcription factor occupancy is sufficient to predict spatio-temporal cis-regulatory activity (Zinzen et al., Nature *2009).*

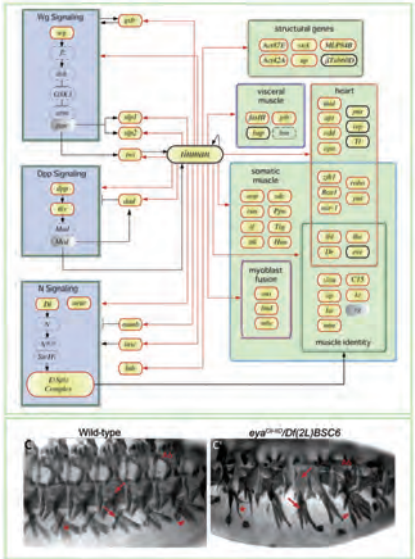


Figure 2 (left): The global transcriptional network reveals new functions for well characterised transcription factors (Liu et al., Dev. Cell *2009).*

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 Head of Unit since 2009.

Systems genetics

Previous and current research

Individuals differ at thousands of positions in the genome. 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 elucidating the genetic basis of resistance to malaria parasites in mosquitoes to the level of single alleles; studying the function of pervasive transcription of non-coding RNAs and the mechanisms of how they are generated; and genotyping single-nucleotide polymorphisms across entire yeast genomes to infer meiotic recombination-activity distributions that define trait inheritance.

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. Ultimately, by integrating genetics, genomics, systems biology and computational modeling 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.

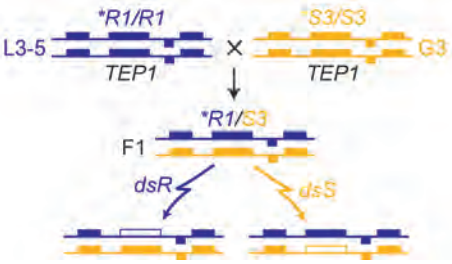


Figure 1 (left): Reciprocal allele-specific RNAi identifies the mosquito allele TEP1 as a major contributor to resistance to malaria parasites (Blandin et al.*, 2009,* Science*).*

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

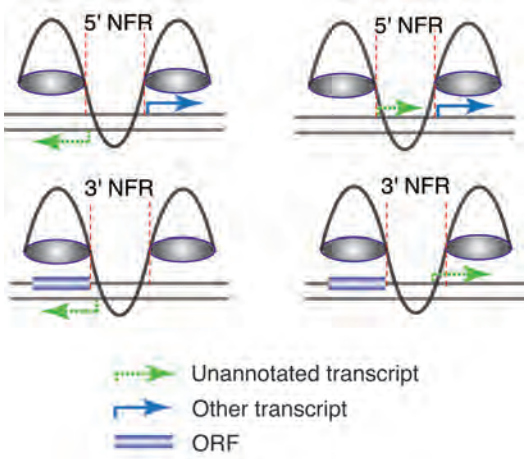
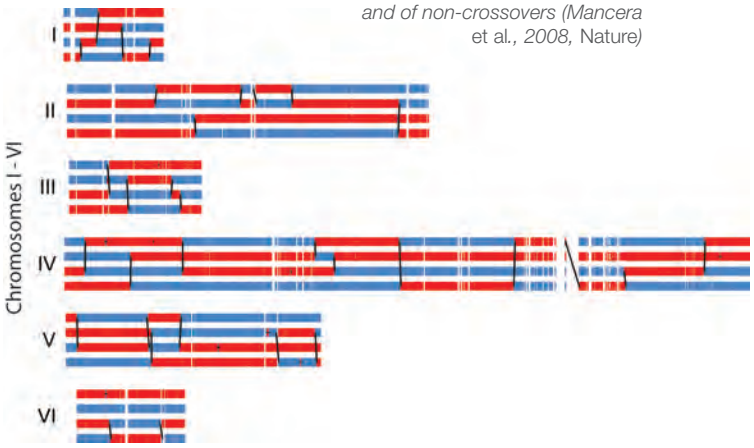


Figure 3 (below): High-resolution map of meiotic recombination identifies hotspots of crossovers and of non-crossovers (Mancera et al.*, 2008,* Nature*)*



Selected references

Blandin, S.A., Wang-Sattler, R., Lamacchia, M., Gagneur, J., Lycett, G., Ning, Y., Levashina, E.A. & Steinmetz, L.M. (2009). Dissecting the genetic basis of resistance to malaria parasites in *Anopheles gambiae*. *Science*, 326, 147-150

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, 1033-1037

Gagneur, J., Sinha, H., Perocchi, F., Bourgon, R., Huber, W. & Steinmetz, L.M. (2009). Genome-wide allele- and strand-specific expression profiling. *Mol. Syst. Biol.*, 5, 274

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

Wolfgang Huber

PhD 1998, Statistical Physics, University of Freiburg.

Postdoctoral research in cheminformatics at IBM Research, San Jose, California, 1998-99, and in microarray statistics at DKFZ Heidelberg, 2000-2004.

Group leader at EMBL-EBI 2004-2009.

Group leader at EMBL Heidelberg since 2009.

Selected references

Bourgon, R., Mancera, E., Brozzi, A., Steinmetz, L.M. & Huber, W. (2009). Array-based genotyping in *S. cerevisiae* using semi-supervised clustering. *Bioinformatics*, 25, 1056-62

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

Durinck, S., Spellman, P.T., Birney, E. & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.*, 4, 1184-1191

Anders, S. (2009). Visualisation of genomic data with the Hilbert curve. *Bioinformatics* 25, 1231-5

Computational Genetics

Previous and current research

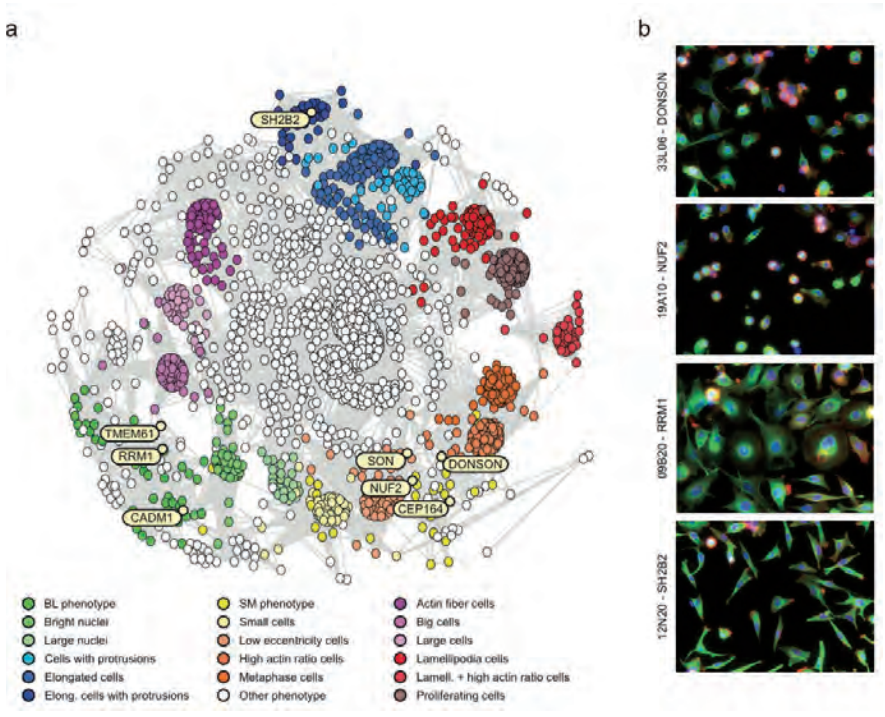
The group studies genotypes and phenotypes on a genome-wide scale: how do variations in the genomes of individuals shape their phenotypes? To this end, we develop computational methods in statistics, probabilistic modeling, image analysis and bioinformatics. We work with experimental labs in systems genetics and functional genomics to design and analyse genome-wide experiments, with the aim of unravelling the mechanisms of genetic inheritance, gene expression, molecular interactions, signal transduction and how they shape phenotypes. Most phenotypes, including many human diseases, are governed by large sets of genes and regulatory elements interacting in complex, combinatorial networks. Our aim is to map and quantitatively understand these complex systems; and, eventually, to devise strategies for engineering phenotypes.

Our research is stimulated by new technologies, and we employ data from high-throughput sequencing (RNA-seq, ChIP-seq, genotyping, polymorphism discovery), 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 of phenotypic outcomes based on individual genomes. To get there, we need a better understanding of the spectrum of genetic variations in a species, and how, combinatorially and together with environmental variations, they affect phenotype.

Progress in biology will continue to be driven by advances in technology. Sequencing technology will allow us to know the genomes of each individual person and model organism, and will provide transcription and DNA-protein interaction data for many different cellular systems at unprecedented depth. Light microscopy of single, life cells will provide data on molecular interactions and life-cycles within the cell, and on the dynamics of signalling, cell cycle, migration both under normal conditions and perturbations at large scale, systems level. We aim to develop the computational methods needed to understand this wealth of data and to help guide experimentation. Our emphasis lies on project-oriented collaborations with experimenters. We work on the methods in statistical computing, integrative bioinformatics and mathematical modelling to turn these data into biology.



Genomewide phenotypic map. a. Each of the 1,839 nodes represents an siRNA perturbation of cultured cells whose shape and morphology was monitored by automated microscopy. Nodes are linked by an edge when they are phenotypically similar. The graph is a two-dimensional representation of phenotypic diversity and similarity. b. Representative images for four siRNA perturbations. Cells were stained at the nuclei (DAPI, blue), actin (red) and tubulin (green).

Maja Köhn

PhD 2005 MPI for Molecular Physiology, Dortmund.
Postdoctoral research 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 signaling 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 in general 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 interested in phosphatases that promote diseases, focusing on protein tyrosine phosphatases (PTPs) and dual specificity phosphatases (DSPs). Furthermore, in collaboration with Mathieu Bollen at KU Leuven, Belgium, we are employing new ways to modulate the serine/threonine phosphatase PP1.

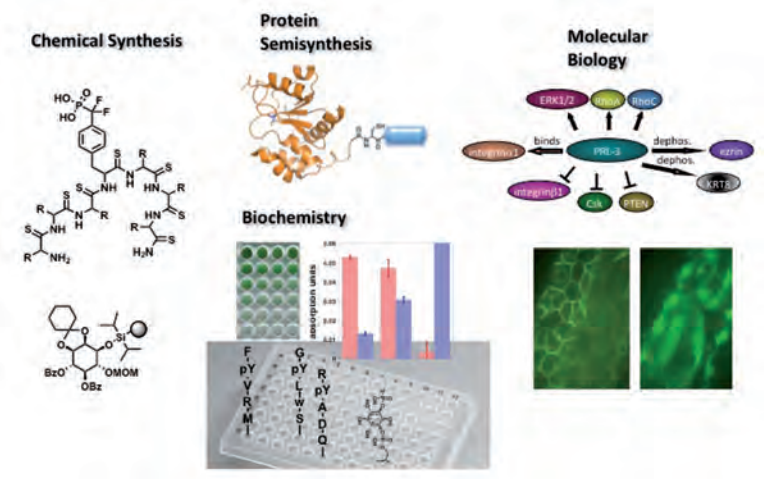
We are working on the design of inhibitors for PTPs based on chemical modification of protein/peptide substrates. Thus, upon binding, the phosphatase cannot fulfill its function and is bound to the modified substrate with natural high affinity. In this way, one does not have to rely on the random discovery of effector molecules by for example exhaustive screening of large compound libraries. In addition, we are looking into phosphoinositides as natural substrates of lipid phosphatases and DSPs. We are working on new synthetic strategies to simplify access to these compounds as well as their analogues to achieve a detailed picture of substrate specificities of these phosphatases in biochemical structure-activity relationship (SAR) studies.

The PRL family of phosphatases is of particular interest to us, because it is involved in several types of cancer. We apply protein semisynthesis and molecular biology approaches to obtain information about natural substrates, regulation and networks of these oncogenic phosphatases in cell culture and zebrafish embryos, the latter in collaboration with Darren Gilmour (page 9).

Future projects and goals

We are interested in further developing chemical methods to stabilise peptides as well as inositides and in working on novel cell penetration concepts. Another goal is to control and investigate the function and interactions of human lipid phosphatases in cells by applying the modulators resulting from our SAR studies. Developing effector molecules for the highly non-specific PSTPs is a long-term goal. The activity of these phosphatases is controlled not only by active-site specificity, but majorly 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.

The lab consists of an equal number of molecular biologists and organic chemists on the graduate student and postdoctoral level. The combination of molecular biology, biochemistry and synthetic chemistry not only opens up new ways to approach challenging phosphatase research, but also broadens the views and skills of every lab member.



Investigation of disease-promoting phosphatases.

Selected references

Weinrich, D., Köhn, M., Jonkheijm, P., Westerlind, U., Dehmelt, L. *et al*. (2010). Preparation of biomolecule microstructures and microarrays by thiol-ene photoimmobilization. *Chembiochem.*, 11, 235-247

McGinty, R.K., Köhn, M., Chatterjee, C., Chiang, K.P., Pratt, M.R. & Muir, T.W. (2009). Structure-activity analysis of semisynthetic nucleosomes: mechanistic insights into the stimulation of Dot1L by ubiquitylated histone H2B. *ACS Chem. Biol.*, 4, 958-968

Köhn, M. (2009). Immobilization strategies for small molecule, peptide and protein microarrays. *J. Pept. Sci.*, 15, 393-397

Köhn, M., Gutierrez-Rodriguez, M., Jonkheijm, P., Wetzel, S. *et al*. (2007). A microarray strategy for mapping the substrate specificity of protein tyrosine phosphatase. *Angew Chem. Int. Ed. Engl.*, 46, 7700-7703

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.

Selected references

Lam, H.Y., Mu, X.J., Stutz, A.M., Tanzer, A., Cayting, P.D. *et al.* (2010). Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. *Nat Biotechnol.*, 28, 47-55

Korbel, J.O., Tirosh-Wagner, T., Urban, A.E., Chen, X.N., Kasowski, M., Dai, L. *et al.* (2009). The genetic architecture of Down syndrome phenotypes revealed by high-resolution analysis of human segmental trisomies. *Proc. Natl. Acad. Sci. USA*, 106, 12031-12036

Hasin, Y., Olender, T., Khen, M., Gonzaga-Jauregui, C. *et al.* (2008). High-resolution copy-number variation map reflects human olfactory receptor diversity and evolution. *PLoS Genet.*, 4, e1000249

Korbel, J.O., Urban, A.E. *et al.* (2007). Paired-end mapping reveals extensive structural variation in the human genome. *Science*, 318, 420-426

Genome variation, dynamics and evolution

Previous and current research

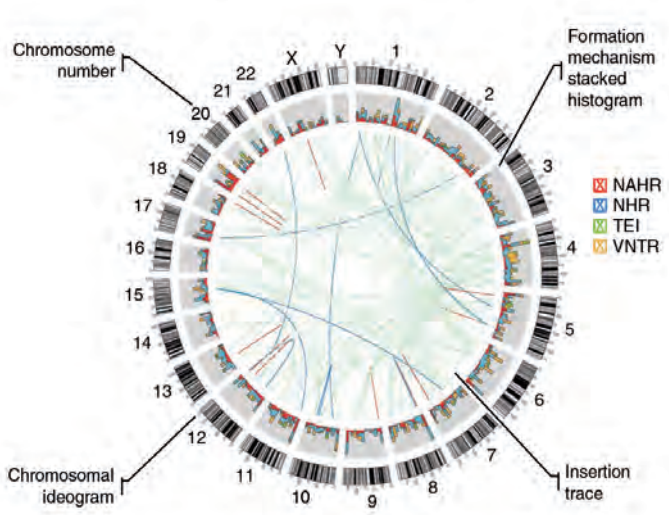
We are an interdisciplinary group that combines experimental and computational approaches for studying the extent, functional impact and mutational origins of genetic variation, with a focus on genomic structural variation (SV). SVs, also known as large (>1kb) copy-number variants (CNVs), translocations, and inversions, are responsible for most genetic variation in the human genome. Recent advances in massively parallel DNA sequencing enable us, for example, to decipher the impact of genetic variation by sequencing and analysing entire genomes. Our group has adopted a systems biology rationale in which computational biology research feeds into the experimental laboratory, and vice versa.

We recently developed high-resolution and massive paired-end mapping (PEM), an approach involving next-generation DNA sequencing of the end-stretches of genomic fragments and the massive alignment of these against a reference genome to globally map SVs. Our lab uses PEM, and recent extensions of the approach such as breakpoint junction-library analysis, to determine the extent of SVs in the genomes of healthy humans as well as in those suffering from cancer. Our central aims involve inferring the evolutionary and functional impact of SVs, and deducing the molecular mechanisms that cause SV-formation in the genome. For example, we recently learned through sequence analysis that the molecular mechanisms causing SV-formation mainly involve transposable elements as well as meiotic and DNA repair-associated recombination (e.g. non-allelic homologous recombination, as well as mechanisms that do not require homology, such as non-homologous end-joining). Our recent findings further involved the discovery of SVs on chromosome 21 as susceptibility factors for congenital heart disease. Furthermore, we are presently examining at genome-wide scale the effects of human genetic variations, particularly SVs, on the regulation of gene expression.

Future projects and goals

The extent to which genomes of healthy individuals differ due to SVs, the impact of SVs on common phenotypes, 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 high-throughput experiments and data-mining.

Furthermore, as members of the 1000 Genomes Project and the International Cancer Genome Consortium (for which we recently have begun carrying out human genome sequencing and data mining in an effort to analyse paediatric tumours), we are continuing the development of approaches for SV mapping and analysis. One goal is to facilitate the analysis and interpretation of personal genome sequencing data. Furthermore, our approaches may in the near future have an impact on personalised medicine, e.g. in helping facilitate the design and interpretation of next-generation sequencing based studies for associating phenotypes with genetic variation.



Abbreviations: NAHR, non-allelic homologous recombination; NHR, non-homologous recombination; TEI, transposable element insertion; VNTR, variable number of tandem repeat expansion

Left: origin of recent insertion variants into the human genome.

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 on mass spectrometry-based proteomics. Mass spectrometry coupled to liquid chromatography has matured to the stage 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 targeting 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 approach combines biochemistry, analytical chemistry, mass spectrometry and bioinformatics. Specifically, we use 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 focusing (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.

We have three main topics of biological interest. One is developmental biology, with a particular focus on stem cell biology. Over the past years, we have studied the dynamics of protein expression during differentiation of embryonic stem cells. We have also studied how protein phosphorylation changes as ES cells leave the pluripotent state, giving insight in (in)activation of phosphorylation networks. Currently, we are currently expanding our technology to iPS cells.

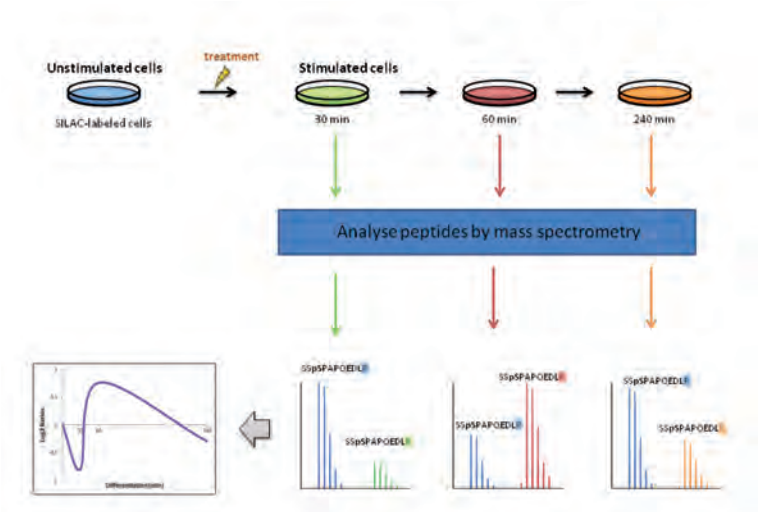
A second topic is in the area of transcriptional regulation. We are interested in the interaction of proteins with DNA, both with regards to the identification of transcription factors, along with their associated proteins, and to specific regulatory elements. More broadly, we are investigating the dynamics of chromatin, with a particular interest in the proteins that are not considered to belong to the core, but which might play an important regulatory role in the response to changing environmental conditions.

In a third research line, we investigate protein synthesis and turnover in yeast and mammalian cells, which is a missing link in our understanding of protein regulation.

Future projects and goals

Our future work can be divided into three major areas:

- We will apply quantitative proteomic techniques for global analysis of protein expression during embryonic development (particularly stem cells).
- We will focus on the identification of protein-protein and protein-DNA complexes to understand regulatory principles of transcription under various biological conditions.
- We will develop biochemical and bioinformatic tools to analyse and understand protein turnover.



Generalized experimental workflow for the quantitative analysis of protein expression in treated vs. untreated cells. Cells that are labeled with heavy isotopes (SILAC) are mixed with non-labelled cells that have recieved a treatment (e.g. induction to differentiate). Quantitative analysis of peptide extracts by mass spectrometry results in the identification of thousands of peptides whose expression level can be quantified over time.

Selected references

Van Hoof, D., Munoz, J., Braam, S.R., Pinkse, M.W., Linding, R., Heck, A.J., Mummery, C.L. & Krijgsveld, J. (2009). Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell*, 5, 214-226

Gouw, J.W., Pinkse, M.W., Vos, H.R., Moshkin, Y., Verrijzer, C.P., Heck, A.J. & Krijgsveld, J. (2009). *In vivo* stable isotope labeling of fruit flies reveals post-transcriptional regulation in the maternal-to-zygotic transition. *Mol. Cell Proteomics*,

Van Hoof, D. *et al.* (2007). An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nat. Methods*, 4, 677-678

Van Hoof, D. *et al.* (2006). A quest for human and mouse embryonic stem cell-specific proteins. *Mol. Cell Proteomics*, 5, 1261-1273

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.

Chromatin-based remodelling of genome function

Previous and current research

Chromatin packages our genome. Whenever cells change their genetic programme, such as during development or upon environmental changes, chromatin needs to be remodelled. Our team identifies, characterises and exploits novel molecular mechanisms that underlie the plasticity of chromatin structure.

We focus on the role of distinct post-translational modifications, cellular metabolites and signalling-activated enzymes in regulating the assembly and remodelling of histone proteins in the nucleosome, the fundamental repeating unit of chromatin structure. For example, we discovered the first example of a protein module capable of recognising a post-translational modification in a histone protein (the so-called bromodomain) and reported the first example of metabolite binding function in a histone and other nuclear proteins (the so-called macrodomain). By carrying out this research, we are uncovering the fundamental mechanisms that govern the function of our complex genome and its ability to adapt to new environments.

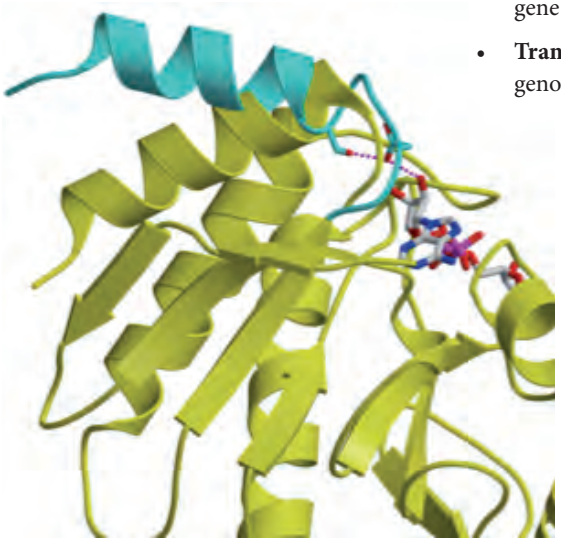
Our approach is defined by a multidisciplinary combination of genetics, genomics, biochemistry, cell biology, biophysics, structural biology (X-ray crystallography) and the use of selected model organisms. This allows us to answer fundamental biological questions in chromatin biology and to identify novel paradigms of molecular recognition and biological regulation in a comprehensive manner.

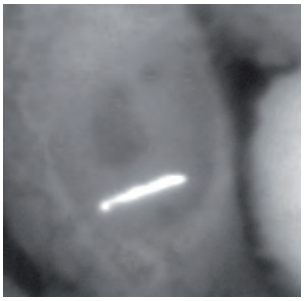
Our current research is focused on three complementary areas of chromatin plasticity. We are dissecting the structure and function of important chromatin remodelling enzymes using a combination of high-resolution X-ray crystallography, biochemistry, protein engineering and *in vivo* approaches. Secondly, we are studying the biological role and cell biology of ADP-ribosylation, a post-translational modification involved in regulating chromatin structure and transcription upon a variety of environmental stresses, where we have pioneered the discovery of the ADP-ribose-sensing macrodomain proteins. Last but not least, we complement our studies in yeast and mammalian cells by studying memory formation in the fruit fly. By applying knowledge from the field of transcription and epigenetics, we are now dissecting the role of chromatin dynamics in the formation, consolidation and maintenance of organismal memory.

Future project and goals

- **Structure and function of chromatin remodelling nanomachines:** histone chaperones and signalling-activated enzymes;
- **Molecular dissection of cellular ADP-ribosylation:** metabolic control of gene activity associated with ADP-ribosyl-recognition;
- **Transcriptional and epigenetic basis for organismal long-term memory:** genomic tools for the systematic dissection of tissue-specific gene activity.

Below: the macrodomain is an ADP-ribose-binding module. We pioneered the identification of a domain that specifically recognises this and related nuclear NAD metabolites.





Left: macrodomains rapidly respond upon DNA damage activation by relocalizing to nuclear sites of PARP1 activity.

Selected references

Sporn, J.C., Kustatscher, G., Hothorn, T., Collado, M., Serrano, M. *et al.* (2009). Histone macroH2A isoforms predict the risk of lung cancer recurrence. *Oncogene*, 28, 3423-8

Gottschalk, A.J., Timinszky, G., Kong, S.E., Jin, J. *et al.* (2009). Poly(ADP-ribosyl)ation directs recruitment and activation of an ATP-dependent chromatin remodeler. *Proc. Natl. Acad. Sci. USA*, 106, 13770-13774

Timinszky, G., Till, S., Hassa, P.O., Hothorn, M., Kustatscher, G. et al. (2009). A macrodomain-containing histone rearranges chromatin upon sensing PARP1 activation. *Nat. Struct. Mol. Biol.*,16, 923-9

Hothorn, M., Neumann, H., Lenherr, E.D., Wehner, M., Rybin, V. *et al.* (2009). Catalytic core of a membrane-associated eukaryotic polyphosphate polymerase. *Science*, 324, 513-516

Christoph A. Merten

PhD 2004, University of Frankfurt.

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

Junior group leader at the Institut de Science et d'Ingénierie Supramoléculaire, Strasbourg.

Group leader at EMBL since 2010.

Microfluidic systems for large-scale screens in biology and chemistry

Previous and current research

We are conducting multidisciplinary research at the interface between biology, chemistry and engineering. The overall goal is to develop enabling technology for large-scale screens in the field of medical biotechnology, developmental biology and synthetic chemistry.

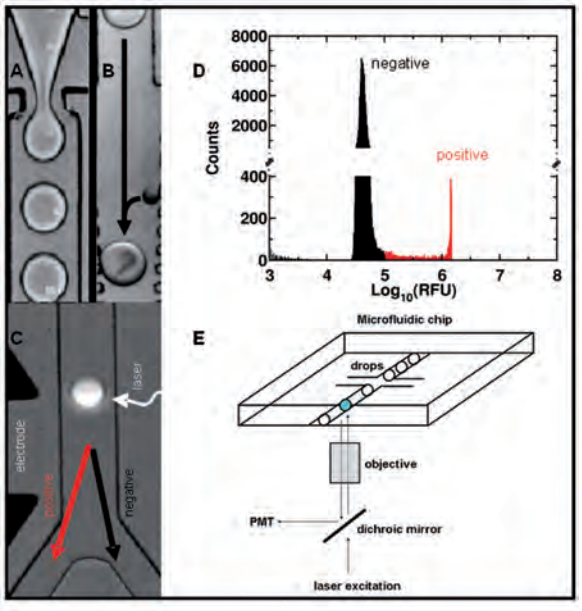
The research in our group focuses on novel microfluidic approaches for biological and chemical applications. Recently, we have established droplet-based microfluidic platforms for the incubation and screening of human cells and multicellular organisms. In these systems, aqueous droplets within an immiscible oil phase serve as miniaturised reaction vessels. Compared to conventional microtitre plate formats, this technology allows massively increased throughput (up to 500 samples per second can be processed) and more than 1000-fold smaller sample volumes (pico- to nanoliters). The miniaturisation not only enables the use of highly valuable samples which generally cannot be obtained on the scale required for high-throughput screening (e.g. primary cells, patient material), but also facilitates assays on the single cell/single animal level.

For screening purposes, we have also developed systems allowing the co-encapsulation of different biomolecules or chemical compounds into the droplets. This can be achieved by using cell libraries in which each individually encapsulated cell releases a different protein variant (e.g. antibodies, peptides), by interfacing robotic systems injecting different compounds into the microfluidic chips, or by encapsulating beads displaying immobilised compound libraries (one-bead–one-compound libraries).

Future projects and goals

Now that we can rapidly generate chemically-distinct droplets, our future research will focus on applications in biology and chemistry. Since the droplet-based technology allows the mixing of high sample numbers in a truly combinatorial fashion (by droplet fusion), we will put special emphasis on combinatorial screens. In particular, we plan:

- **combinatorial drug screens.** We want to identify drug combinations minimising the number of non-responding cells. This is of special interest for highly heterogeneous populations such as tumour cells, in which individual cells that do not respond to a given drug can give rise to a lethal outcome (e.g. tumour regrowth).
- **combinatorial RNAi screens.** Our systems should allow the monitoring of systematic double knockouts on a genome-wide scale; interactions that can not be revealed by silencing individual genes become visible.
- **combinatorial chemistry.** The possibility of rapidly generating and mixing huge sample numbers should allow the exploration of large areas of chemical structure space in search of new bioactive molecules.



Encapsulation of human cells into 660 pL droplets at rates of up to 800s-1 (A). Addition of further compounds to each individual sample by droplet fusion (B). Droplet sorting based on fluorescence intensities (C, D). Optical setup (E).

Selected references

Clausell-Tormos, J., Griffiths, A.D. & Merten, C.A. (2010). An automated two-phase microfluidic system for kinetic analyses and the screening of compound libraries. *Lab Chip*, in press

Upert, G., Merten, C.A. & Wennemers, H. (2010). Nanoliter Plates – versatile tools for the screening of split-and-mix libraries on-bead and off-bead. *Chem. Comm.*, in press

Köster, S., Angilè, F.E., Duan, H, Agresti, J.J., Wintner, A. *et al.* (2008) Drop-based microfluidic devices for encapsulation of single cells. *Lab Chip.,* 8, 1110-1115

Clausell-Tormos, J., Lieber, D., Baret, J.-C., El Harrak, A., Miller, O.J. *et al.* (2008). Droplet-based microfluidic platforms for the encapsulation and screening of mammalian cells and multicellular organisms. *Chem. Biol.,* 15, 427–437 (research highlight in *Nat. Methods,* 5, 580-581)

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 focuses on the molecular mechanisms by which chromatin proteins encoded by the Polycomb group (PcG) and the trithorax group (trxG) of genes maintain transcriptional states of target genes. PcG and trxG proteins are two evolutionary conserved sets of regulatory factors 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 to understand their molecular mechanisms. We found that PcG protein complexes contain enzymatic activities that add or remove particular post-translational modifications at specific lysine residues in histone proteins. These include 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 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 the 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

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*. Examples of current studies in the lab include dissecting the mechanisms of chromatin complexes *in vitro*, measuring the stability of post-translational modifications on histone proteins *in vivo*, analysing the role of protein GlcNAcylation in Polycomb repression, and functional testing of the 'histone code' in *Drosophila*. Our long-term goal is to understand how gene transcription states are controlled by the Polycomb /trithorax system and how they are propagated through replication and cell division.

Selected references

Gambetta, M.C., Oktaba, K. & Muller, J. (2009). Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. *Science*, 325, 93-6

Muller, J. & Verrijzer, P. (2009). Biochemical mechanisms of gene regulation by polycomb group protein complexes. *Curr. Opin. Genet. Dev.*, 19, 150-158

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

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

Figure 1 (above): Polycomb protein complex purified from Drosophila *embryos (Nekrasov* et al.*, 2007,* EMBO J.*).*

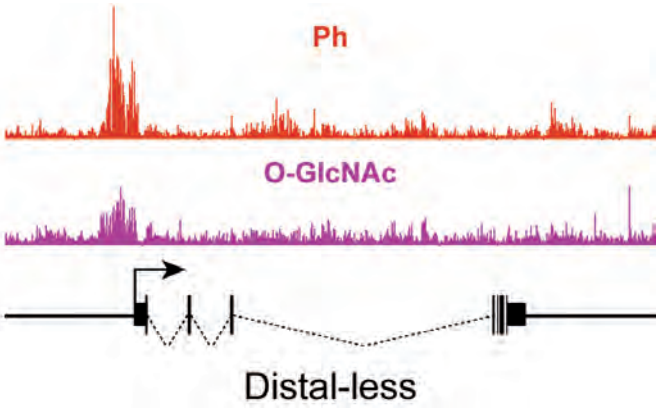


Figure 2 (right): Genome-wide profiling reveals co-localisation of the sugar modification GlcNAc and the Polycomb protein Ph at target genes (Gambetta et al.*, 2009,* Science*).*

Structural and Computational Biology Unit

The Structural and Computational Biology Unit pursues an ambitious research programme with a strong basis in integrated structural systems biology and a far-reaching computational component that bridges into various areas of biology. A wide spectrum of expertise allows the unit to tackle problems at different ranges of spatial 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 unit pursue several common large projects that require the input of different skill sets. One example is the elucidation of a comprehensive structural and temporal description of an entire cell at almost molecular resolution. It goes hand in hand with the application of and integration of various '-omics' approaches to the small bacterium *Mycoplasma pneumoniae*, 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, Genome Biology (page 34) and the Nédélec group, Cell Biology (page 14)).

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 – including a recently purchased high-throughput Titan Krios microscope – and scanning micro-densitometers for single particle cryo-electron microscopy and cryo-electron tomography. The unit also has facilities for single-molecule light microscopy, isothermal calorimetry, circular dichroism 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 Heads of the Structural and Computational Biology Unit

Peer Bork

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

Selected references

Yamada, T. & Bork, P. (2009). Evolution of biomolecular networks: lessons from metabolic and protein interactions. *Nat. Rev. Mol. Cell Biol.*, 10, 791-803

Gianoulis, T.A., Raes, J., Patel, P.V., Bjornson, R., Korbel, J.O. *et al.* (2009). Quantifying environmental adaptation of metabolic pathways in metagenomics. *Proc. Natl. Acad. Sci. USA*, 106, 1374-1379

Kühner S., van Noort V., Betts M.J., Leo-Macias A., Batisse C., Rode M., Yamada T., Maier T., Bader S., Beltran-Alvarez P. *et al.* (2009). Proteome organization in a genome-reduced bacterium. *Science*, 326, 1235-1240

Campillos, M., Kuhn, M., Gavin, A.C., Jensen, L.J. & Bork, P. (2008). Drug target identification using side-effect similarity. *Science*, 321, 263-266

Deciphering function and evolution of biological systems

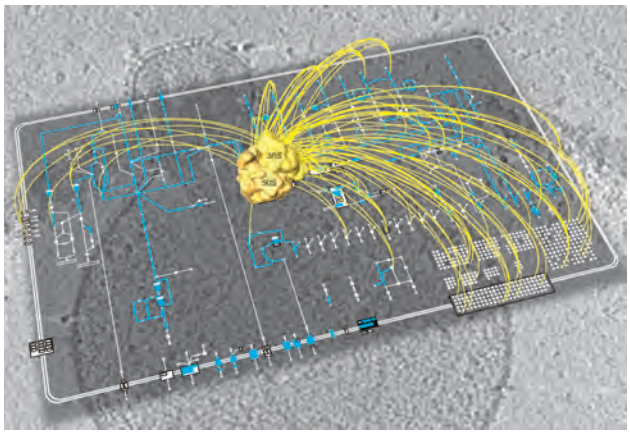
Previous and current research

The main focus of our computational biology group is to gain insights into biological systems and their evolution by comparative analysis and integration of complex molecular data. The group currently works on three different spatial scales, but with common underlying methodological frameworks:

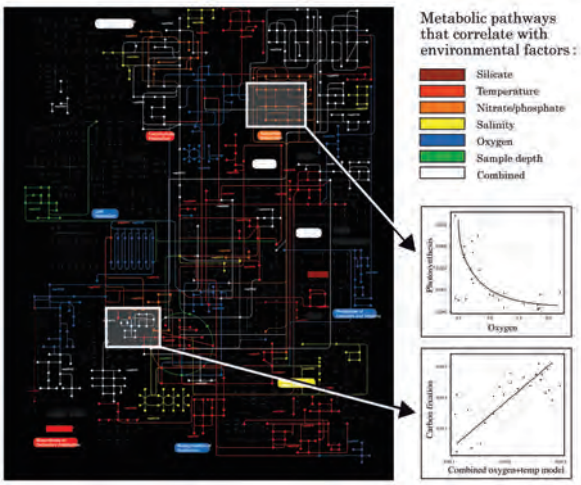
- genes, proteins and small molecules;
- networks and cellular processes;
- phenotypes and environments, often related to diseases.

We are aiming at biological discoveries and often develop tools and resources to make this happen. We usually work in new or emerging areas of biology; for example we have projects that integrate drugs (and other small molecules) with cellular and phenotypic information to predict new uses for old drugs (e.g. Campillos *et al.*, 2008, *Science*) or find biomolecules that cause disease or side effects. We study temporal and spatial aspect of protein networks to identify biological principles that determine function and evolution (e.g. de Lichtenberg *et al.*, 2005, *Science*; Jensen *et al.*, 2006, *Nature*; Kuehner *et al.*, 2009, *Nature*). We also trace the evolution of the animal gene repertoire (e.g. Ciccarelli *et al.*, 2006, *Science*) and, for example, connect gene losses and duplications with morphological or life style changes. We study environmental aspects via comparative metagenomics (Tringe *et al.*, 2005, *Science*; von Mering *et al.*, 2007, *Science*; Qin *et al.*, 2010, *Nature*) and hope to find marker genes for various diseases like obesity but also to understand microbial community interactions. All our projects 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 Heidelberg university.



Integration of various -omics data from a genome-reduced bacterium, Mycoplasma pneumonia*. Together with other SCB groups, we overlay genomic, transcriptomic, proteomic, metabolic and structural data to establish a model organism for systems biology and discover lots of exciting biology on the way (see Kuehner* et al.*, 2009, Guell* et al.*, 2009 and Yus* et al.*, 2009, all* Science*). The figure depicts a tomographic snapshot, a single particle EM of the ribosome (many proteins of which have unexpected links to various cellular processes indicated by lconnectors) and a metabolic reconstruction in which the correspondence to operon organisation is shown (blue).*



Integration of metagenomics data with environmental factors. Using novel visualization concepts and statistical approaches we can correlate the abundance of molecular functions to external data (e.g. Gianoulis et al.*, 2009,* PNAS*; Qin* et al.*, 2010,* Nature*). For example, many distant ocean samples are analysed and the abundance of some pathways significantly correlate with temperature or oxygen concentration of both. In human, we find correlations of gut genes from metagenomes with several diseases.*

Christoph Müller

PhD 1991, University of Freiburg.

Postdoctoral work at Harvard University, Cambridge, Massachusetts.

At EMBL Grenoble since 1995.

Joint Head of Unit at EMBL Heidelberg since 2007.
Joint appointment with the Genome Biology Unit.

Molecular mechanisms of transcriptional regulation and epigenetics

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 aim to gain insight into how DNA sequence information and epigenetic modifications act together to regulate gene transcription.

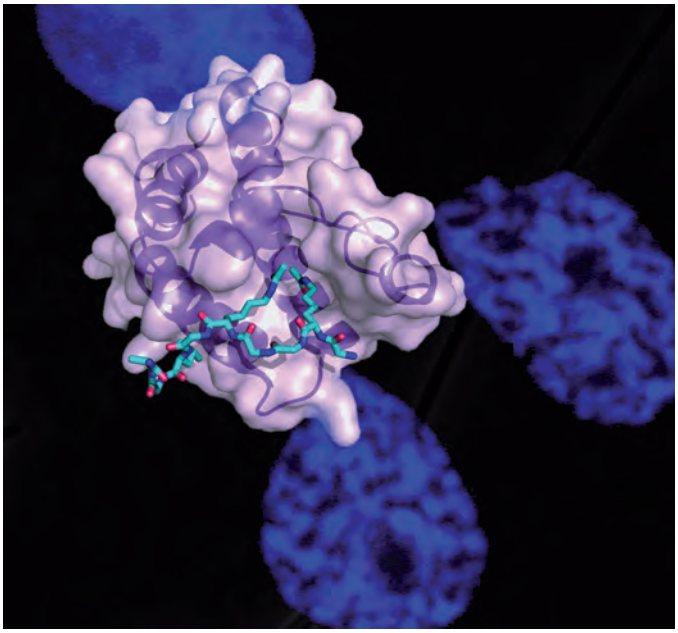
To achieve these goals, we use structural information obtained by X-ray crystallography and electron microscopy combined with other biophysical and biochemical techniques. Systems currently under investigation include transcription factor/DNA complexes, yeast RNA polymerase III and multi-protein complexes involved in chromatin targeting, remodelling and histone modifications.

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. 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.

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.

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 cell biology, biochemical and structural biology approaches.



Crystal structure of the first bromodomain of Brdt that cooperatively recognises two acetylated lysine residues. During spermatogenesis binding of hyper-acetylated histone tails to Brdt causes chromatin inside the nucleus to compact and clump together (stained blue inside the nuclei of two cells in the background image).

Selected references

Moriniere, J., Rousseaux, S., Steuerwald, U., Soler-Lopez, M., Curtet, S., Vitte, A.L., Govin, J. *et al.* (2009). Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature*, 461, 664-668

Grimm, C., Matos, R., Ly-Hartig, N., Steuerwald, U., Lindner, D., Rybin, V., Muller, J. & Muller, C.W. (2009). Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. *EMBO J.*, 28, 1965-77

Hames, C., Ptchelkine, D., Grimm, C., Thevenon, E., Moyroud, E., Gerard, F. *et al.* (2008). Structural basis for LEAFY floral switch function and similarity with helix-turn-helix proteins. *EMBO J.*, 27, 2628-2637

Fernandez-Tornero, C., Bottcher, B., Riva, M., Carles, C. *et al.* (2007). Insights into transcription initiation and termination from the electron microscopy structure of yeast RNA polymerase III. *Mol. Cell*, 25, 813-823

PhD 2005, Eötvös Loránd University, Budapest, Hungary.

Postdoctoral research at the National Institutes of Health, Bethesda, USA.

Group leader at EMBL since 2009.

Orsolya Barabas

Mechanism of DNA recombination and its applications for research and therapy

Previous and current research

Our lab is interested in how DNA recombination is carried out by various systems, with a particular focus on DNA transposons, as these mobile genetic elements offer attractive genetic tools and are starting to be applied as gene delivery vehicles in mammalian functional genomics and human gene therapy. To support the development of transposon-based genetic tools we study their mechanism of movement using structural biology (mainly X-ray crystallography), molecular biology and biochemistry techniques and cell culture assays.

There are major problems to overcome before gene therapy can become a common medical technique. There is a need for gene delivery vehicles that provide safe and efficient integration of the therapeutic gene into the host chromosome and allow long-term expression without causing harmful DNA rearrangements.

DNA transposons are mobile DNA elements that have the ability to move from one genomic location to another. They contain specific DNA sequences at their ends, and encode a transposase enzyme that catalyses all necessary DNA cleavage and joining reactions. Transposons can also be engineered to carry desired genetic information, offering stable and heritable modifications of a target genome. We aim to improve our understanding of the mechanism of transfer, target site selection and cellular control of various DNA transposition systems on molecular and cellular levels, and two major current interests include the Sleeping Beauty (SB) transposon system and site specific bacterial transposons.

Due to its relatively high transposition efficiency in human cells SB, which was reconstructed from fish genomes, has been successfully applied to identify oncogenes and tumour suppressors (Dupuy *et al.*, 2005), and to characterise unknown genes. More recently, SB is being applied in a gene therapy clinical trial (Williams, 2008). However, the mechanism of SB transposition and its interactions with the host have yet to be unravelled. Our ongoing structural and *in vitro* functional studies wil offer a mechanistic understanding and invaluable insights for rational design of transposition cassettes.

One of the main obstacles of gene therapy is the occurrence of harmful genome perturbations due to integration at unwanted locations. Site specific recombinases may offer a solution, and our recent work revealed the mechanism of the bacterial Insertion Sequence IS608 that integrates site specifically at short DNA sequences. We found that the transposase co-opts a segment of the transposon DNA to be part of its active site, and uses it to recognise the site of insertion (Barabas *et al.,* 2008). A consequence of the observed target site recognition strategy is that the site of insertion can be easily altered by changing a few nucleotides in the transposon (Guynet *et al.*, 2009). This DNA based recognition might also be expandable with the aim of targeting longer potentially unique genomic sites. We are currently investigating this possibility.

We are also studying a newly discovered site specific mobile element, the so-called plasticity zone transposon (TnPZ) from *Helicobacter pylori* (Kersulyte *et al.*, 2009). This mobile element appears to move with a mechanism similar to bacterial conjugation and integrates to 7nt long specific sites. We are curious to find out how this transposon moves and selects its target site.

The structure of the IS608 transpososome, modelled based on a series of crystal structures.

Future projects and goals

- We will initiate structural and mechanistic studies with the site-specific eukaryotic transposons, the Helitrons.
- We will select representative targets from several transposon and recombinase families to capture a broader picture of the alternative recombination pathways applied by nature.
- In general, we are fascinated by the enormous information content of DNA and the sophisticated machinery maintaining, interpreting and manipulating this code. In the long term, we intend to broaden the range of our projects in these directions.

Selected references

Guynet, C., Achard, A., Hoang, B.T., Barabas, O., Hickman, A.B., Dyda, F. & Chandler, M. (2009). Resetting the site: redirecting integration of an insertion sequence in a predictable way. *Mol. Cell.,* 12, 612-9

Kersulyte, D., Lee, W., Subramaniam, D., Anant, S., Herrera, P. *et al*. (2009). *Helicobacter pylori*'s plasticity zones are novel transposable elements. *PLoS One*, 3, e6859

Palazzoli, F., Carnus, E., Wells, D.J. & Bigot, Y. (2008). Sustained transgene expression using non-viral enzymatic systems for stable chromosomal integration. *Curr Gene Ther.*, 8, 367-90

Barabas, O., Ronning, D.R., Guynet, C., Hickman, A.B., Ton-Hoang, B., Chandler, M. & Dyda, F. (2008). Mechanism of IS200/IS605 family DNA transposases: activation and transposon-directed target site selection. *Cell,* 25, 208-20

Martin Beck

PhD 2006, Max-Planck-Institute of Biochemistry, Martinsried, Germany.

Postdoctoral research at the Institute for Molecular Systems Biology, ETH Zurich, Switzerland.

Group leader at EMBL since 2010.

Structure and function of large macromolecular assemblies

Previous and current research

Research in our laboratory combines biochemical approaches, proteomics and cryo-electron microscopy to study the structure and function of large macromolecular assemblies.

Cryo-electron tomography is the ideal tool to observe molecular machines at work in their native environment (figure 1). In combination with single particle analysis and averaging techniques the overall structure of macromolecular assemblies can be determined (figure 2). Since the attainable resolution of the resulting three-dimensional maps is moderate, the challenge ahead is to integrate information provided by complementary techniques and, in particular, to bridge the resolution gap towards the high resolution techniques (NMR, X-ray crystallography).

Proteomics approaches can provide the auxiliary information that is necessary to tackle this challenge. Targeted mass spectrometry can handle complex protein mixtures and, in combination with heavy labelled reference peptides, provides quantitative information about protein stoichiometries within macromolecular assemblies. Together with cross-linking techniques the protein interfaces are revealed. The spatial information obtained in this way facilitates the fitting of high-resolution structures into cryo-EM maps in order to build atomic models of entire molecular machines.

Megadalton protein complexes are involved in a number of fundamental cellular processes such as cell division, vesicular trafficking and nucleocytoplasmic exchange. In most cases such molecular machines consist of a multitude of different proteins that can occur in several copies within an individual assembly. Studying their structure and function is a challenging task, not only due to their compositional complexity, but also because of their sheer size that, in many cases, makes them inaccessible to biochemical purification.

We believe that the overall structure of intricate megadalton complexes can be elucidated through (i) studying isolated protein subcomplexes that make up individual building blocks and (ii) understanding the stepwise assembly or disassembly process.

Future projects and goals

Our goals are:

- to develop integrated workflows for structure determination of large macromolecular assemblies such as the nuclear pore complex (figure 2);
- to study their function by imaging them in action;
- to reveal individual steps of their assembly and disassembly process.

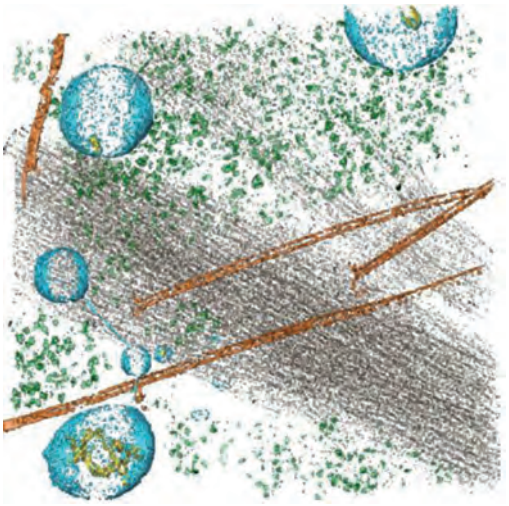


Figure 1. Cryo electron tomogram of a fraction of the cytoplasm of a human cell. Microtubules are colored in orange, stress fibres in grey, protein complexes in green, membranes in cyan and vesicular contents in yellow.

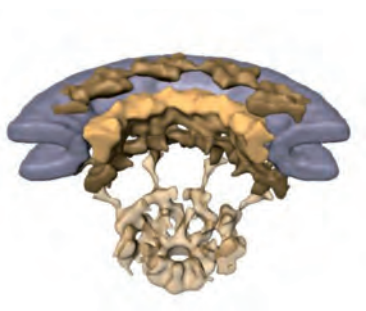


Figure 2. Structure of the nuclear pore complex. Membranes are colored in grey, the scaffold structure in yellow and the nuclear basket in transparent brown.

Selected references

Beck, M., Malmström, J., Lange, V., Schmidt, A., Deutsch, E. & Aebersold, R. (2009). Visual proteomics of the human pathogen *Leptospira interrogans*. *Nature Methods,* 6, 11, 817-23

Malmström, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E. & Aebersold, R. (2009). Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature,* 460, 762-765

Beck, M., Lucic, V., Forster, F., Baumeister, W. & Medalia, O. (2007). Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature*, 449, 611-5

Beck, M., Forster, F., Ecke, M., Plitzko, J.M., Melchior, F., Gerisch, G., Baumeister, W. & Medalia, O. (2004). Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science*, 306, 1387-90

PhD 2004, Oxford University.

Postdoctoral research at the University of Munich.

Group leader at EMBL since 2006.

John Briggs

Enveloped viruses and coated vesicles – cryo-electron microscopy and tomography

Previous and current research

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.

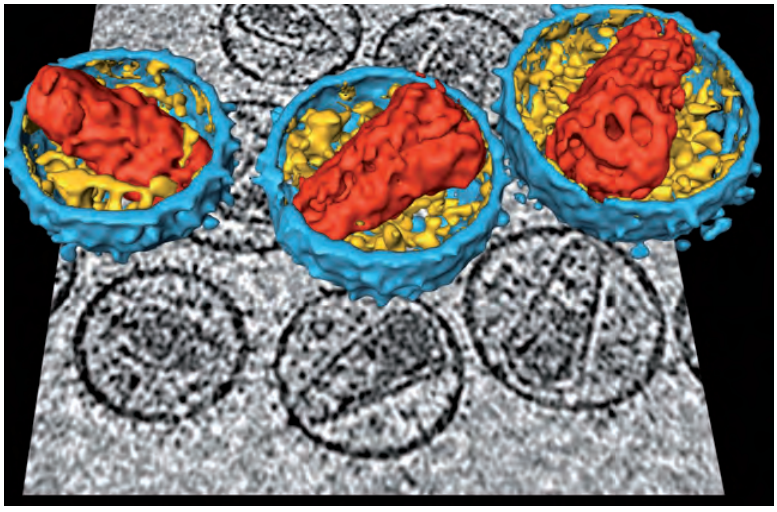
A particular emphasis of our research is the structure and life-cycle of asymmetric membrane viruses such as HIV. The assembly of the virus particles and their subsequent fusion with target cells offer insights into general features of vesicle budding and membrane fusion.

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.

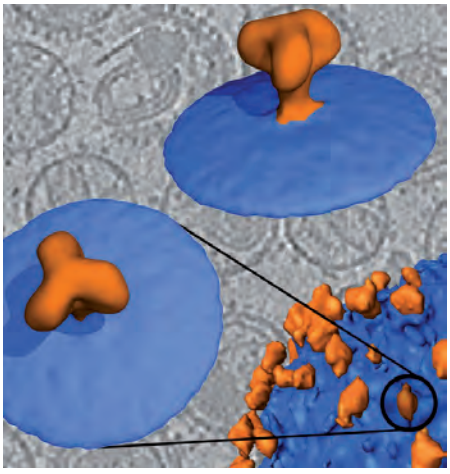
We take a step-by-step approach to understanding the native structure. Fluorescence microscopy can by used to locate and characterise features of interest. Three-dimensional reconstructions of these features can be obtained using cellular electron tomography of the biological system in its native state. 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.



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



3D reconstruction of the SIV glycoprotein spike, generated by averaging sub-tomograms extracted from whole virus tomograms.

Selected references

Zanetti, G., Riches, J.D., Fuller, S.D. & Briggs, J.A. (2009). Contrast transfer function correction applied to cryo-electron tomography and sub-tomogram averaging. *J. Struct. Biol.*, 168, 305-12

Briggs, J.A., Riches, J.D., Glass, B., Bartonova, V., Zanetti, G. & Krausslich, H.G. (2009). Structure and assembly of immature HIV. *Proc. Natl. Acad. Sci. USA*, 106, 11090-5

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

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 Genome Biology Unit.

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

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.

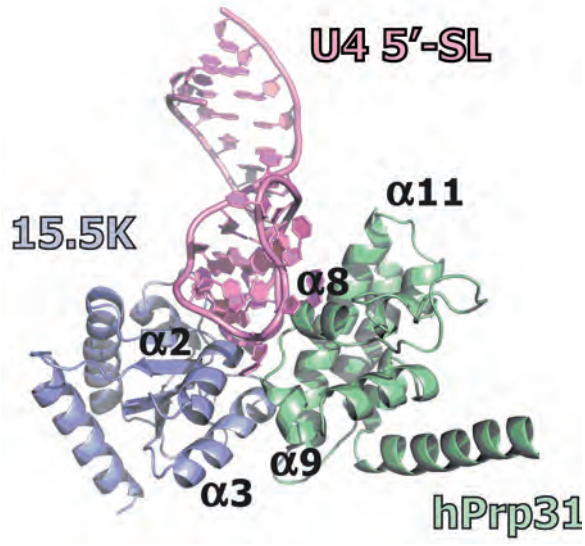
The laboratory focuses on studying 1) the interaction of small drugs with cellular receptors; and 2) structure-activity and dynamics-activity relationship of RNP complexes and catalytic RNAs involved in RNA processing.

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 focus on the development of methods that will allow a ligand-based reconstruction of the receptor binding pocket. We apply our methods to 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 characterize 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 at EMBL Grenoble (page 95) 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) recognizes 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

Rodriguez-Castaneda, F. *et al.* (2010). Modular architecture of Munc13/calmodulin complexes: dual regulation by Ca^{2+} and possible function in short-term synaptic plasticity. *EMBO J.*, 29, 680-91

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.*, 388, 283-98

Orts, J., Tuma, J., Reese, M. *et al*. (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

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.

Selected references

Yus E., Maier T., Michalodimitrakis K., van Noort V. *et al.* (2009). Impact of genome reduction on bacterial metabolism and its regulation. *Science*, 326, 1263-1268

Kühner S., van Noort V., Betts M.J., Leo-Macias A., Batisse C. *et al.* (2009). Proteome organization in a genome-reduced bacterium. *Science*, 326, 1235-1240

Güell M., van Noort V., Yus E., Chen W.-H., Leigh-Bell J. *et al.* (2009). Transcriptome complexity in a genome-reduced bacterium. *Science*, 326, 1268-1271

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

Biomolecular networks

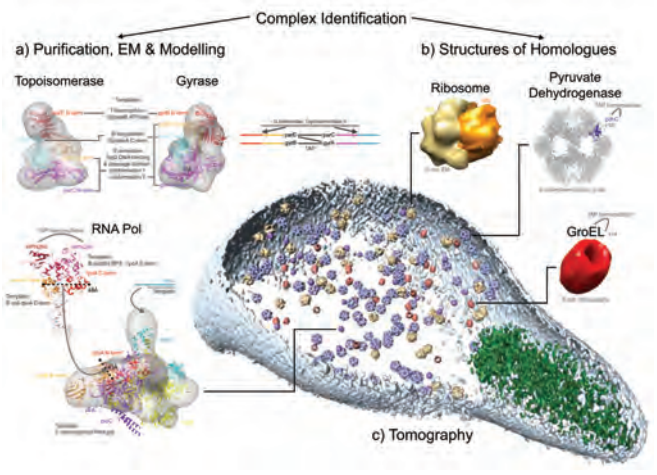
Previous and current research

How is biological matter organised? Can the protein and chemical worlds be matched to understand the cell's inner works? We can now access an unprecedented level of knowledge on the basic components of living systems; an ever-growing number of molecular players and functions are being characterised and localised. Despite this spectacular progress we still don't understand how cellular components work collectively and achieve biological function. Our group's research focuses on three main areas in the detailed and systematic charting of cellular networks and circuitry at molecular levels, both spatially and temporally.

The charting of biological networks: Biological function at cellular levels is achieved by groups of interacting proteins or protein complexes that represent the basic functional and structural units of proteome organisation. The systematic charting of their dynamics has been one of our main focuses, for which we use biochemical and quantitative mass spectrometry (MS) approaches in the eukaryote *S. cerevisiae*, the human pathogen *M. pneumoniae* and, in the future, thermophiles or other extremophiles. The datasets produced allow an unbiased overview of important biological principles. Protein complexes often form larger assemblies, suggesting that sequential steps in biological processes have been captured, and they also often share components, implying protein multifunctionality or pleiotropy. Collaborations with structural groups at EMBL and incorporation of structural models, single-particle EM and cellular electron tomograms provides supporting structural details for this proteome organisation. The group is also part of a network of EMBL groups tackling a range of biological networks in *M. pneumoniae*, for which we generated large-scale quantitative datasets on Mycoplasma transcription, metabolome and proteome organisation.

Development of new methods for charting new types of biological networks: While current protein–protein or protein–DNA (regulatory) networks give spectacular results, huge uncharted areas still need to be tackled. For example, many metabolites have signalling functions and many proteins are allosterically modulated by metabolites. These bindings are sometimes mediated by a variety of specialised domains; to date, though, large-scale, unbiased analyses are still largely missing. The group developed interests in new methods for the systematic charting of interactions between cellular proteomes, small molecules or metabolites. For example, in *S. cerevisiae* we developed a generic biochemical assay based on miniaturised lipid arrays for the systematic study of protein-lipid interactions. New avenues such as affinity chromatography methods using immobilised metabolites as affinity probes are being explored. We are also interested in multiplexing the assays through miniaturisation using integrated microfluidic devices.

Bridging biological networks to phenotypes: Because biological function arises from extensively interacting biomolecules, it is in the context of biological networks that information encoded in genomes must be decrypted. We use networks as a molecular frame for the interpretation of phenotypic data recorded after systematic cell perturbations; these include small molecule inhibitors, gene knock-outs and mutations. We also use network analyses to design models, predictions and perturbations that can be challenged experimentally.



Future projects and goals

- Further development of chemical biology methods based on affinity purification to monitor protein-metabolites interaction.
- Global screen aiming at the systematic charting of the interactions taking place between the proteome and the metabolome in the model organisms *S. cerevisiae* and *M. pneumoniae*.
- Develop new and existing collaborations with computational and structural biology groups at EMBL and elsewhere to tackle the structural and functional aspects of biomolecular recognition.

Toby Gibson

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

Biological sequence analysis

Previous and current research

It is now clear that the regulatory decisions during eukaryotic cell signalling are made within large dynamic protein complexes (see Gibson, 2009). Cell regulation is networked, redundant and – above all – cooperative. The deeply misleading 'kinase cascade' metaphor needs to be retired and the sooner, the better. The proteins that take part in regulatory systems make remarkable numbers of interactions, with the corollary that they also have highly modular architectures. Therefore our main recent focus has been to develop and deploy computational tools for protein architecture analysis, working mostly at the sequence level.

Thus we coordinate development of ELM, the Eukaryotic Linear Motif resource for functional sites in modular protein sequences. Linear motifs (LMs) are short functional sites used for the dynamic assembly and regulation of large cellular protein complexes and their characterisation is essential if we are to understand cell signalling. So-called 'hub' proteins that make many contacts in interaction networks are being found to have abundant LMs in large segments of IUP (intrinsically unstructured protein segments). The freely available ELM resource data are now used by many bioinformatics groups to improve prediction of LM interactions, examples being the NetworKIN kinase-substrate predictor and the DILIMOT and SLIMFinder novel motif predictors.

Recent ELM developments include the addition of structure and conservation filtering. We are now actively hunting for new LM candidates; we recently proposed new candidate KEN boxes, a sequence motif that targets cell cycle proteins for destruction in anaphase (see figure), as well as KEPE, a motif of unknown function that is superposed on many sumoylation sites. We have worked closely with groups undertaking validation experiments: Michael Sattler (Munich) in the characterisation of ULM-UHM interactions used in alternative splicing, and Annalisa Pastore (Mill Hill) on a fascinating 3-way molecular switch in Ataxin-1.

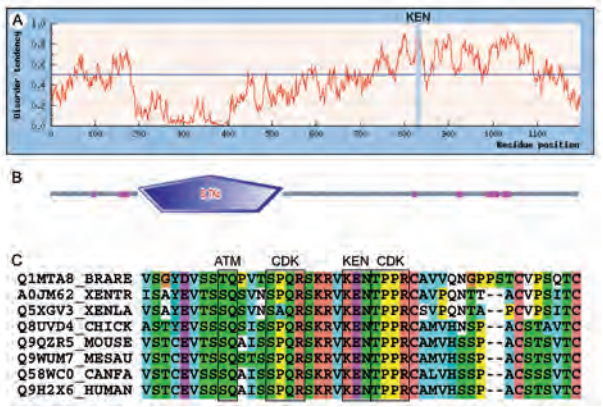
We also undertake more general computational analyses of biological macromolecules. Where possible, we contribute to multidisciplinary projects involving structural and experimental groups at EMBL and elsewhere. We collaborate with Des Higgins (Dublin) and Julie Thompson (Strasbourg) to maintain and develop the ClustalW and Clustal X programs that are widely used for multiple sequence alignment. We maintain several public web servers at EMBL, including ELM, the protein linear motif resource; Phospho.ELM, a collection of some 20,000 reported phosphorylation sites; and GlobPlot, a tool for exploring protein disorder.

Future projects and goals

We will continue to hunt for regulatory motifs and may survey individual gene families in depth and will undertake proteome surveys when we have specific questions to answer. Protein interaction networks are anticipated to become increasingly important to our work. Molecular evolution is also one of the group's interests, especially when it has practical applications.

With our collaborators, we will look to build up the protein architecture tools, especially the unique ELM resource, taking them to a new level of power and applicability. We will apply the tools in the investigation of modular protein function and may deploy them in proteome 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.

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

Gould, C.M., Diella, F., Via, A., Puntervoll, P., Gemund, C., Chabanis-Davidson, S., Michael, S., Sayadi, A., Bryne, J.C., Chica, C., Seiler, M., Davey, N.E., Haslam, N. *et al.* (2009). ELM: the status of the 2010 eukaryotic linear motif resource. *Nucleic Acids Res.*, 38, D167-80

Gibson, T.J. (2009). Cell regulation: determined to signal discrete cooperation. *Trends Biochem Sci.*, 34, 471-482

Diella, F., Chabanis, S., Luck, K., Chica, C., Ramu, C., Nerlov, C. & Gibson, T.J. (2009). KEPE—a motif frequently superimposed on sumoylation sites in metazoan chromatin proteins and transcription factors. *Bioinformatics*, 25, 1-5

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

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.

Edward Lemke

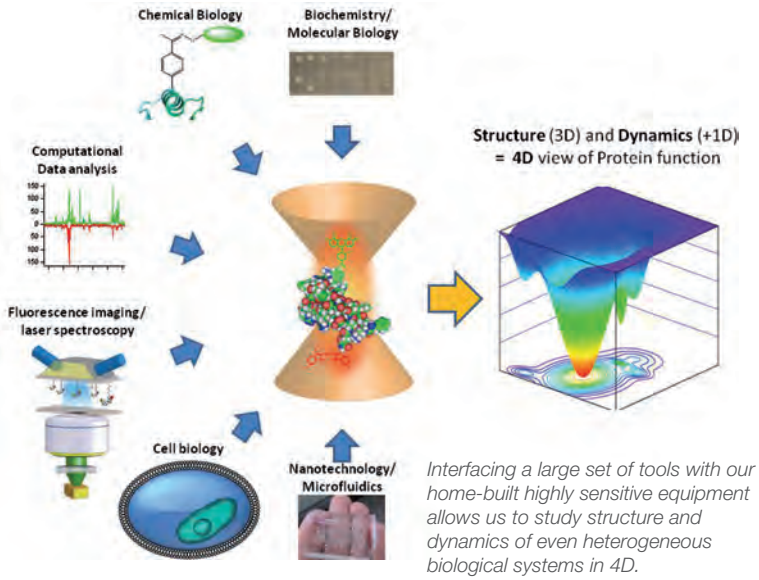
Structural light microscopy – single molecule spectroscopy

Previous and current research

We combine biochemistry, molecular biology, cell biology and modern chemical 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 would 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 signaling 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, in any complex biological system a mosaic of molecular states and reaction pathways exist simultaneously, further complicating the situation to measure these systems. For example, some proteins might behave differently, giving rise to new and unexpected phenotypes; for example 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 such a system, ignoring coexisting populations and rare events. This can easily lead to false or insufficient models, which may further impede our understanding of the biological processes and disease mechanisms.

In contrast, single molecule techniques, which directly probe the distribution of molecular events, can reveal important mechanisms that otherwise remain obscure. In particular, single molecule fluorescence studies allow probing of molecular structures and dynamics at near atomic scale with exceptional time resolution. While such experiments are even possible in the natural environment of the entire cell, single molecule fluorescence studies require labelling with fluorescent dyes which still hampers the broad application of this technique. Our lab uses a large spectrum of chemical biology and state-of-the-art protein engineering tools to overcome this, genetically encoding unnatural amino acids being one of our primary strategies. We also continue to develop new methods and recruit techniques from other disciplines, such as microfluidics.



Interfacing a large set of tools with our home-built highly sensitive equipment allows us to study structure and dynamics of even heterogeneous biological systems in 4D.

Future projects and goals

Recent studies have shown that even the building blocks of some of the most complex and precise machines with a critical role to survival of the cell, such as DNA packing, epigenetics and nuclear 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 combine protein structure and dynamics into a 4D view of biological function within its naturally complex environment.

Selected references

Lemke, E.A., Gambin, Y., Vandelinder, V., Brustad, E.M., Liu, H.W., Schultz, P.G., Groisman, A. & Deniz, A.A. (2009). Microfluidic device for single-molecule experiments with enhanced photostability. *J. Am. Chem. Soc.*, 131, 13610-13612

Ferreon, A.C., Gambin, Y., Lemke, E.A. & Deniz, A.A. (2009). Interplay of alpha-synuclein binding and conformational switching probed by single-molecule fluorescence. *Proc. Natl. Acad. Sci. USA*, 106, 5645-5650

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, 769-772

Carsten Sachse

PhD 2007, University of Jena/FLI–Leibniz-Institute for Age Research and Brandeis University, Waltham, Massachusetts.

Postdoctoral research at Max Planck Research Unit for Enzymology of Protein Folding, Halle, and at MRC Laboratory of Molecular Biology, Cambridge.

Group leader at EMBL since 2010.

Single-particle electron cryo-microscopy of the machinery involved in abnormal protein aggregation

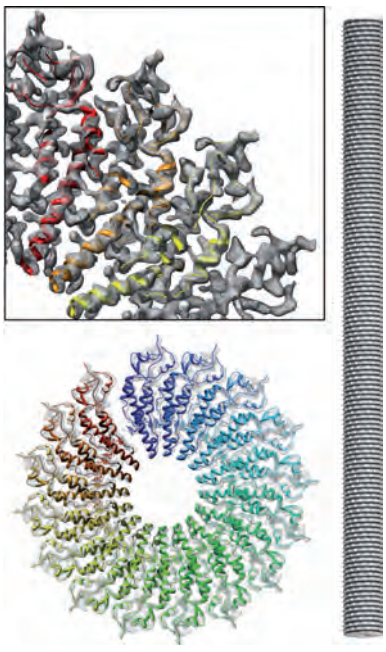
Previous and current research goals

The molecular hallmark of neurodegenerative diseases, such as Alzheimer's and Parkinson's, is the formation of large protein aggregates called amyloid fibrils. The net buildup of these fibrillar aggregates is a result of an imbalance in the cellular production and clearance of misfolded polypeptides. In our group, we are investigating the molecular structures involved in these aberrant processes using single-particle electron cryo-microscopy (cryo-EM) as they are fundamental to our understanding of ageing and neuronal dysfunction.

Single-particle cryo-EM is becoming a powerful tool of three-dimensional structure determination, because large macromolecular structures and multi-protein complexes can be studied in their near-native environment without the need for crystals. Small amounts of material are sufficient to obtain snapshots of 'single particles' in the electron cryo-microscope and subsequent computer-aided image processing enables 3D image reconstruction. Despite the recent success of the technique that enabled structure determination at near-atomic resolution, there is still a great need for hardware-based improvements and software enhancements in the scientific community. We are therefore also interested in developing techniques ranging from sample preparation to data processing with the goal of ultimately increasing the obtainable resolution of single-particle cryo-EM and realising its promising potential to make it a routine tool for structural biology of large macromolecules.

Future projects and goals

Autophagy (from the Greek, meaning 'to eat oneself') is the cell's housekeeping mechanism to engulf and degrade large protein aggregates, damaged organelles and even microbes in double-membrane vesicles called autophagosomes. Multiprotein complexes are essential mediators in the events leading to autophagy. On the structural level, little is known about their 3D architecture and thus fundamental questions on the nature of these complexes need to be addressed. What are the shapes of these multiprotein assemblies? What do they look like in the presence of lipids? How do they give rise to the cellular structure of the autophagosome? Therefore, we want to use cryo-EM and 3D image processing techniques to obtain insights into the molecular mechanisms of autophagy. We are aiming at understanding the molecular basis of aggregation buildup and delineating the structures of the pathways that mediate clearance of these otherwise harmful protein depositions associated with cellular ageing and neurodegeneration.



Right: Three-dimensional image reconstruction of an Alzheimer's Aβ(1-40) fibril superimposed on an electron cryomicrograph.

Far right: High-resolution helical reconstruction of tobacco mosaic virus at near-atomic resolution using single-particle cryo-EM.

Selected references

Sachse, C., Chen, J. Z., Coureux, P. D., Stroupe, M. E., Fändrich, M. & Grigorieff, N. (2007). High-resolution electron microscopy of helical specimens: a fresh look at tobacco mosaic virus. *J. Mol. Biol.*, 371, 812-35

Sachse, C., Fändrich, M. & Grigorieff, N. (2008). Paired beta-sheet structure of an Abeta(1-40) amyloid fibril revealed by electron microscopy. *Proc. Natl. Acad. Sci. USA*, 105, 7462-7466

Low H. H., Sachse C., Amos L. A., Löwe J., (2009). Structure of a bacterial dynamin-like protein lipid tube provides a mechanism for assembly and membrane curving. *Cell,* 139, 1342-1352

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.

Data integration and knowledge management

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, and new challenges from the standpoint of storage, indexing, retrieval and system scalability over disparate types of data is central to large-scale efforts in understanding biological systems.

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.

Selected references

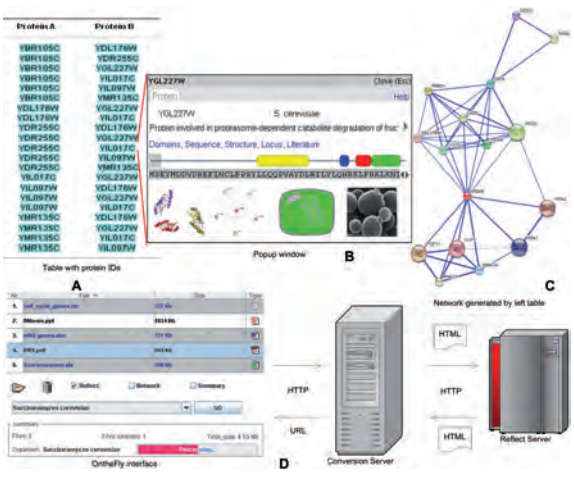
Neumann, B., Walter, T., Heriché, J-K,, Bulkescher, J., Erfle, H., Conrad, C., Rogers, P., Poser, I. *et al.* (2010). Phenotypic profiling of the human genome by time-lapse microscopy reveals genes with functions in cell division, survival or migration. *Nature*, in press

Pafilis, E., O'Donoghue, S.I., Jensen, L.J., Horn, H., Kuhn, M., Brown, N.P. & Schneider, R. (2009). Reflect: augmented browsing for the life scientist. *Nat. Biotechnol.*, 27, 508-510

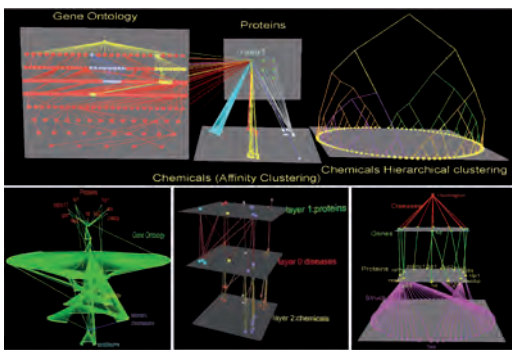
Pavlopoulos, G.A., Moschopoulos, C.N., Hooper, S.D., Schneider, R. & Kossida, S. (2009). jClust: A clustering and visualization toolbox. *Bioinformatics*, 25, 1994-6

Bromberg, Y., Yachdav, G., Ofran, Y., Schneider, R. & Rost, B. (2009). New in protein structure and function annotation: hotspots, single nucleotide polymorphisms and the 'Deep Web'. *Curr. Opin. Drug Discov. Devel.*, 12, 408-419

Below: 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.



Right: 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.



Directors' Research

Directors' Research is unlike other EMBL units in that it covers three thematically distinct research groups, headed by the Director General and Associate Director of EMBL and the Director of EMBO. 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 studies diverse processes that are under the control of the Ran GTPase. During mitosis, Ran is needed for both mitotic spindle assembly and nuclear envelope (NE) formation. 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 focused on identifying the factors that are involved in NE assembly – a multi-stage process – and their modes of action. 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 a 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 post-transcriptional control work 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 MMPU research on the regulation of the iron hormone hepcidin and its involvement in iron overload and deficiency diseases.

The Leptin group uses two cell types to study the processes determining cell shapes. The branched terminal cells of the *Drosophila* tracheal system contain an intracellular membranous tube that transports oxygen to cells. The mechanisms involved in establishing the cells' architecture, especially the oxygen-carrying lumen, are poorly understood, and the group studies the role of vesicle trafficking and fusion, polarity protein complexes and the cytoskeleton. As subcellular localisation of mRNA and the local control of translation are recognised as significant for polarised cellular functions, the group have also developed an *in vivo* screening method to identify genes with asymmetrically localised RNA in tracheal cells. This has already led to the discovery of new genes which will be used to build a sequence database of mRNAs with tissue-specific polar distributions. Collaborating laboratories will use the insertion stocks to screen other cell types, resulting in a valuable bioinformatic resource. Secondly, the epithelial cells in the embryo mesoderm are used to study the dynamics of epithelial junctions during cell shape changes and epidermal-to-mesenchymal transitions, using laser microsurgery, live imaging, genetic and cell biological methods.

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.

Selected references

Santarella-Mellwig, R., Franke, J., Jaedicke, A., Gorjanacz, M., Bauer, U., Budd, A., Mattaj, I.W. & Devos, D.P. (2010). The compartmentalized bacteria of the planctomycetes-verrucomicrobia-chlamydiae superphylum have membrane coat-like proteins. *PLoS Biol.*, 8, e1000281

Gorjanacz, M. & Mattaj, I.W. (2009). Lipin is required for efficient breakdown of the nuclear envelope in *C. elegans*. *J Cell Sci.*, 122, 1963-1969

Casanova, C.M., Rybina, S., Yokoyama, H., Karsenti, E. & Mattaj, I.W. (2008). Hepatoma up-regulated protein is required for chromatin-induced microtubule assembly independently of TPX2. *Mol. Biol. Cell*, 19, 4900-4908

Yokoyama, H., Gruss, O.J., Rybina, S. *et al.* (2008). Cdk11 is a RanGTP-dependent microtubule stabilization factor that regulates spindle assembly rate. *J. Cell Biol.*, 180, 867-875

The RanGTPase as a spatial regulator

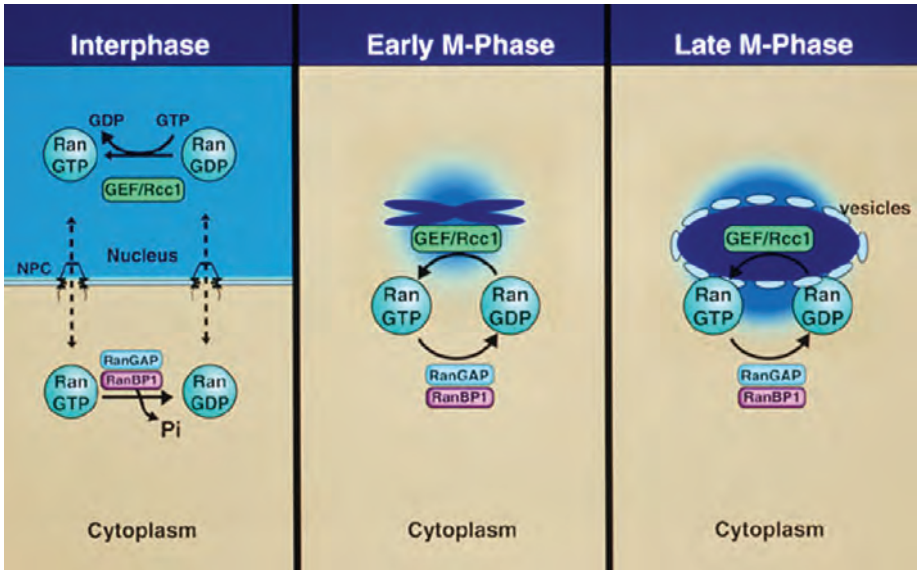
Previous and current research

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

Future projects and goals

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

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



The 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.

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.

Cytoplasmic gene regulation and molecular medicine

Previous and current research

Important steps in the control of gene expression are executed in the cytoplasm: the regulation of mRNA translation and stability. We are elucidating these regulatory mechanisms, including the function of miRNAs, which has become a very active focus of our work (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).

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.

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

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

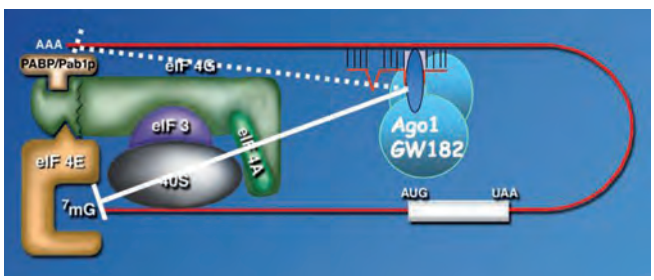
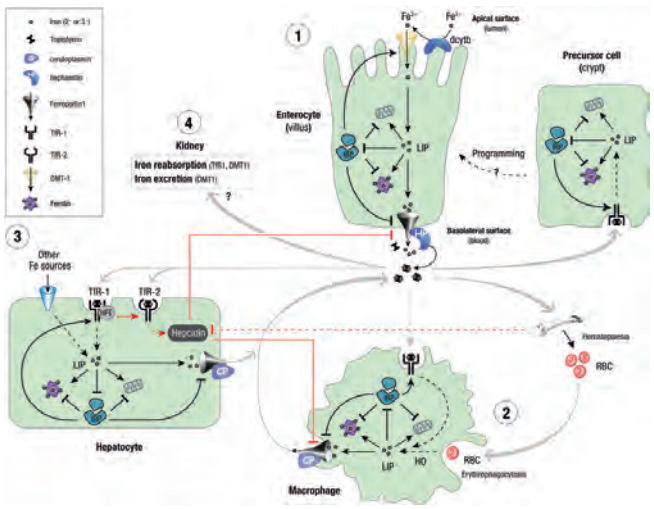


Figure 2: Systems biology of mammalian iron metabolism.



Selected references

Gehring, N.H., Lamprinaki, S., Kulozik, A.E. & Hentze, M.W. (2009). Disassembly of exon junction complexes by PYM. *Cell*, 137, 536-548

Vujic Spasic, M., Kiss, J., Herrmann, T., Galy, B., Martinache, S., Stolte, J., Grone, H.J., Stremmel, W., Hentze, M.W. & Muckenthaler, M.U. (2008). Hfe acts in hepatocytes to prevent hemochromatosis. *Cell Metab.*, 7, 173-178

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

Zdanowicz, A., Thermann, R., Kowalska, J., Jemielity, J., Duncan, K., Preiss, T., Darzynkiewicz, E. & Hentze, M.W. (2009). *Drosophila* miR2 primarily targets the m7GpppN cap structure for translational repression. *Mol. Cell*, 35, 881-888

Maria Leptin

PhD 1983, Basel Institute for Immunology.
Postdoctoral research then Staff Scientist at the MRC Laboratory for Molecular Biology, Cambridge, UK.
Group leader at the MPI for Developmental Biology, Tübingen.
Full Professor, University of Cologne, Institute of Genetics.
Director of EMBO and group leader at EMBL since 2010.

Generation of complex cell shapes: membrane vesicles, polarity proteins and localised mRNAs

Previous and current research

Many cells have highly polarised morphologies that serve specialised cell functions. We use the terminal cells of the *Drosophila* respiratory (tracheal) system and the epithelium of the gastrulating embryo to study the processes that determine cell shapes.

Terminal cells, like neurons, are extensively branched. The mechanisms involved in establishing their cellular architecture, especially the oxygen-carrying lumen, are poorly understood, as is also the case for related events in vertebrate tissues. Tracheal cell morphology depends on environmental signals that trigger events at sites distant from the cell body. We therefore assume that it depends on protein synthesis from locally stored RNA. We have developed an *in vivo* genetic screening method to search for such mRNAs. A pilot screen has identified genes with localised mRNAs that are required for the elaboration of the tracheal branches, and that had not been found by other methods. The analysis of these genes is beginning to reveal the mechanisms by which branches grow in response to local cues and generate the intracellular lumen to transport oxygen. These involve vesicle trafficking and fusion, definition of membrane domains and their polarity, and interactions of membranes with the cytoskeleton.

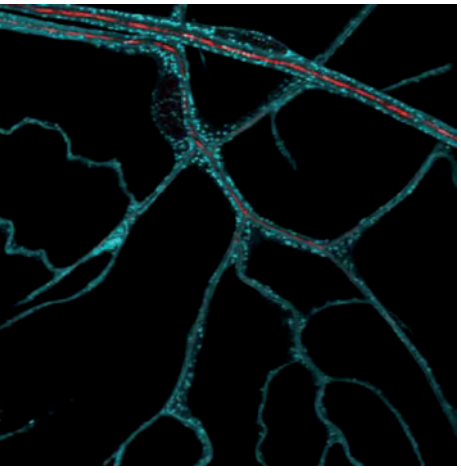
In the embryo, we have found that cell shape changes depend on localised G-protein dependent recruitment of cytoskeletal regulators, and require a dramatic re-distribution of adherens junctions by as yet unknown mechanisms. This redistribution occurs under the control of the transcription factor Snail, which controls epithelial-mesenchymal transitions in normal development and cancer, but the target genes through which Snail acts are largely unknown.

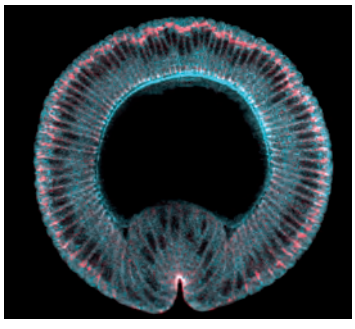
Future projects and goals

Our work on tracheal cells has two aims: the understanding of complex cell morphology and the search for signals involved in mRNA localisation. We will conduct a large-scale screen to identify all genes with localised mRNAs in tracheal cells, with a view to analysing their cell biological functions as well as their localisation signals. Results from testing genes from the pilot screen in other cell types have indicated that there must be cell type specific mRNA recognition and processing systems. To build a sequence database of mRNAs with tissue-specific polar distributions, the data obtained through screening tracheal cells in our own group will be enriched by input from collaborating laboratories who will use the insertion stocks we generate to screen other cell types. The combined datasets will be used for the bioinformatic identification of localisation signals.

To understand how the transcription factor Snail induces epithelial mesenchymal transitions, we will study sets of genes that were found in ChIP-on-chip experiments and by expression profiling to be direct targets of Snail in the mesoderm. In parallel, we will use laser micromanipulations and genetic and cell biological approaches to determine the mechanisms by which the adherens junctions are disassembled and reassembled.

Below: a tracheal terminal cell expressing a fluorescent marker for the endoplasmic reticultum (blue). The lumen of the cell, which carries the oxygen for the surrounding (invisible) tissue, is seen in red (PhD thesis, Jayan Nair).





Cross section through a Drosophila *embryo, in which the cells on the ventral side (bottom) have begun to change their shapes, creating an indentation that will eventually lead to the internalisation of these cells. The embryo is stained with antibodies against beta catenin (pink) and RhoGEF2 (blue). (Image by Verena Kölsch).*

Selected references

Csiszar, A., Vogelsang, E., Beug, H. & Leptin, M. (2010). A novel conserved phosphotyrosine motif in the *Drosophila* FGF-signaling adaptor Dof with a redundant role in signal transmission. *Mol Cell Biol.*, in press

Mathew, S.J., Haubert, D., Kronke, M. & Leptin, M. (2009). Looking beyond death: a morphogenetic role for the TNF signalling pathway. *J. Cell Sci.*, 122, 12, 1939-46

Baer, M.M., Bilstein, A. & Leptin, M. (2007). A clonal genetic screen for mutants causing defects in larval tracheal morphogenesis in *Drosophila*. Genetics, 176, 4, 2279-91

Kolsch, V., Seher, T., Fernandez-Ballester, G.J., Serrano, L. & Leptin, M. (2007). Control of *Drosophila* gastrulation by apical localization of adherens junctions and RhoGEF2. *Science*, 315, 5810, 384-6

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
Head of Core Facilities and Services

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

Rainer Pepperkok

Advanced Light Microscopy Facility

The Advanced Light Microscopy Facility (ALMF) offers a collection of state-of-the-art light microscopy equipment and image processing tools. The facility 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 ALMF also organises regular international courses to teach advanced light microscopy methods.

Major projects and accomplishments

- The ALMF presently manages 19 top-of-the-line microscope systems plus nine High-Content Screening microscopes from leading industrial companies, as well as five image analysis workstations.
- More than 30 visitors per year come to carry out their own experiments in the ALMF or to evaluate microscopy equipment.
- The ALMF was the seed for the European Light Microscopy Initiative (ELMI) that establishes links with light microscopy facilities throughout Europe.
- Usage of the facility has exceeded 30,000 hours per year.

Services provided

- Assisting all EMBL groups and visitors through the microscopic imaging process: project planning, sample preparation, staining, microscope selection and use, image processing, presentation, data transfer and storage.
- Developing accessory software and microscopy equipment, co-developments with industrial partners, pre-evaluation of commercial equipment.
- Facilitating automated microscopy and High Content Screening (HCS) projects.

Technology partners

The ALMF presently has collaborations with the following companies:

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

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

Selected references

Erfle, H., Neumann, B., Rogers, P., Bulkescher, J., Ellenberg, J. & Pepperkok, R. (2008). Work Flow for Multiplexing siRNA Assays by Solid-Phase Reverse Transfection in Multiwell Plates. *J. Biomol. Screen.*, 13, 575-580

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

Pepperkok, R. & Ellenberg, J. (2006). High-throughput fluorescence microscopy for systems biology. *Nat. Rev. Mol. Cell Biol.*, 7, 690-696

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, 1283-1289

Joe Lewis

PhD 1991, Institute of Molecular Pathology, Vienna.
Postdoctoral research at EMBL.
Group leader, 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.

Chemical Biology Core Facility

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

The facility is a collaboration between EMBL, the German Cancer Research Center (DKFZ), 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. We have a very strong pipeline of projects from all three institutes covering biochemical and cell based targets. At the end of 2009 we established a computational chemistry as part of the facility offering. Elara Pharmaceuticals GmbH and Savira Pharmaceuticals GmbH have been founded to further develop and commercialise active compounds indentified in the facility, targeting specific cancer cell signalling pathways and the influenza virus respectively.

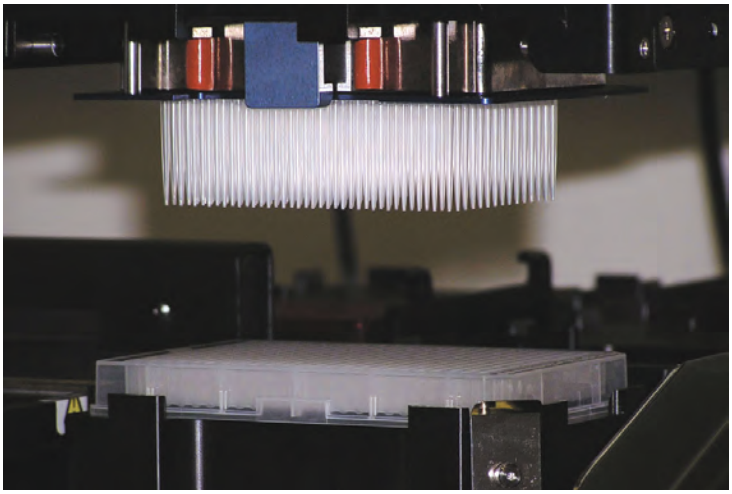
Services provided

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.

Parallel pipetting of samples in 384-well format.



Technology partners

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

Selected references

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T., Sehr, P., Lewis, J.D., 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

Polycarpou-Schwarz, M., Muller, K., Denger, S., Riddell, A. *et al.* (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

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

Equipment available: **http://www.embl.de/services/core_facilities/em/equipment/index.html**

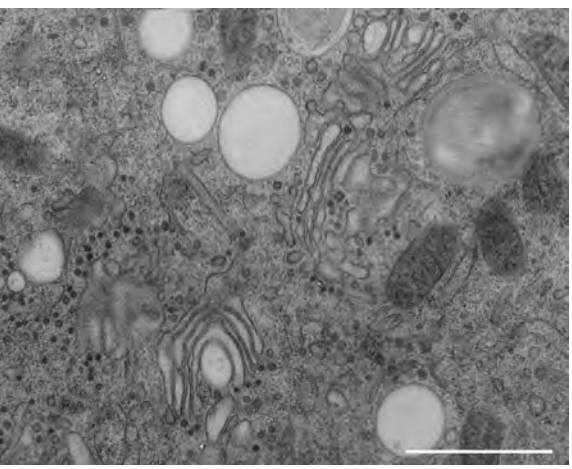
Online booking: **https://webapps.structures.embl.de/EM-booking**

The EMCF gives EMBL scientists access to and training in advanced electron microscopes, 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 develops new EM methods in cellular and developmental biology.

Major projects and accomplishments

Our new electron tomography equipment, which has been operational since the start of 2008, includes a new microscope and computing set-up with programs for 3D reconstruction and cellular modelling. Mostly performing cellular tomography of plastic-embedded samples, the new microscope is a FEI F30 (300 kV microscope with a Field Emission Gun and Eagle FEI 4K camera) which is also equipped as a cryomicroscope. It is managed by specialised EM engineers with expertise in tomography data acquisition and processing, who also train researchers in handling the electron tomography microscope and its applications for cellular structure modelling. CorrelativeMicroscopy technology (a collaboration between EMCF and ALMF, page 54) has been established with conventionally fixed cells grown on coverslips (Colombelli *et al.,* 2008). We are currently collecting data using a similar method adapted to cells grown on sapphire coverslips which are cryofixed by high-pressure freezing after LM visualisation. Other projects include investigations on annexins in plasma membrane injured cells (with the Schultz group, page 16); 3D reconstruction of the dorsal closure in the *Drosophila* embryo (Brunner and Frangakis groups, formerly EMBL) and the microtubule based polarity of the *Drosophila* oocyte (Ephrussi group, page 20). External collaborations concerned in particular the study of kinetochore-derived microtubules in budding yeast (T. Tanaka, Dundee University), SPB duplication in meiotic fission yeast (K. Tanaka, University of Leicester), bacterial proteins expression in fission yeast (M. Balasubramanian, Temasek Life Science Laboratory, Singapore), *in vivo* dengue virus replication (R. Bartenschlager, Molecular Virology, University of Heidelberg) and SPB-microtubules interactions in budding yeast (Schiebel group, DKFZ–ZMBH, Heidelberg).

MDCK cells grown on a sapphire disk were cryofixed by high pressure freezing, freeze-substituted and plastic-embedded. The section reveals Golgi apparatus elements lipid droplets and budding viral particles. (Picture by Uta Haselmann, collaboration with the Hufnagel group (page 11)).



Services provided

- Up-to-date know-how of EM methods for cell biology, immunocytochemistry, cryosectioning and cryofixation applied to various cell types or organisms.
- Maintaining the microscopes and equipment for sample preparation, microtomy and cryogenic methods.
- Supplying a range of reagents specific for EM methods and protocols.
- Electron tomography, image acquisition and data processing for plastic-embedded samples.
- Assisting users in choosing the right methods and protocols.
- Organising courses and lectures on EM methods in cell biology.

Technology partners

- FEI Company (advanced electron microscopes, including the new tomography microscope).
- Leica Microsystems (supplied our portable EMPACT2 HPFreezer, as well as ultramicrotomes units for sample plastic- or cryo-sectioning).

Selected references

Colombelli, J., Tangemo, C., Haselman, U., Antony, C., Stelzer, E.H.K., Pepperkok, R. & Reynaud, E.G. (2008). A correlative light and electron microscopy method based on laser micropatterning and etching. *Methods Mol. Biol.*, 457, 203-213

Welsch, S., Miller, S., Romero-Brey, I., Merz, A., Bleck, C.K., Walther, P., Fuller, S.D., Antony, C., Krijnse-Locker, J. & Bartenschlager, R. (2009). Composition and 3D architecture of the dengue virus replication and assembly sites. *Cell Host Microbe,* 5, 365-75

Kitamura, E., Tanaka, K., Komoto, S., Kitamura, Y., Antony, C. & Tanaka, T.U. (2010). Microtubules generated at kinetochores with distal plus ends: their roles and limited lifetime in mitosis. *Dev. Cell*, in press

Roque, H. & Antony, C. (2010). Electron microscopy of model systems: the fission yeast *Schizossacharomyces pombe*. *Methods in Cell Biology*, in press

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.

Flow Cytometry Core Facility

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

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

Major projects and accomplishments

- The analysis of algae life cycle project relied on the intrinsic fluorophores in algae to identify life cycle stages. It required photosaturation of the photosynthetic units in the algae.
- Cell cloning by Darwinian selection required a series of single cell sorts of a target population into 96-well plate in order to select a stable integrated gene into a cell line.
- There is a project investigating a bi-stable state of a reworked bacterial signalling cascades that requires precise and accurate instrument measurements of the bacteria in order for them to be identified.
- The apoptosis project 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.

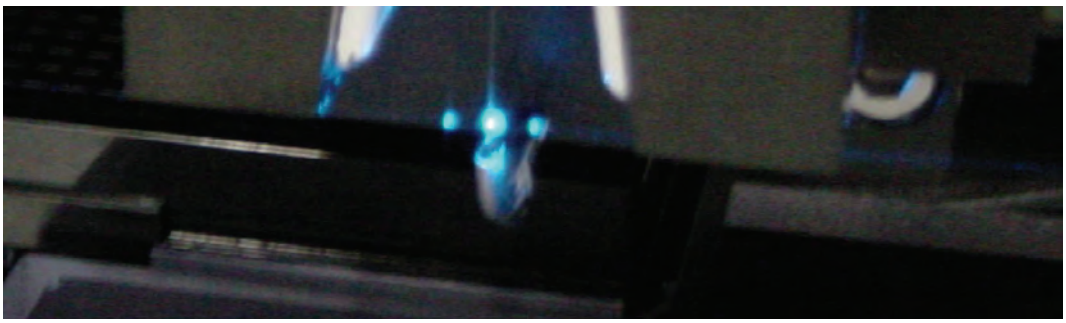
Selected references

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

Deflection illumination for calibrating droplet break-off point.



PhD 1994, Czech Academy of Sciences, Prague.
Postdoctoral research at EMBL.
Facility head since 2001.

Vladimír Beneš

Genomics Core Facility www.genecore.embl.de

GeneCore is the in-house genomics service centre at EMBL, equipped with state-of-the-art technologies required for functional genomics analyses and embracing the latest high-throughput methodologies. Training is a vital part of our activities, so the highly-qualified staff also tutor individual researchers and organise practical courses on complementary subjects.

Major projects and accomplishments

The acquisition of new-generation sequencing technology has been a vital step to keep EMBL at the forefront of European research. Currently, our massively parallel sequencing (MPS) suite consists of three Illumina GAIIx and two cluster generation stations. Preparation of MPS libraries is facilitated by a robust infrastructure of other instruments (e.g. Covaris, Bioanalyzer, Qubit).

We provide the following analyses in a single-read or a paired-end sequencing mode, including barcoding:

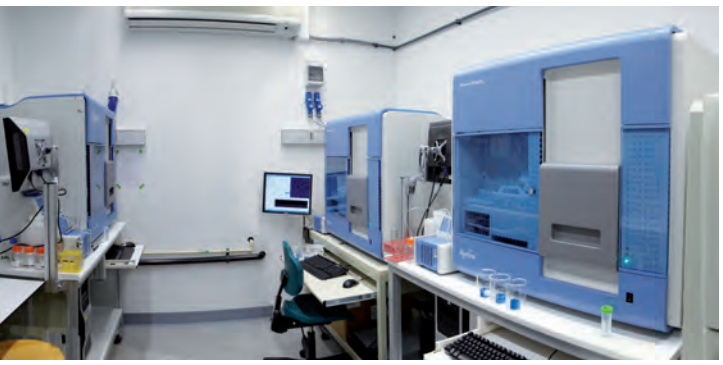
- Genome-wide location analysis of nucleic acids-protein interactions – ChIP-Seq, CLIP-Seq;
- Transcriptome sequencing – RNA-Seq;
- Discovery of small non-coding RNAs – ncRNA-Seq;
- Genome-wide DNA methylation analysis – Methyl-Seq;
- De novo sequencing & resequencing of genomic DNA;
- Targeted enrichment (sequence capture) on arrays or in solution coupled with MPS.

For analysis of large volumes of data generated with MPS technology, GeneCore has obtained a Genome Analyzer and Mapping Station from Genomatix. We also interact with EMBL's bioinformatics community towards implementation of tools developed in-house.

Services provided

- Microarrays (home-made, commercial), liquid handling robotics;
- MPS sequencing, Bioanalyzer; miRNA qPCR profiling;
- access to instruments and complete support (qPCR, NanoDrop, PCR cyclers, microarray spotters and scanners, high-capacity vacuum concentrator).

GeneCore offers processing of samples for a range of applications (gene, miRNA and other ncRNA expression profiling, comparative genome hybridisation, occupancy profiling) suitable for hybridisation to various microarrays including Affymetrix and Agilent. Acquisition of a highly sensitive Agilent scanner further enhanced the facility's capacity to process microarrays in a microscope slide format.



Massively parallel sequencing suite.

microRNAs (miRNAs) are now recognised as an important player in regulation of gene expression, but detailed analysis of miRNAs is technically demanding. Expression profiling of mature miRNAs is facilitated by locked nucleic acids-based oligonucleotide array miChip. This valuable platform is complemented by a panel of qPCR TaqMan miRNA assays for human and mouse and an opened miQPCR system co-invented with Mirco Castoldi (MMPU, see page 51).

qPCR instruments are primarily used for gene transcript quantification to corroborate microarray results, as well as for detailed DNA occupancy profiling after chromatin immunoprecipitations. With our assistance, EMBL researchers analysed about 200,000 qPCR assayed points in 2009. We also implemented a new application, high-resolution melting analysis, which enables determination of sample homogeneity and can also be used for the analysis of proteins.

Selected references

Ibberson, D., Benes, V., Muckenthaler, M.U. & Castoldi, M. (2009). RNA degradation compromises the reliability of microRNA expression profiling. *BMC Biotechnol.*, 9, 102

Güell M., van Noort V., Yus E., Chen W.-H., Leigh-Bell J., Michalodimitrakis K., Yamada T. *et al.* (2009). Transcriptome complexity in a genome-reduced bacterium. *Science*, 326, 1268-1271

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

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

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.

Protein Expression and Purification Core Facility

Our facility produces and purifies proteins from *E. coli*, insect cells and sera using a variety of chromatographic methods. Following purification, we perform biophysical analyses to ensure quality in terms of correct folding and stability. We also develop and evaluate new techniques and advanced protocols, as well as time-saving solutions, for protein production and purification. Moreover, we keep stocks of a large number of expression vectors and bacterial strains as well as frequently used enzymes, which reduces costs for our users considerably.

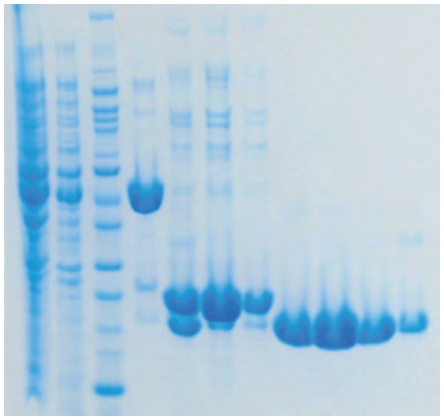
Major projects and accomplishments

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), and 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 re-designing 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.
- Preparing injection material for immunisations and purification of antibodies from serum and hybridoma supernatants.
- 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 use at EMBL.
- Developing and testing new vectors and protocols.
- Providing access to protocols and vector sequence information on the web.
- Giving scientific and technical advice to internal and external users.
- Caring for protein production and analysis equipment and the 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 technologies that could be helpful for improving the services of our group. We are frequently approached by company representatives when there are new products to be tested, which we perform according to requirements.

Selected references

Grimm, C., Matos, R., Ly-Hartig, N., Steuerwald, U., Lindner, D., Rybin, V., Muller, J. & Muller, C.W. (2009). Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. *EMBO J.*, 28, 1965-77

Hothorn, M., Neumann, H., Lenherr, E.D., Wehner, M. *et al.* (2009). Catalytic core of a membrane-associated eukaryotic polyphosphate polymerase. *Science*, 324, 513-516

Kuehnel, M.P., Rybin, V., Anand, P.K., Anes, E. & Griffiths, G. (2009). Lipids regulate P2X7-receptor-dependent actin assembly by phagosomes via ADP translocation and ATP synthesis in the phagosome lumen. *J. Cell Sci.*, 122, 499-504

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

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.

Proteomics Core Facility

We provide a full proteomic infrastructure for the identification and characterisation of proteins both in-house and for external visitors. This includes various platforms for protein and peptide separation and state-of-the-art mass spectrometry for MS and LC-MSMS experiments, including medium-throughput protein and peptide identification.

Major projects and developments

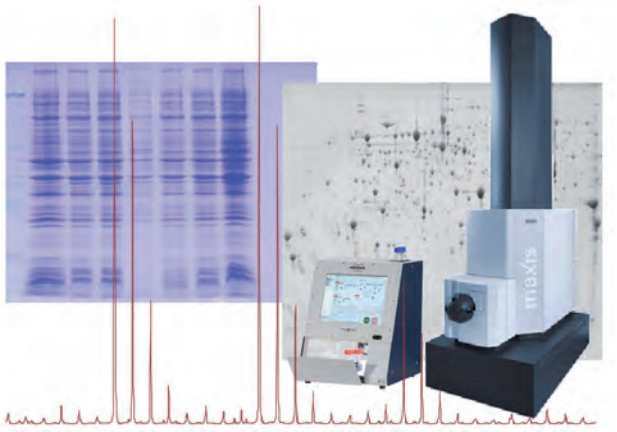
Determination of molecular weights of intact proteins; nano-flow liquid chromatography coupled to high-resolution mass spectrometry with high mass-accuracy for the identification of proteins in complex mixtures; routine identification of proteins from coomassie and silver-stained gels; developed new 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

- Iso-electric focusing of proteins and peptides, using a wide range of pH gradients. Proteins can be separated as a first step in 2D gel electrophoresis. Peptide IEF (using the OFFgel fractionators) is a relatively new addition which is typically used as a pre-fractionation step in peptide-based proteomics prior to LC-MSMS.
- SDS-PAGE gel electrophoresis on gels from 7 to 24 cm, either in 1D or 2D gel electrophoresis.
- Various staining protocols: Coomassie, silver, fluorescent staining by Sypro Ruby or Flamingo, ProQ-Diamond phosphostain.
- Densitometry of coomassie and silver stained gels; laser fluorescence imaging of fluorescently stained gels; PDQuest software for differential analysis of 2D gels.
- Differential analysis by DIGE, including fluorescent staining, quantitation and statistical analysis of 2D spot patterns. Automatic excision of spots from 2D gels.
- Protein digestion in gel or in solution using a variety of proteases (trypsin, LysC, ArgC, chymotrypsin)
- Peptide purification using ZIP tips or Stage-tips.
- Protein identification by MALDI peptide mass fingerprinting.



- ESI-MS for protein molecular weight determination under denaturing or non-denaturing conditions.
- Nano-flow reversed-phase chromatography coupled in-line with ESI mass spectrometry (ion trap and MaXis Q-tof).
- Ion trap MS and MSMS for routine identification of proteins from coomassie and silver-stained gels.
- Peptide fragmentation by CID and ETD.
- High-resolution and high mass-accuracy MS and MSMS using MaXis Q-tof for detailed characterisation of proteins, including posttranslational modifications.
- Protein identification from peptide fingerprints and LC-MSMS using Mascot. Databases can be customised to the user's needs.
- Reporting to users in appropriate detail.

Selected references

Pekova, S., Ivanek, R., Dvorak, M., Rueggeberg, S., Leicht, S., Li, X., Franz, T., Kozak, T., Vrba, J., Koza, V. *et al.* (2009). Molecular variability of FLT3/ITD mutants and their impact on the differentiation program of 32D cells: implications for the biological properties of AML blasts. *Leuk. Res.*, 33, 1409-1416

Ruggeberg, S., Horn, P., Li, X., Vajkoczy, P. & Franz, T. (2008). Detection of a gamma-carboxy-glutamate as novel post-translational modification of human transthyretin. *Protein Pept. Lett.*, 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,

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

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; large-scale analysis of regulatory systems and differentiation, RNA genomics; mathematical analyses of evolutionary models; computational modelling of neuronal signalling; text mining; and signalling circuits in health and disease. 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 at www.embl.org/jobs.

Janet Thornton
Director, EMBL-EBI

Janet Thornton

PhD 1973, King's College & Nat. 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, UCL, since 1990. Bernal Professor at Birkbeck College, 1996-2002. Director of 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

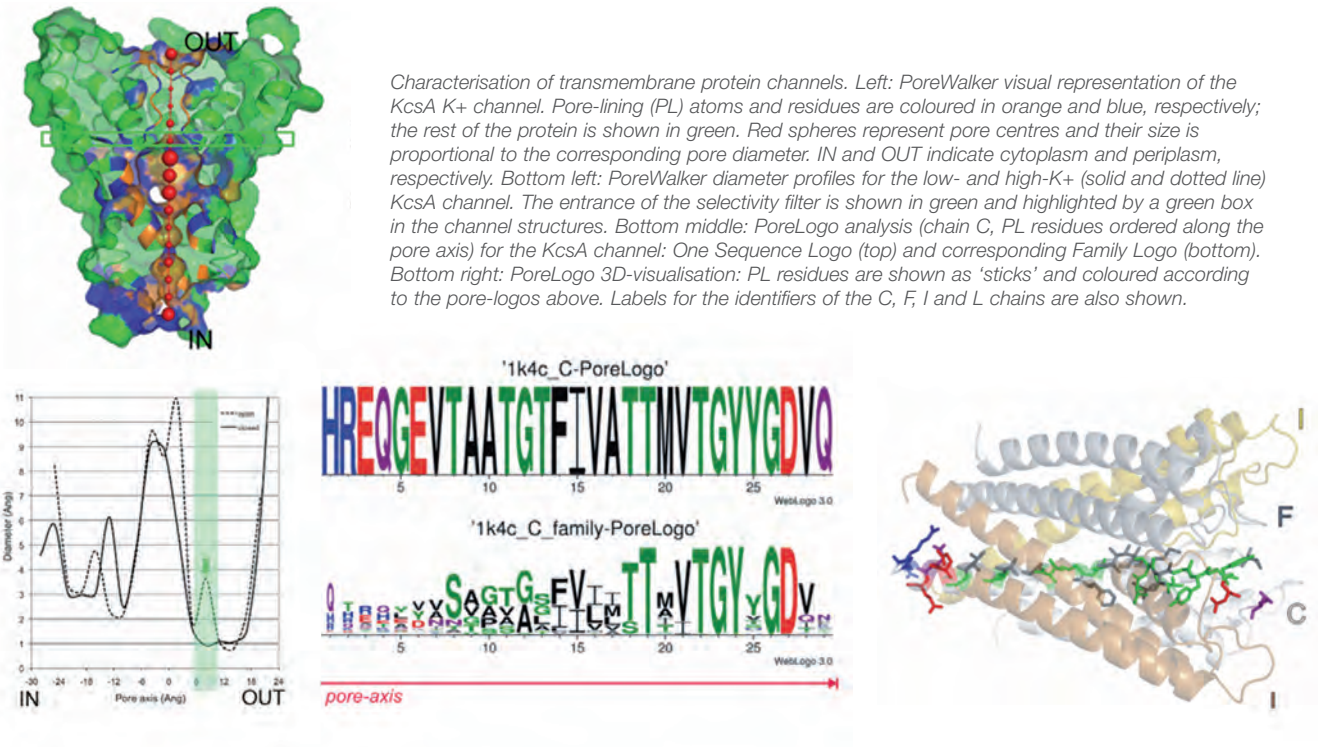
Previous and current research

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 2009 our major contributions have been in the following areas:

- enzyme structure and function;
- using structural data to predict protein function;
- functional genomics analysis of ageing.

Future projects and goals

We will continue our work on understanding more about enzymes and their mechanisms using structural and chemical information. This will include a study of how the enzymes, their families and their pathways have evolved and how genetic variations in individuals impact on structure, function and disease. We will apply the new computational tools we have developed to improve the handling of mechanisms and their reactions in order to gain a better understanding of reaction space and its impact on pathways. This will also allow improved chemistry queries across our databases. We will continue to use evolutionary approaches to improve our prediction of protein function from sequence and structure. To this end, we have developed a fully automated method that detects and fully characterises channels in transmembrane proteins from their three-dimensional structures. In the ageing project we are interested in tissue specificity and combining human public transcriptome data sets with results from flies, worms and mice to explore effects related to human variation and age.



Characterisation of transmembrane protein channels. Left: PoreWalker visual representation of the KcsA K+ channel. Pore-lining (PL) atoms and residues are coloured in orange and blue, respectively; the rest of the protein is shown in green. Red spheres represent pore centres and their size is proportional to the corresponding pore diameter. IN and OUT indicate cytoplasm and periplasm, respectively. Bottom left: PoreWalker diameter profiles for the low- and high-K+ (solid and dotted line) KcsA channel. The entrance of the selectivity filter is shown in green and highlighted by a green box in the channel structures. Bottom middle: PoreLogo analysis (chain C, PL residues ordered along the pore axis) for the KcsA channel: One Sequence Logo (top) and corresponding Family Logo (bottom). Bottom right: PoreLogo 3D-visualisation: PL residues are shown as 'sticks' and coloured according to the pore-logos above. Labels for the identifiers of the C, F, I and L chains are also shown.

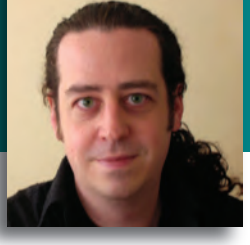
Selected references

Holliday, G.L. *et al*. (2009). Understanding the functional roles of amino acid residues in enzyme catalysis. *J. Mol. Biol.*, 390, 560-577

Nobeli, I. *et al*. (2009). Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.*, 27, 157-167

Pellegrini-Calace, M. *et al*. (2009). PoreWalker: A novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure. *PLoS Comput. Biol.*, 5, e1000440

Selman, C. *et al*. (2009). Ribosomal protein S6 kinase 1 signaling regulates mammalian life span. *Science*, 326, 140-144



Paul Bertone

PhD 2005, Yale University.
At EMBL-EBI since 2005. Group leader since 2006.
Joint appointment with the Genome Biology and Developmental Biology Units.

Differentiation and development

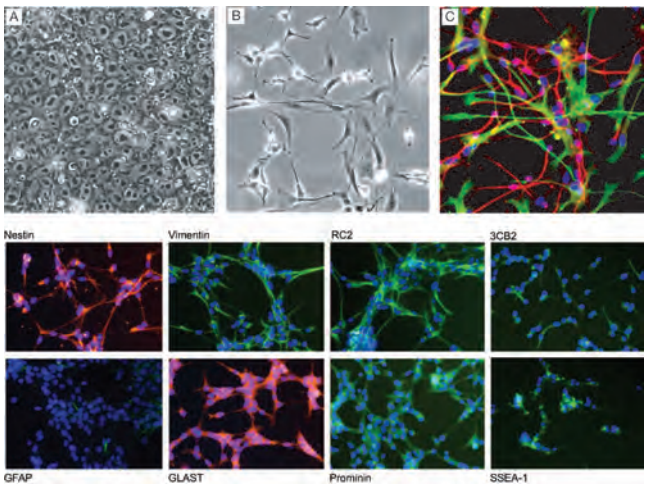
Previous and current research

We investigate the cellular and molecular processes underlying mammalian stem cell differentiation. Embryonic stem (ES) cells are similar to the transient population of self-renewing cells within the inner cell mass of the pre-implantation blastocyst (epiblast), capable of pluripotential differentiation to all specialised cell types comprising the adult organism. 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. 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. We are using high-throughput RNA sequencing to define the comprehensive transcriptional status of stem cells during neural lineage commitment and differentiation to neurons and oligodendrocytes. Using an efficient induced pluripotent stem cell model we will perform the reciprocal analysis to measure global RNA expression during the reversion of neural stem cells to induced pluripotent stem cells. We also plan to use the ChIP-seq approach to capture the epigenetic status of cells undergoing reversion to pluripotency. It is believed that a stabilising process in lineage selection involves progressive restriction of the transcriptional potential of cells as they transition through the lineage hierarchy, mediated through chromatin modifications. This hypothesis suggests that subsequent induction of somatic cells to a pluripotent state would then invoke widespread epigenetic erasure, in order to restore the cell to a state where global lineage commitment options are available. We then aim to elucidate how the functions of several key transcription factors relate mechanistically to their roles in established ES cell self-renewal. Analysing and integrating these genome-wide datasets will allow us to build transcriptional regulatory networks that convey the key mechanistic process of reprogramming neural stem cells to a pluripotent state.



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

Dvinge, H. & Bertone, P. (2009). HTqPCR: High-throughput analysis and visualization of quantitative real-time PCR data in R. *Bioinformatics*, 25, 3325-3326

Tan, D.J.L. *et al.* (2009). Mapping organelle proteins and protein complexes in *Drosophila melanogaster*. *J. Proteome Res.,* 8, 2667-2678

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

Conti, L. *et al.* (2005). Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. *PLoS Biol.*, 3, e283

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

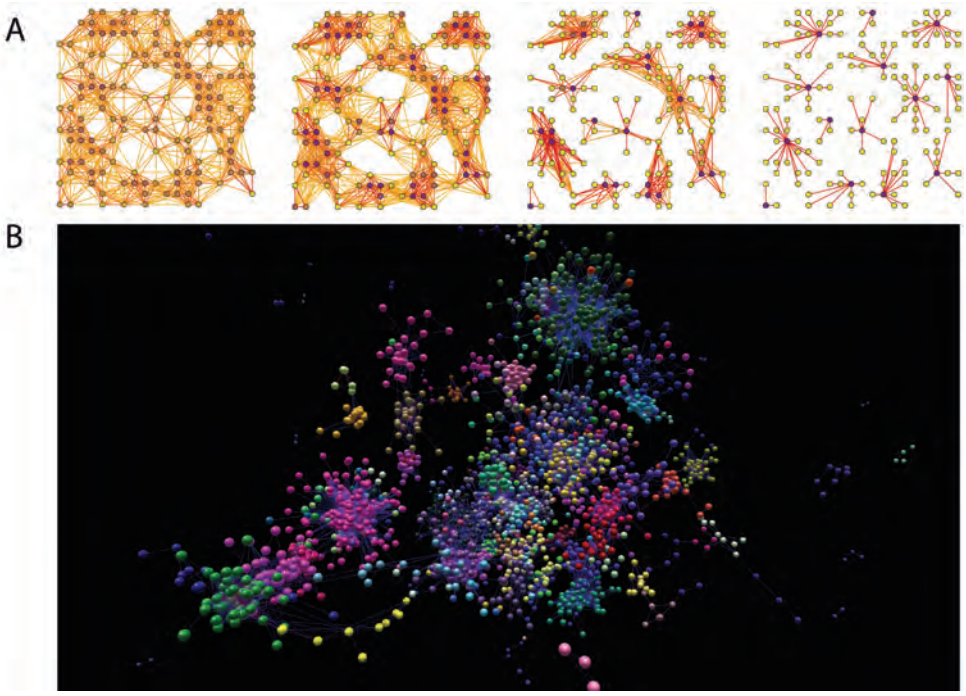
Previous and current research

Complete genome sequencing projects are generating enormous amounts of data. Although progress has been rapid, a significant proportion of genes in any given genome are either not annotated or possess a poorly characterised function. The goal of our group is to predict and describe the functions of genes, proteins, and in particular regulatory RNAs and their interactions in living organisms. Regulatory RNAs have recently entered the limelight as the roles of a number of novel classes of non-coding RNAs have been uncovered.

Our work is computational and involves the development of algorithms, such as the Markov Cluster algorithm (MCL, see figure), protocols and datasets for functional genomics. Our research currently focuses on determining the functions of regulatory RNAs. We are also interested in analysis of biological networks, protein–protein interactions, clustering algorithms and visualisation techniques. We collaborate extensively with experimental laboratories on both the commissioning of experiments and analysis of experimental data. Some laboratory members take advantage of these close collaborations to gain hands-on experience in the wet-lab or perform relevant experiments that support their computational projects.

Future projects and goals

Our long-term goal is to combine regulatory RNA target prediction, secondary effects and upstream regulation into complex regulatory networks. We hope that by building these integrated networks we will be able to place miRNAs into a functional context that will help us to better understand the function and importance of these regulatory molecules. We are extremely interested in the evolution of regulatory RNAs and in developing phylogenetic techniques appropriate for short non-coding RNA. We will continue to build strong links with experimental laboratories working on miRNAs in different systems. In particular such work allows us to build better datasets with which to train and validate our computational approaches. The use of visualisation techniques to assist with the interpretation and display of complex multi-dimensional data will continue to be an important parallel aspect of our work.



A. The MCL graph clustering algorithm alternates dissipation and reinforcement steps. This process is shaped by the connectivity structure in the initial network and leads eventually to a segmented set of smaller networks, interpreted as a clustering of the input. B. A network of the mouse transcriptome linked according to degree of co-expression correlation, clustered using MCL within the BioLayout 3D visualisation system.

Selected references

Lewis, M.A. *et al*. (2009). An ENU-induced mutation of miR-96 associated with progressive hearing loss in mice. *Nat. Genet.*, 41, 614-618

van Dongen, S. *et al*. (2008). Detecting microRNA binding and siRNA off-target effects from expression data. *Nat. Methods*, 5, 1023-1025

Rodriguez, A. *et al*. (2007). Requirement of bic/microRNA-155 for normal immune function. *Science*, 316, 608-611

Saini, H.K., Griffiths-Jones, S. & Enright, A.J. (2008). Genomic analysis of human microRNA transcripts. *Proc. Natl Acad. Sci. USA*, 45, 17719-24

Nick Goldman

PhD, 1992, University of Cambridge.

Postdoctoral research 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.

Evolutionary tools for sequence analysis

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 make predictions about the function of genomic sequence. The group maintains a balance between phylogenetic methodology development and the use of such techniques, focusing on comparative genomics and the bulk analysis of biological sequence data. Continued fruitful collaborations with major sequencing consortia provide the essential state-of-the-art data and challenges to inspire and confront these new methods of sequence analysis. Intra-group collaborations between members involved in theoretical development and those who carry out 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 2009, the group has confirmed its growing strength in analysing biological data, helping bring new insight in the evolution of organisms ranging from bacteria to human, while still developing data analysis theory. Our aim is to continue to increase our understanding of the process of evolution and to provide new tools to elucidate the changing function of biological molecules.

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 (Goldman & Yang, 2008), and we will continue to contribute to this theoretical field. 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.

During 2009 we have become increasingly involved in the study of transcriptional regulation. This will continue and will include consideration of both proteins and non-coding DNA and extension of current TFBS work to vertebrates using ChIP-seq data available through external collaboration. We will continue to work with Martin Taylor, who leaves the EBI at the end of 2009 to establish an independent research group at the MRC Human Genetics Unit in Edinburgh. Building on work initiated at the EBI, he will focus on the detection of selection and mutation patterns using both evolutionary and population genetic resources.

There is no doubt that our involvement with the analysis of data arising from next generation sequencing projects will increase. We also have high hopes for our work on graph alignments. This representation of sequences can be used to account for uncertainty in the input data (e.g. due to high insertion/deletion rates or elevated error rates from some next-generation sequencing technologies), or to capture variation within populations (e.g. for probabilistic modelling of a 'reference sequence' against which to align a population sample).

The TSPAN6 dataset is aligned using the webPRANK alignment server (www.ebi.ac.uk/goldman-srv/webprank/) and the result is displayed in a web browser window. In addition to automatic colour-coding, display of the evolutionary tree and horizontal scrolling, the alignment browser allows for post-processing of the results. Here, alignment columns with low reliability score (lighter shades in the track at the bottom) are selected and shaded light grey. The filtered set of columns can be exported in several different alignment formats for further analyses.

Selected references

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

Goldman, N. & Yang, Z. (2008). Introduction. Statistical and computational challenges in molecular phylogenetics and evolution. *Philos. Trans. R. Soc. B-Biol. Sci.*, 363, 3889-3892

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

Washietl, S. *et al*. (2008). Evolutionary footprints of nucleosome positions in yeast. *Trends Genet.*, 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

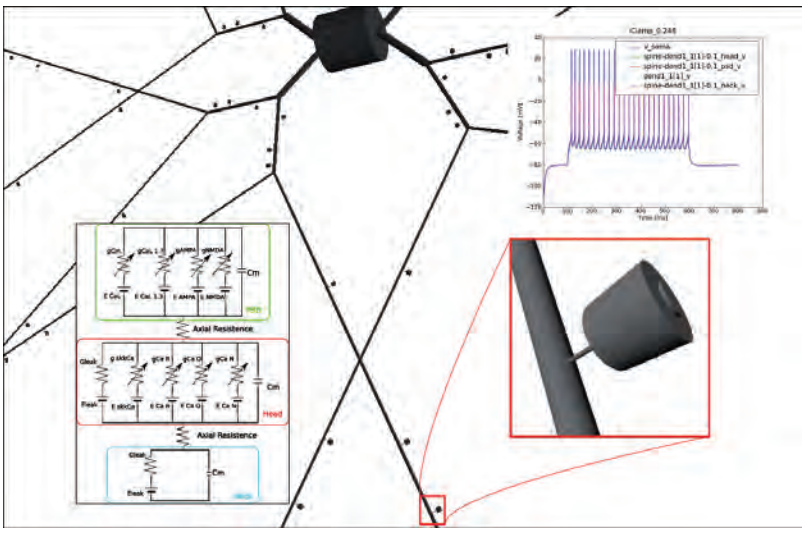
Previous and current projects

The Le Novère group's research interests revolve around signal transduction in neurons, ranging from the molecular structure of proteins involved in neurotransmission to signalling pathways and electrophysiology. In particular, we focus on the molecular and cellular basis of neuroadaptation in neurons of the basal ganglia. By building detailed and realistic computational models, we try to understand how neurotransmitter-receptor movement, clustering and activity, 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. The Systems Biology Graphical Notation (SBGN) 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), a format for describing simulation experiments (SED-ML), and controlled vocabularies (the Systems Biology Ontology, the TErminology for the Description of DYnamics etc.) to improve the models. Finally, a model is only useful if it can be easily accessed and reused. BioModels Database 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 developing 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.



Model of a medium-spiny neuron of the striatum based on the cable approximation. The background image shows the cell body, the dendrites and the dendritic spines. The enlargement shows the structure of a spine. The electrical circuit on the left represents the electrical model of the spine. The plot on the right shows a train of action potentials caused by a current clamp.

Selected references

Li, C. *et al.* (2010) BioModels.net Web Services, a free and integrated toolkit for computational modelling software. *Briefings in Bioinformatics*, in press

Edelstein, S.J., Stefan, M.I., & Le Novère, N. (2010). Ligand depletion *in vivo* modulates the dynamic range and cooperativity of signal transduction. *PLoS One*, 5, e8449

Le Novère, N. *et al*. (2009). The Systems Biology Graphical Notation. *Nat. Biotechnol.*, 27, 735-741

Stefan, M.I. *et al.* (2009). Computing phenomenologic Adair-Klotz constants from microscopic MWC parameters. *BMC Sys. Biol.*, 3, 68

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 Genome Biology Unit, EMBL Heidelberg.

Genome-scale analysis of regulatory systems

Previous and current research

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

Much of our basic knowledge of transcription regulation has derived from molecular biological and genetic investigations. In the past decade, the availability of genome sequences and development of new laboratory techniques have generated (and 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.

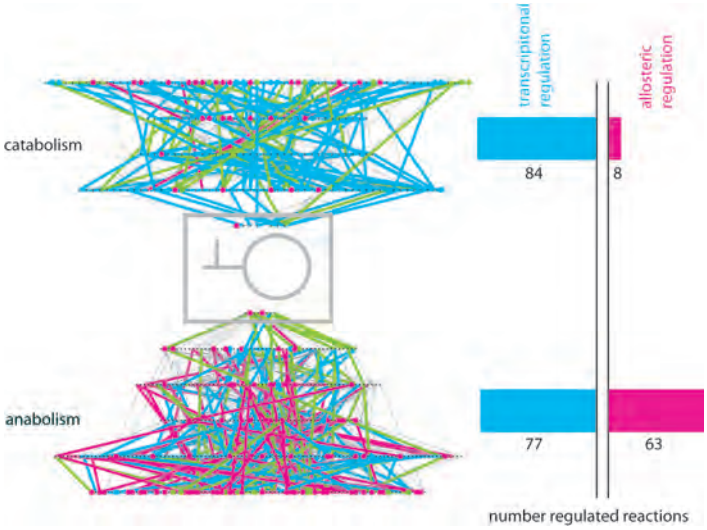
By integrating diverse data sources – from genome sequence to the results of functional genomics experiments – we study the regulatory system at a genomic scale. Since the start of the group in 2005, we have focused our interests to understanding bacterial and eukaryotic gene regulation. Below we describe some of our new findings in these areas.

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 bacterial pathogenicity;
- wet/dry collaborations to understand the epigenetic control of sex determination.

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. We will continue to interact closely with research groups performing functional genomics 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 both (green). Allosteric feedback predominantly regulates anabolic pathways, whereas transcriptional feedback controls both anabolic and catabolic pathways.*

Selected references

Vaquerizas, J.M. *et al.* (2010). Nuclear pore proteins Nup153 and Megator define transcriptionally active regions in the *Drosophila* genome. *PloS Genet.*, in press

Vaquerizas, J.M. *et al.* (2009). A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.*, 10, 252-263

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

Luscombe, N.M. *et al.* (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature*, 431, 308-312

Master in Medicine, 1988 and PhD in immunology, 1989, University of Düsseldorf.

Master in Computer Science, 1993, Passau.

Senior scientist at gsf, Munich and LION bioscience AG, Heidelberg.
Group leader at EMBL-EBI since 2003.

Dietrich Rebholz-Schuhmann

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 and into services for the general public, for example delivery of annotated full text documents as part of UK Pubmed Central (UKPMC).

The research work of the group is concerned with the integration of the scientific literature into the bioinformatics data resources. One important part of this research work is the identification of named entities, e.g. genes, proteins, diseases, species, from the scientific literature, and subsequently linking the entities to a database entry in a reference database, for example UniProtKB for proteins. Both steps are challenging and require the use of natural language processing techniques as well as statistical methods.

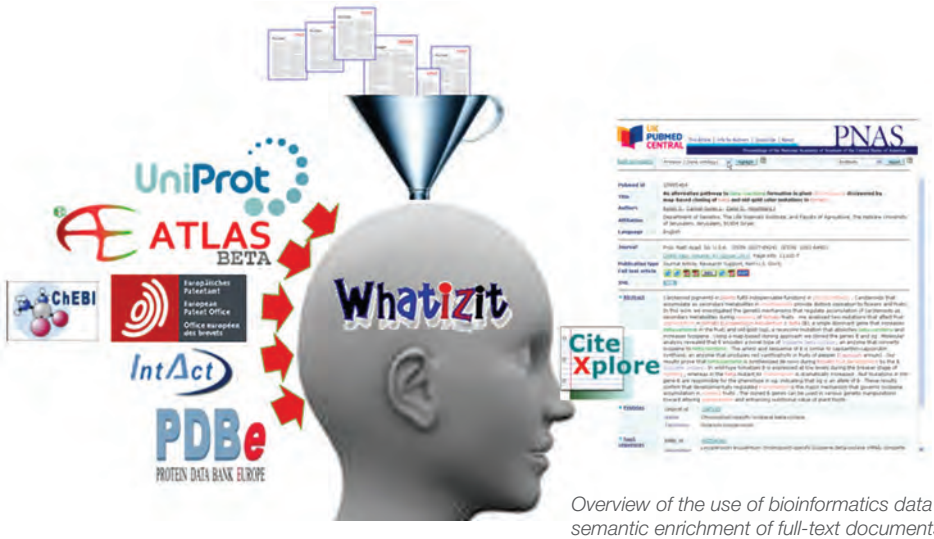
The work in the research group is split into different parts: 1) research work in named entity recognition and its quality control (e.g. UKPMC project); 2) knowledge discovery tasks, e.g. for the identification of gene–disease associations; and 3) further development of the IT infrastructure for information extraction. 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. This work is constantly monitored by automatic means to measure and evaluate the results to identify the most promising solutions (UKPMC project).

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. Disambiguation of semantic types requires special solutions.



Overview of the use of bioinformatics data resources for the standardisation and semantic enrichment of full-text documents as part of the UKPMC project.

Selected references

Jimeno-Yepes, A., Berlanga-Llavori, R. & Rebholz-Schuhmann, D. (2010). Exploitation of ontological resources for scientific literature analysis: searching genes and related diseases. *IEEE Eng. Med. Biol.*, in press

Trieschnigg, D. *et al*. (2009). MeSH Up: effective MeSH text classification for improved document retrieval. *Bioinformatics*, 25, 1412-1418

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

Kim, J.J., Pezik, P. & Rebholz-Schuhmann, D. (2008). MedEvi: Retrieving textual evidence of relations between biomedical concepts from Medline. *Bioinformatics*, 24, 1410-1412.

Julio Saez-Rodriguez

PhD 2007, Max-Planck-Institute & University of Magdeburg.

Postdoctoral research at Harvard Medical School & MIT, 2007-2010.

Group leader at EMBL-EBI since 2010.

Computational analysis of information transfer within signalling networks implicated in disease

Previous and current research

Human cells are equipped with complex signalling networks that allow them to receive and process the information encoded in myriad extracellular stimuli. Understanding how these networks function is a rich scientific challenge, but also of great practical importance, since alterations in the functioning of these networks underlies the development of diseases such as cancer or diabetes. Considerable effort has been devoted to identifying proteins that can be targeted to reverse this deregulation, but their benefit is often not the expected one: it is hard to assess their influence on the signalling network as a whole and thus their net effect on the behaviour of the diseased cell. Such a global understanding can only be achieved by a combination of experimental and computational analysis.

Our group develops computational methods and tools to analyse signalling networks, and we collaborate closely with experimental groups to jointly tackle specific biological questions. We develop mathematical models that integrate high-throughput biochemical data with various sources of prior knowledge, with an emphasis on providing both predictive power of new experiments and insight into the functioning of the signalling network. Towards this end, we combine statistical methods with models describing the mechanisms of signal transduction either as logical or physico-chemical systems. We then use these models to better understand how signalling is altered in human disease and predict effective therapeutic targets.

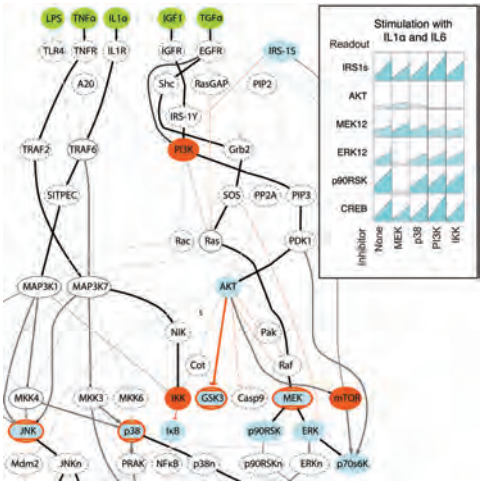
Productive integration of data and computation requires an effective workflow that pulls together all the steps that link experiments to mathematical models and analysis. We are developing, in collaboration with other groups, a platform to facilitate this process by creating a set of interoperable software tools incorporating public standards. We are also involved in a community effort to advance the inference of mathematical models of cellular networks: DREAM (Dialogue for Reverse Engineering Assessments and Methods).

Future projects and goals

Our goals are to continue the development of methods and tools, and their application, to yield insight of medical relevance. We plan to expand the palette of formalisms and levels of detail of the models, from very simple Boolean logic models that coarsely describe signalling networks to differential equation systems describing in detail the dynamics of the underlying biochemical processes. There is no 'right' model formalism and the best option depends on the question and the information available. We will also investigate integration of knowledge from sources such as public databases and data from transcriptomics and genomics.

With these methods we hope to address questions such as:

- What are the origins of the profound differences in signal transduction between healthy and diseased cells and in particular, in the context of cancer, between normal and transformed cells?
- What are the differences in signal transduction among cancer types, and from patient to patient? Can we use these differences to predict disease progression?
- Do these differences reveal valuable targets for drug development? Can we study the side effects of drugs using these models?



Section of a Boolean logical model of the signalling network downstream of seven receptors in the hepatocellular carcinoma HepG2 cell line, obtained by training a general signalling network to data describing the phosphorylation of key proteins upon exposure of combinations of extracellular stimuli and small-molecule inhibitors. Green ellipses denote stimuli, red ellipses species blocked by inhibitors, and blue ellipses readouts. The upper-right inset shows a subset of the data used to validate the model. The model was constructed with CellNetOptimizer and the data processed and visualised with DataRail. Adapted from Saez-Rodriguez et al.*, 2009.*

Selected references

Saez-Rodriguez, J. *et al.* (2009). Discrete logic modeling as a means to link protein signaling networks with functional analysis of mammalian signal transduction. *Mol. Syst. Biol.*, 5, 331

Saez-Rodriguez, J. *et al.* (2008). Flexible informatics for linking experimental data to mathematical models via DataRail. *Bioinformatics*, 24, 840-847

Saez-Rodriguez, J. *et al.* (2007). A logical model provides insights into T-cell signaling. *PLoS Comp. Biol.*, 3, 163

Kremling A. & Saez-Rodriguez, J. (2007). Systems Biology – an Engineering Perspective. *J. Biotechnol.*, 129, 329-351

PhD 1994, University of Heidelberg, Germany.
At EMBL since 1987.
Team Leader at EMBL-EBI since 1994.

Rolf Apweiler

PANDA Proteins and the Apweiler research group

Previous and current research

The activities of the PANDA group are focused on the production of protein sequence, protein family and nucleotide sequence databases at EMBL-EBI. We maintain and host the European Nucleotide Archive, the Ensembl and Ensembl Genomes resources, the UniProt protein resource, the InterPro domain 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 77). In addition, we add information extracted from the scientific literature and curator-evaluated computational analysis whenever possible. The combined InterPro and 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 approaches to improve protein identification from mass spectrometry data and improved data mining of large biological datasets.

Future projects and goals

It is our intention to work on improved integration and synchronisation of all PANDA resources. In addition to major improvements of our current systems, we intend to add mining of high-throughput genomics and proteomics datasets to our automatic annotation toolset. Despite the abundance of data from large-scale experimentation on a genome-wide level, such as expression profiling, protein–protein interaction screens or protein localisation, the systematic and integrated use of this type of information for high-throughput annotation of proteins remains largely unexplored. We therefore intend to build upon 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 the new methodologies to UniProtKB/TrEMBL annotation.

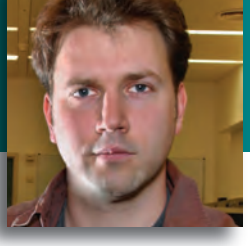
Selected references

Apweiler, R. *et al.* (2009). The Universal Protein resource (UniProt) 2009. *Nucleic Acids Res.*, 37, D169-D174

Barrell, D. *et al.* (2009). The GOA database in 2009 – An integrated Gene Ontology Annotation resource. *Nucleic Acids Res.*, 37, D396-D403

Ilsley, G.R. *et al.* (2009). Know your limits: Assumptions, constraints and interpretation in systems biology. *Biochim. Biophys. Acta*, 1794, 1280-1287

Martens, L. & Apweiler, R. (2009). Algorithms and databases. *Methods Mol. Biol.*, 564, 245-259



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 strategic oversight of the major DNA projects: Vertebrate Genomics, which includes Ensembl (under the leadership of Paul Flicek, page 74); Ensembl Genomes (Paul Kersey, page 78) and the European Nucleotide Archive (Guy Cochrane, page 73). In addition, the HUGO Gene Nomenclature Committee (HGNC), a smaller group coordinated by Elspeth Burford, is part of PANDA Nucleotides.

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 2009 we have seen a striking growth in four 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 association between genetics and disease. We have also launched Ensembl Genomes, a high-quality, community-led genomic information resource for non-vertebrate species. Finally we have launched the European component of the Short Read Archive (SRA). 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 focuses on developing algorithmic methods for genome analysis. Research projects include elucidating the evolution of human genomic sequence since modern humans split from Neanderthals, investigating the role of weak binding motifs in transcription and the evolution of regulatory regions, utilising a novel method for large scale association studies to discover loci associated with various diseases, developing a new tool for DNA analysis, and studying the evolution of development and organogenesis in *Drosophila*.

Future goals and projects

The Birney research group's future projects will encompass two themes: the development of new genomic algorithms and harnessing population variation. Algorithm development will be based around short read assembly and string based methods for structural variation, allowing researchers to handle the large volume of data produced by next generation sequence technology.

Our group, along with several collaborators, will explore intraspecies variation in the context of human disease. This will produce data relating to both basic molecular biology (for example, the binding of transcription factors) and complex disease-related phenotypes (such as MRI scans of beating hearts). Our goal will be to integrate these different data types in the context of the underlying individual genetic differences.

We will also study other species to gain a more basic understanding of how variation in genotype gives rise to differences in phenotype, for example in the development of *Drosophila melanogaster*.

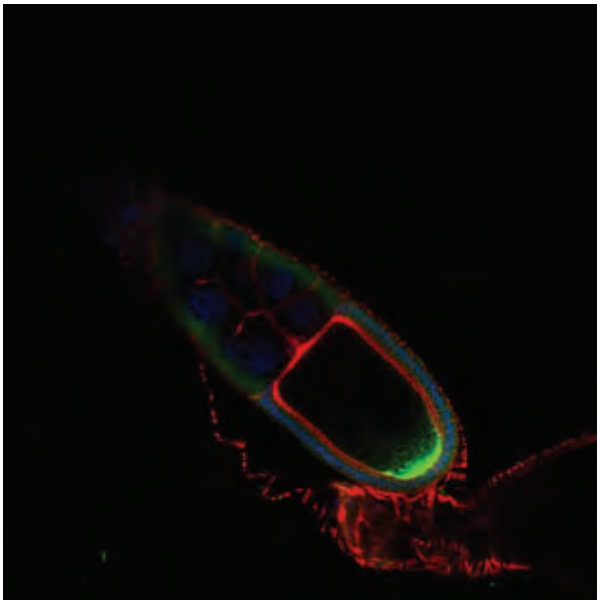


Image of a Drosophila *oocyte, which has been stained for actin (red), nuclei (blue) and oskar protein (green). The amount of each component varies in a complex and consistent way between different wild* Drosophila *individuals.*

Selected references

Paten, B. *et al.* (2008). Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.*, 18, 1829-1843

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

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

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.

The Microarray Informatics Team

Previous and current research

Over the last ten years the focus of bioinformatics has changed dramatically, moving from assembling and analysing reference genomes to looking at individual genome variation and understanding gene function. This was largely driven by the development of new technologies, such as microarrays, and more recently next generation sequencing. Now, the expression of all genes in the genome can be measured in a single experiment. Tens of thousands of such experiments are performed yearly, and the resulting data are used to study complex biological and biomedical problems.

Our team focuses on collecting, curating, integrating, and analysing data from functional genomics experiments, and making these data and the analysis results available to scientists worldwide. We are responsible for two of the EBI's core resources:

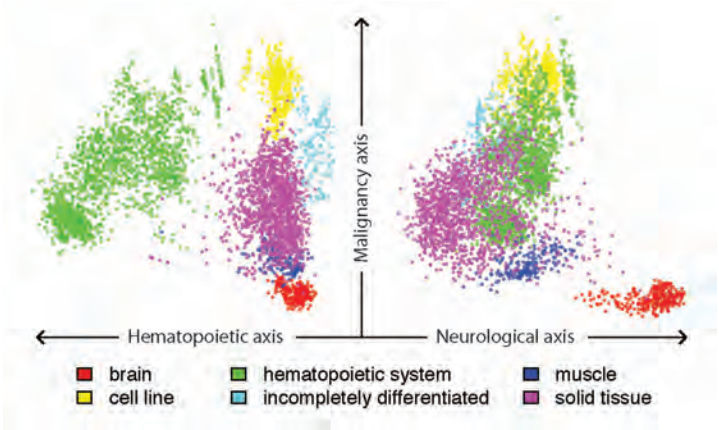
- ArrayExpress Archive of Functional Genomics Data, currently hosting publication-related data from more than 250,000 microarray or next-generation sequencing based assays;
- Gene Expression Atlas (Kapushesky *et al.*, 2010), an added value database making these data accessible to 'bench biologists' by enabling queries about gene expression in different cell and tissue types, or under various disease and other biological conditions.

Our team was among the first to use microarray data to study transcription regulation mechanisms on a genomic scale (Brazma *et al.*, 1998). In collaboration with other groups, we have been building models for systems biology (e.g. Rustici *et al.*, 2004), and developing new methods and algorithms for data analysis. By integrating data from over 5,000 assays of human samples, performed in 163 different laboratories and representing 369 different biological conditions, we have built a global map of gene expression and shown that there are only six major classes of gene expression profiles, representing six major 'continents' (see figure).

Future projects and goals

Our research will continue towards understanding gene expression. What are the main determining factors? How much of gene expression variability is driven by environment and how much by individual genome variation? How is gene expression affected by a disease? How do genes work together to determine the biological state of a cell? We are also developing methods and algorithms for data analysis, mining and visualisation, including developing methods for analysing next generation sequencing data to derive gene expression. We are addressing many of these themes in close collaboration with biologists from laboratories in Cambridge and elsewhere.

On the service side, we have started working on developing the EBI Sample Database, which eventually will hold information about all samples and phenotypes deposited in any of the core databases at EMBL-EBI. We will continue developing the ArrayExpress Archive and Atlas, enriching them with new functionality and data, including the next-generation sequencing data and data from protein expression. We will continue participating in many medically relevant collaborative projects to develop tools for data management and analysis, and to assist our collaborators in data analysis.



Analysis of expression of ~14,000 genes in over 5,000 samples representing 369 different cell and tissue types, disease states and cell lines. Each dot represents a sample in a multidimensional space projected onto the plains defined by the first two principal components (left), and second and third principal component (right). The first three principal components have biological interpretation; we call them Hematopoietic, Malignancy and Neurological axes. Samples belonging to each of the six major transcription profile classes are coloured differently, each representing a 'continent' on the global map of human gene expression. Solid tissue samples that are located higher on the plot (close to cell lines) are predominantly malignant. Adapted from Lukk et al.*, 2010.*

Selected references

Brazma, A. *et al.* (1998). Predicting gene regulatory elements *in silico* on a genomic scale. *Genome Res.*, 8,1202-1215G

Rustici, J. *et al*. (2004). Periodic gene expression program of the fission yeast cell cycle. *Nat. Genet.*, 36, 809-817

Kapushesky, M. *et al.* (2010). Gene Expression Atlas at the European Bioinformatics Institute. *Nucleic Acids Res.*, 38, D690–D698

Lukk, M. *et al.* (2010). A global map of human gene expression. *Nat. Biotechnol.*, in press

Guy Cochrane

PhD, 1999, University of East Anglia.
At EMBL-EBI since 2002.
Team leader since 2009.

The European Nucleotide Archive Team

Previous and current research

The European Nucleotide Archive (ENA) provides a comprehensive repository for public nucleotide sequence data, attracting external users from a multitude of research disciplines and serving as underlying data infrastructure for PANDA services such as Ensembl, Ensembl Genomes and UniProt, as well as broader services such as ArrayExpress. The foundation for the ENA was the EMBL Data Library, which was established in EMBL Heidelberg in the early 1980s. While this component continues to be operated to this day, the mandate of the ENA has expanded enormously as sequencing technology has advanced and the breadth of applications to which sequencing can now be applied has grown.

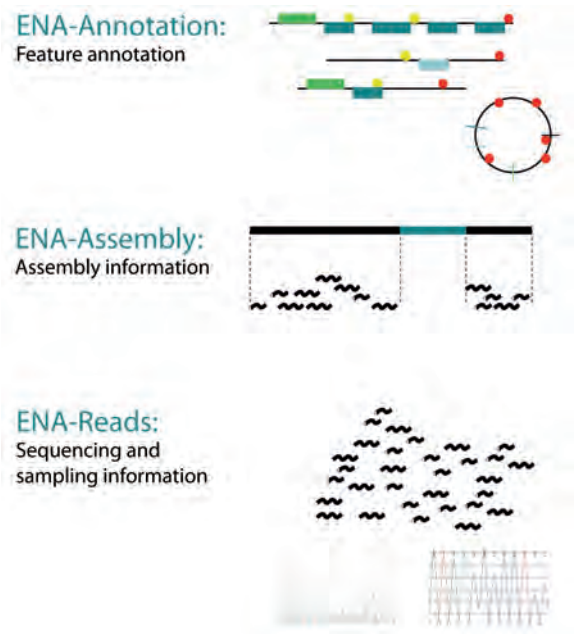
There are three conceptual tiers within the ENA, which reflect an underlying legacy infrastructure that has resulted from the integration of three databases: the original EMBL Data Library (latterly known as EMBL-Bank), the Trace Archive that was established at the Wellcome Trust Sanger Institute in the early 2000s and the newly established Sequence Read Archive (SRA). The three tiers are defined as follows:

- 'Reads': sequencing machine output, base calls and quality scores;
- 'Assembly': information relating overlapping fragmented sequence reads to contigs and covering higher order structures where contigs are structured into representations of complete biological molecules, such as chromosomes;
- 'Annotation': where interpretations of biological function are projected onto coordinate-defined regions of assembled sequence in the form of annotation.

The ENA achieves comprehensive coverage through partnership with the other global bioinformatics service providers, namely NCBI in the US and DDBJ in Japan. The longest running ENA collaboration, the International Nucleotide Sequence Database Collaboration (INSDC; www.insdc.org/), has been underway for over a quarter of a century and now serves as a model for data sharing in the life sciences.

Future projects and goals

The focus of current development work is the presentation of data. A comprehensive browser for all ENA data types is under development, and is expected to be launched in 2010. Accompanying this will be a data retrieval web service and text-based search functionality using the EB-eye search engine. We are currently working on sequence similarity search tools optimised for unassembled next-generation sequence reads, based on de Bruijn graph representations that have been successfully used at EMBL-EBI and elsewhere in sequence assembly work. In addition, we will deliver further submission components as part of our ongoing submissions infrastructure replacement programme and will continue to develop our scalable approach to face the challenging growth in global nucleotide sequencing activity.



Organisation of the ENA into 'Reads', 'Assembly' and 'Annotation' components. Published in Cochrane,G., Martin, M.J. & Apweiler, R. (2010),'Public Data Resources as the Foundation for a Worldwide Metagenomics Data Infrastructure' in Metagenomics: Theory, Methods and Applications, *Caister Academic Press.*

Selected references

Leinonen, R. *et al.* (2010). Improvements to services at the European Nucleotide Archive. *Nucleic Acids Res.*, 38, D39-45

Shumway, M., Cochrane, G. & Sugawara, H. (2010). Archiving next generation sequencing data. *Nucleic Acids Res.*, 38, D870-871

Cochrane, G. *et al.* (2009). Petabyte-scale innovations at the European Nucleotide Archive. *Nucleic Acids Res.*, 37, D19-25

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, D5-12

Paul Flicek

DSc 2004, Washington University, St. Louis, Missouri.
At EMBL-EBI since 2005.
Honorary Faculty Member, Wellcome Trust Sanger Institute since 2008.
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 focusing on genome annotation and variation. The team's research is on computational epigenomics with a particular focus on the integration of diverse data types such as DNA–protein interactions, epigenetic modifications, and the DNA sequence itself in the context of comparative genomics. The major service projects of the Vertebrate Genomics team are Ensembl, the European Genome-phenome Archive, the data coordination centre for the 1000 Genomes Project and the mouse informatics team. In support of these projects, we develop large-scale and novel bioinformatics infrastructure aimed at integrated data analysis and provision of data to the scientific community.

Ensembl (www.ensembl.org) provides an integrated set of tools for chordate genome annotation, data mining and visualisation. A significant achievement for Ensembl over the past year was the launch of the new and more easily navigable Ensembl web interface. From September 2009, Ensembl contained 57 chordate genomes with seven new species added during the year, including the pig, anole lizard, zebra finch and marmoset.

The European Genome-phenome Archive (EGA) is a permanent repository for all types of identifiable personal data, including phenotype information, genotypes, genome sequences and other sequence-based assays. In 2009 the team developed an infrastructure for secure data storage and distribution, as well as an extensive data consistency suite to ensure that the EGA's datasets are internally coherent.

The 1000 Genomes Project (www.1000genomes.org) aims to create the most comprehensive public catalogue of human variation in major world populations by using next-generation sequencing technology. The project's data coordination centre (DCC) is led by the Vertebrate Genomics team in collaboration with the NCBI. During 2009, the data production and initial analysis associated with the pilot phase of the 1000 Genomes Project was completed and the main sequencing phase of the project is expected to be completed in 2010. The Vertebrate Genomics team launched a project-specific genome browser based on the Ensembl platform at http://browser.1000genomes.org.

Future projects and goals

Our work on the analysis of DNA–protein interactions is continuing with the comprehensive and evolutionary-based analysis of individual transcription factors in matched tissues from several species. We are also exploring the utility of DNA methylation profiles for the prediction of genome function.

Developments in services, including support of a public copy number and structural variation (CNV/SV) databases, have been ongoing as part of the larger genome variation effort. These developments will result in the launch of the CNV/SV database in late 2009/early 2010. This will also provide data to Ensembl to ensure that that the CNV/SV data generated over the past several years is made as widely available as possible. The database infrastructure will also be leveraged by the EGA project to provide support for CNV/SV data generated from cohort and disease studies. Ensembl continues to adapt to high-throughput sequencing data and next year we expect to release gene sets and multi-species alignments including genome assemblies created entirely with next-generation sequencing data. Ensembl will also focus on the display and annotation of variation data. This effort is supported by our participation in the Locus Reference Genomic (LRG) consortium (www.lrg-sequence.org), though which we plan to incorporate specific summary data from Locus Specific Databases (LSDBs).

Selected references

Bruce, A.W. *et al.* (2009). Functional diversity for REST (NRSF) is defined by *in vivo* binding affinity hierarchies at the DNA sequence level. *Genome Res.*, 19, 994-1005

Hubbard, T.J.P. *et al.* (2009). Ensembl 2009. *Nucleic Acids Res.*, 37, D690-697

Morley, R.H., *et al.* (2009). A gene regulatory network directed by zebrafish No tail accounts for its roles in mesoderm formation. *Proc. Natl Acad. Sci. USA*, 106, 3829-3834

Vilella, A.J. *et al.* (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.*, 19, 327-335

Midori Harris

PhD 1997, Cornell University, Ithaca, NY.

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

GO Editor at EMBL-EBI since 2001.

The GO Editorial Office

Previous and current research

The Gene Ontology (GO) project (www.geneontology.org) is a collaborative effort to construct and use ontologies to facilitate the biologically meaningful annotation of genes and their products in a wide variety of organisms. At EMBL-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:

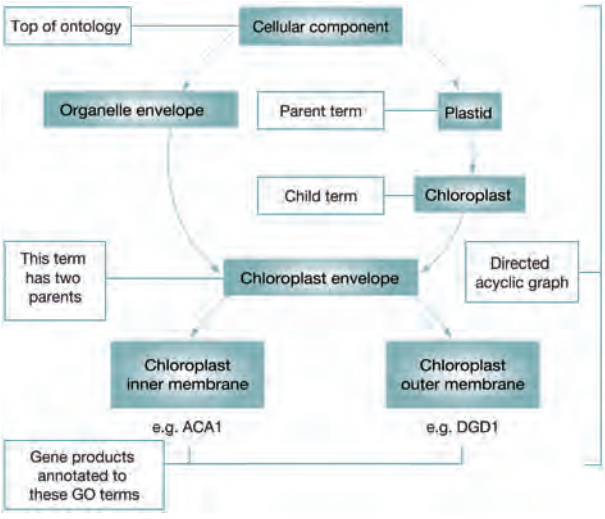
- molecular function defines the tasks performed by individual gene products; examples include aminoacyl-tRNA ligase activity and translation elongation factor activity;
- biological process defines broad biological goals, such as signal transduction or ribosome assembly, that are accomplished by ordered 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 relationships between terms are specified. The GO vocabularies define several semantic relationships between terms: is_a, part_of, and three relations representing biological regulation. The is_a relationship means that a term is a subclass of another; part_of may mean 'physically part of' (as in the cellular component ontology) or 'subprocess of' (as in the biological process ontology). The figure shows a portion of the GO cellular component DAG.

Future projects and goals

The GO Editorial Office will continue to work closely with the rest of the GO Consortium and with biological experts to ensure that the ontologies are comprehensive, logically rigorous and biologically accurate. Improvements begun or continued in 2009 on signal transduction, viral biology, heart development, and other topics will therefore continue. Additional links between the biological process and molecular function ontologies, will be created, using new process-specific function terms. Work on creating cross-products definitions for GO terms will continue, expanding to include orthogonal ontologies such as the ChEBI ontology and the cell ontology.



GO terms are organised in directed acyclic graphs (DAGs) – hierarchical structures in which any 'child' (more specialised term) can have many 'parents' (less specialised terms). For example, the cellular component term chloroplast envelope has two parents, reflecting the fact that it is a part of the chloroplast and a type of membrane. Any gene that is annotated to this term is automatically annotated to both chloroplast and membrane. Some terms and relationships have been omitted for clarity.

Selected references

The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. *Nat Genet.*, 25, 25-29

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

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

The Gene Ontology Consortium (2010). The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res.*, 38, D331-D335

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.

Henning Hermjakob

The Proteomics Services Team

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 New York University and the Ontario Institute for Cancer Research.

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, LipidomicNet, SLING, ENFIN, and ProteomeBinders) 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 further development.

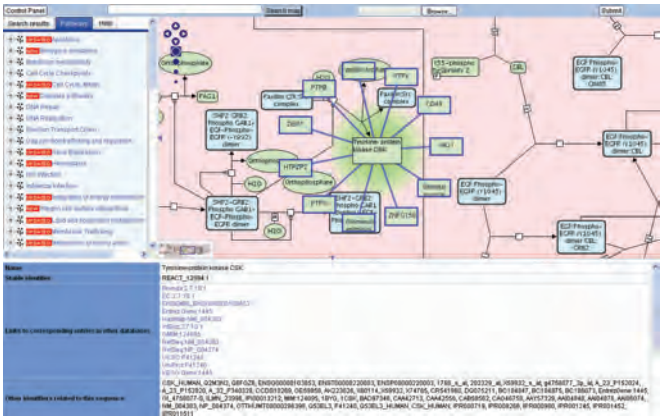
The Proteomics Services team follows an open source, open data approach; all resources we develop are freely available.

Future project and goals

In 2007, our molecular interactions 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 was achieved in the domain of protein identifications, with three published MIAPE modules and the release of the mzML format for mass spectrometry data representation. In 2009, this prior work was recognised by the award of the PSIMEx grant, which for the first time specifically funds our international integration activities. For 2010, we plan to build on these strengths, and expect to initiate regular production mode for international molecular interaction data exchange.

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 and the Distributed Annotation System (DAS). We will also provide closer integration between Reactome pathways and IntAct molecular interactions, by moving current prototypes into production mode.

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.



Reactome prototype Entity Level Viewer, showing the pathway 'Signalling by EGFR', with interactions from the IntAct database shown as a 'halo' around tyrosine-protein kinase CSK.

Selected references

Vizcaíno, J.A. *et al.* (2010). The Proteomics Identifications database: 2010 update. *Nucleic Acids Res.*, 38, D736-742

Aranda, B. *et al.* (2010). The IntAct molecular interaction database in 2010. *Nucleic Acids Res.*, 38, D525-531

Gloriam, D.E. *et al.* (2010). Report: A community standard format for the representation of protein affinity reagents. *Mol. Cell Proteomics,* 9, 1-10

Salwinski, L. *et al.* (2009). Recurated protein interaction datasets. *Nat. Methods*, 6, 860-861

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 currently coordinates the InterPro, CluSTr and Gaia projects at EMBL-EBI.

InterPro is an integrated documentation resource for protein families, domains and functional sites (Hunter *et al.*, 2009). 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, CATH-Gene3D, PANTHER and HAMAP.

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, building relationships between entries where applicable. Additional biological annotation is included, together with links to external databases such as GO, PDB, SCOP and CATH. InterPro pre-computes 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 Ensembl and in the GOA project to provide large-scale mapping of proteins to GO terms.

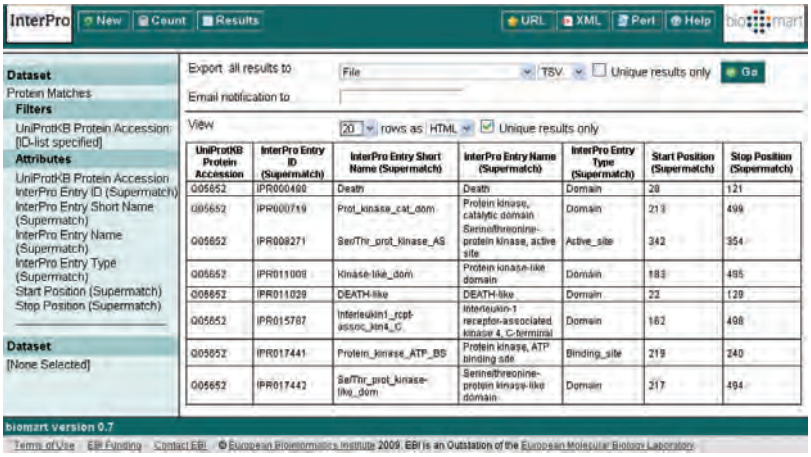
Users access the data from InterPro in various ways, including a web interface at www.ebi.ac.uk/interpro and by utilising the InterProScan protein search software (Quevillon *et al.*, 2005). Recently we have added two new sources of information in the form of a BioMart and via DAS (Distributed Annotation Service). A view of the BioMart can be seen in the figure.

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. It also provides best reciprocal hit orthologue data for a number of complete genomes.

Future projects and goals

An overhaul of the InterPro web interface and web services is ongoing to allow more users easy access to our data. Our intention is to allow for more complex querying and more user-friendly web pages. The InterProScan software package has been re-written to improve its flexibility and modularity and bring it in line with our internal production pipelines and it is our intention to release the new version of the software to the public in early 2010.

The InterPro team is also now responsible for the provision of a resource (known as 'Gaia') to serve the meta-genomics community with the tools necessary for the archiving and analysis of their data, eventually integrating these data with other resources at EBI. A web-based public interface to Gaia will be available by the end of 2010.



A BioMart view of the InterPro entries which contain protein signatures matching the protein Q05652, a Serine/Threonine Kinase. BioMart is a very powerful tool which allows users to federate queries across multiple databases in a very flexible way.

Selected references

Hunter, S. *et al.* (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res.*, 37, D211-215

Quevillon, E. *et al.* (2005). InterProScan: protein domains identifier. *Nucleic Acids Res.*, 33, W116-120

Paul Kersey

PhD 2005, University of Edinburgh.
At EMBL-EBI since 1999.
Team leader at EMBL-EBI since 2008.

Ensembl Genomes

Previous and current research

The Ensembl Genomes team is responsible for the provision of services based on the genomes of non-vertebrate species. 2009 saw the initial release of the five new portals that provide public access to the data under this strategy: Ensembl Bacteria, Ensembl Protists, Ensembl Fungi, Ensembl Plants and Ensembl Metazoa, complementing the coverage of vertebrate genomes in Ensembl. The launch of these sites is in part a response to the huge explosion in data generation triggered by the development of high-throughput sequencing technologies. The Ensembl software system is a modular platform, with different modules capable of handling the analysis, display and distribution of different data types (e.g. variation data, comparative and functional genomics data). Often the raw data is hard to interpret unless integrated into its genomic context. The use of Ensembl technologies enable this, with a data model and browser centred on standard biological concepts (e.g. genes, SNPs, orthologue sets) that can be linked back to the underlying data that provides the evidence for their existence and functional annotation.

The team works collaboratively with a range of experts on individual species or collections of species to ensure that the database continues to provide high quality data. By combining their biological expertise and our own strength in infrastructure, it becomes possible to envisage an Ensembl for every species of significant scientific interest. Through being involved with many communities simultaneously, we have the potential to influence annotation standards, perform comparative analyses of our own and facilitate further analyses by others, in a way that would not be possible if each community developed its own resources in isolation.

We are already working in tight collaboration with other groups producing Ensembl databases including VectorBase (in which consortium we are direct participants), WormBase, the Central *Aspergillus* Database Repository, and Gramene; and additionally represent gene models produced by model organism databases including TAIR, SGD, FlyBase and GeneDB_Spombe. In addition, the group is involved in maintaining four legacy databases that represent different aspects of genomic data: Integr8, Genome Reviews, IPI and ASTD.

Future projects and goals

We are aiming to increase the coverage of Ensembl Genomes to cover all important model species. Species scheduled for inclusion in the first quarter of 2010 include the red bread mould *Neurospora crassa*, the slime mould *Dictyostelium discoides* and the body louse *Pediculus humanus*. Further species will be added throughout the year.

Two EC funded projects have recently started: INFRAVEC, a 30-partner project to produce a research infrastructure for the genetic control of mosquitoes, and Microme, a 14-partner project to produce a new resource for bacterial metabolic pathways. INFRAVEC is one of many projects (alongside others in every domain of Ensembl Genomes) in which we will be ramping up the use of the Ensembl variation infrastructure as population-wide resequencing becomes increasingly common in all species. Microme will be hosted at the EBI and its development offers the potential to improve the annotation of bacterial genomes in the context of our knowledge of metabolism.

As the number of sequenced genomes increases, we are reworking our comparative genomics pipelines to ensure the scalability of the analysis and to integrate the presentation of analyses with different taxonomic scope. The main Ensembl code base was developed for eukaryotic genomes and we are working on improving the representation of genome, gene structure and variation to provide a better fit to the biology of bacterial species. We are also working on the development of Curtain, a memory-efficient pipelining system for the *de novo* assembly of large genomes from short read data.

Selected references

Kersey, P. *et al.* (2010). Ensembl genomes: Extending Ensembl across the taxonomic space. *Nucleic Acids Res.*, 38, D563-D569

Kersey, P. *et al.* (2005). Integr8 and Genome Reviews: integrated views of complete genomes and proteomes. *Nucleic Acids Res.*, 33, D297-D302

Kersey, P. *et al.* (2004). The International Protein Index: An integrated database for proteomics experiments. *Proteomics*, 4, 1985-1988

Lawson, D. *et al.* (2009). VectorBase: a data resource for invertebrate vector genomics. *Nucleic Acids Res.*, 37, D583-587

Gerard Kleywegt

PhD 1991, University of Utrecht.
Postdoctoral research, University of Uppsala.
Coordinator and then programme director of the Swedish Structural Biology Network (SBNet), 1996–2009.
Appointed Professor of Structural Molecular Biology, University of Uppsala, 2009.
Team leader at EMBL-EBI since 2009.

The Protein Data Bank in Europe (PDBe) Team

Previous and current research

The Protein Data Bank in Europe (PDBe, www.ebi.ac.uk/pdbe) is one of the five core molecular databases (genomes, nucleotides, proteins, 3D structures, and expression data) hosted by EMBL-EBI. PDBe holds detailed knowledge of structure and function of biological macromolecules and access to this information is vital for many different users, for example, in the identification of potential targets for therapeutic intervention as well as of lead structures for pharmaceutical use. PDBe usage averages approximately1.8 million web hits per month from around 22,000 unique web hosts. In addition, we serve over 1,350 FTP addresses with 375 GB of data on average per month. A combination of EMBL and Wellcome Trust funding supports the core staff and the computer hardware that is essential for the PDBe to store, give access to and integrate the deluge of data relating to 3D molecular structures.

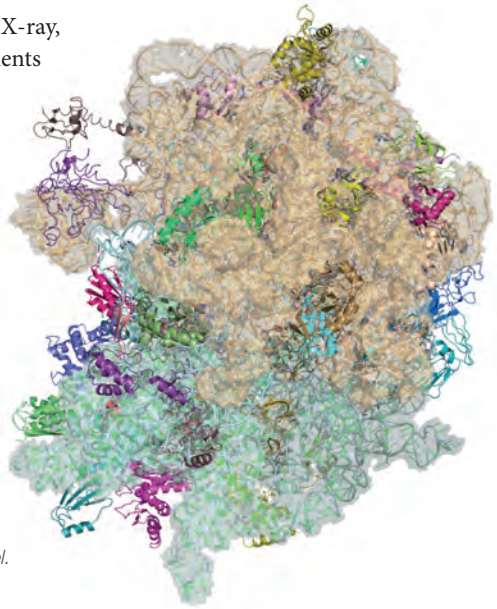
Through our membership of the Worldwide Protein Data Bank (wwPDB) organisation we are an equal PDB partner with the United States (RCSB) and Japan (PDBj) and all partners work closely to maintain the single international archive for structural data. We also integrate the experimental data derived by 3D cryo-electron microscopy and electron tomography techniques, and derive the molecular biological assemblies of structures held in the PDB.

Future projects and goals

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, the aim of PDBe is to continue to provide integrated data resources that evolve with the needs of structural biologists. To achieve this goal, the main objectives for PDBe in 2009 and onwards are to:

- expertly handle deposition and annotation of structural data as one of the wwPDB deposition sites. We aim for an average turn-around time of one working day or less, employ expert annotators with experience in structure determination, and eventually intend to handle approximately one third of all depositions worldwide;
- provide an integrated resource of high-quality macromolecular structures and related data. This is implemented by developing and maintaining advanced structural bioinformatics databases and services that are, or even define, the state-of-the-art. They should be kept up to date to keep pace with the growth of the PDB archive and ideally be available on a 24/7 basis for over 360 days per year;
- maintain in-house expertise in all the major structure-determination techniques (X-ray, NMR and EM) in order to stay abreast of technical and methodological developments in these fields, and to work with the community on issues of mutual interest (e.g. data representation, harvesting, formats and standards, or validation of structural data).

The wwPDB partners are committed to the development of a common deposition and annotation tool. The steering committee reviewed the proof-of-concept testing phase in July 2009. The key infrastructure architecture components covering the data-sharing technology and workflow technology were demonstrated and the team will produce a significant workflow example by January 2010 and a completed component by June 2010. The project is expected to be finished in 2011.



In October 2009, the Nobel Prize for Chemistry was awarded to three prominent structural biologists, Venki Ramakrishnan, Tom Steitz and Ada Yonath, for their studies on the structure and function of ribosomes. All the results of their structural studies (and those of many others) are available from the PDB and EMDB. This figure, created by Jawahar Swaminathan (PDBe), shows the structure of the Thermus thermophilus *70S ribosome (PDB entries 2WDI and 2WDG). The PDBe has a website with information about the prize-winning structures at www.ebi.ac.uk/pdbe/docs/nobel.*

Selected references

Velankar, S. *et al.* (2010). PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.*, 38, D308-317

Dutta, S. *et al.* (2009). Data deposition and annotation at the worldwide protein data bank. *Mol. Biotechnol.*, 42, 1-13

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

Henrick, K. *et al.* (2008). Remediation of the protein data bank archive. *Nucleic Acids Res.*, 36, D426-433

Vet. Med Degree 1984, Oslo Vet. Hoyskole.
Cand. Scient. Molecular Toxicology and Informatics 1987, University of Oslo.
Team leader at EMBL-EBI since 1995.

Rodrigo Lopez

The EMBL-EBI External Services Team

Previous and current research

The External Services (ES) team is responsible for web development, web administration and services frameworks at EMBL-EBI.

Web development includes developing and maintaining the main EBI web portal as well as several Wellcome Trust, BBSRC and EU-funded projects' websites. The group is currently working on a migration plan that uses DRUPAL (www.drupal.org) as the core system for both www.ebi.ac.uk as well as all the EU project portals.

During 2009 the web developers, in collaboration with the University of Manchester, launched the BioCatalogue, a curated catalogue of life science web services (www.biocatalogue.org). This project, funded by the BBSRC, focuses on capturing, annotating and cataloguing web services in bioinformatics from around the globe.

Web administration focuses on monitoring and reporting the status of the various EBI web portals and services. ES have recently developed a usage reporting system based on mySQL database and the AWSTATS package.

Services frameworks comprise tools for generating core services, such as the EBI search engine, EB-eye, and applications for the maintenance of core analytical tools, especially in the domain of nucleotide and protein sequence analysis. The team developed a new version of the dispatcher application framework. This now provides a single point of maintenance for all core sequence analysis services, including those from remote as well as in-house collaborators. These new frameworks can be accessed at www.ebi.ac.uk/Tools/sss (for sequence similarity search applications) and www.ebi.ac.uk/Tools/msa (for multiple sequence alignment tools).

Future projects and goals

The developers in the group are collaborating with various groups and teams across the institute on a data integration effort that aims to provide a more descriptive view of results obtained via search engines, such as the EB-eye. The goal is to summarise key information in the core data resources about genes and their products in well characterised organisms. The data types on which the system is being built represent the central dogma of molecular biology: genomes contain genes, genes are transcribed, transcripts translate to proteins and proteins form structures. These represent data from Ensembl and ENA, ArrayExpress, UniProtKB and PDBe, respectively. Secondary data resources, such as InterPro, PRIDE and new resources such as ChEMBL are also part of the effort. Domain specific literature is also being built into the system, to provide the user with a brief, but comprehensive view of publications pertinent to the various data types.

There are several focal points for the development and implementation of the system that currently include: the EB-eye Web Services, Distributed Annotation System (DAS), as well as Web Services from literature and secondary data resources.

During 2010, the group will focus on the deployment of services in two data centres based in London. These will house all the compute resource required for service provision from EBI as well from service providers under the upcoming ELIXIR project. The move to London will result in improved accessibility and reliability of services and signifies a major change for developers and service maintainers at the EBI and its collaborators.

Selected references

Li, W., *et al.* (2010). Non-redundant patent sequence databases with value-added annotations at two levels. *Nucleic Acids Res.*, 38, D52-56

McWilliam, H. *et al.* (2009). Web services at the European Bioinformatics Institute – 2009. *Nucleic Acids Res.*, 37, W6-10

Goble, C. *et al.* (2009). BioCatalogue: A Curated Web Service Registry For The Life Science Community. *Nature Precedings*, DOI:10.1038/npre.2009.3132.1

Johanna McEntyre

PhD Plant Biotechnology, 1990, Manchester Metropolitan University.

Editor, Trends in Biochemical Sciences, Elsevier, Cambridge.

Staff Scientist, NCBI, National Library of Medicine, NIH, USA.

Team leader at EMBL-EBI since 2009.

Literature resource development

Previous and current research

The biomedical literature is the formal record of achievement and scientific understanding of the biomedical research community. As such, searching, reading and browsing the literature are key components of all research strategies. Furthermore, many biological and biomedical databases, including those produced at EMBL-EBI, are heavily linked to relevant portions of the scientific literature to provide functional information and supporting evidence.

The EBI currently hosts a database of biomedical abstracts called CiteXplore, which provides a good foundation for further growth of the literature resources. CiteXplore content is imported from a number of sources, including: PubMed (from the US National Library of Medicine), AGRICOLA (from the USDA-National Agriculture Library), patents (from the European Patent Office), Chinese Biological Abstracts (CAS-SICLS) and CiteSeer. There are currently over 22 million abstracts available, of which about 19 million are PubMed abstracts. There are approximately 1.9 million patents, with the remaining data sources completing the full dataset.

UKPMC is a freely accessible archive of the full text of peer-reviewed, biomedical research articles, supported by all major UK funding agencies. It is based on PubMed Central, built at the National Library of Medicine (NLM), USA. The role of the Literature Services group in the development of UKPMC functionality has been to broker the search and retrieval mechanisms between the data (i.e. full-text research articles) and the web interface, include added value from text mining, and make new metadata content available via CiteXplore.

Future projects and goals

Improvements to CiteXplore: In collaboration with the British Library, the Literature Services team will identify more content for UKPMC, such as theses, which will be added to CiteXplore and exposed in the UKPMC interface. We also plan to add all non-PubMed indexed UKPMC records to CiteXplore. Improvements will also be made to CiteXplore search logic and the team will also index and expand the database and citation link data to extend browsing and cross database integration.

Full text capabilities: The UKPMC collaboration will continue to move towards a search and retrieval interface that incorporates new text mining features. The Literature Services group has a key role to play in this process, as a provider of the search index as well as helping to productise the results of text mining from the Rebholz-Schuhmann research group (page 68). A significant piece of work within this remit will be to form a robust pipeline for content, text mining and indexing updates within the context of UKPMC.

The Literature Services group is also a SLING partner, with a deliverable of indexing the full text of patents, supplied by the European Patent Office. This will augment the UKPMC full-text research articles and provide further breadth and depth to the full-text search services.

Leveraging text mining: We plan to maximise the use of the Whatizit text-mining resource and other resources built by the Rebholz-Schuhmann group, with related components (such as vocabularies or ontologies) to enhance search, retrieval, and browsing, and to improve integration of the literature resources with other databases at EMBL-EBI.

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.

Computational chemical biology: the ChEMBL Team

Previous and current research

The ChEMBL team's research and service activities centre around the area of computational chemical biology (or chemogenomics), with specific focus on the discovery of chemical probes for biological systems, leading to an understanding of normal and disease biology. A major application of such approaches is in the prioritisation and lead discovery of drug-like molecules for use as innovative therapies.

The group is funded by a five-year Wellcome Trust Strategic Award, and tasked to develop publicly available databases of drug discovery relevant data, specifically large-scale structure activity relationship (SAR) data.

The ChEMBL database (ChEMBLdb) is the primary respository for the bioactivity and drug data for the group. Data is entered manually from the literature using outsourcing contracts – this task is not amenable to automation due to the way in which SAR data is presented in the medicinal chemistry and pharmacological literature. The total size of the ChEMBLdb exceeds 620,000 compound records, covering over 3,600 distinct protein targets. We have also established a network of EBI-based specialist curators with expertise in particular areas of biology. The core SAR data is now freely available, as is an easy to use web front-end (www.ebi.ac.uk/chembldb). Uptake of this has been strong and vigorous, and follows on from an active pre-release data sharing programme with over sixty leading academic and industrial groups, in which we gathered feedback on data organisation, quality and community needs. This has already led to the use of the data in several high impact publications.

Future projects and goals

Toxicity prediction: Given the large volume of binding and functional data within ChEMBLdb, we are starting to address a new and complementary area of drug discovery and development – toxicology. Potential therapeutic compounds are tending to fail at a later stage of clinical trials with toxicity liabilities. Surprisingly, there are few well organised, publicly accessible data on the toxicity properties of clinical development candidates and drugs. We are investigating a number of collaborative approaches to populate new resources with toxicity related data, both for human therapeutics, but also for other life science areas, such as cosmetics, herbicides and so forth.

Pre-competitive data sharing: EMBL-EBI is a core data provider for many resources for industry, and this role is now further extended by the content of ChEMBLdb. We are thus well placed to act as an informatics hub to facilitate pre-competitive data sharing, whilst leveraging our extensive network of existing contacts. This will lead to the development of a generic registration and deposition system for small molecule bioactivity data.

Resource integration: The ChEMBLdb team is also working towards establishing effective data sharing agreements with key resources around the world, ranging from PubChem (http://pubchem.ncbi.nlm.nih.gov/) the chemspider resource (www.chemspider.com), TDR targets (www.tdrtargets.org), NC-IUPHAR (www.iuphar-db.org), Binding-DB (www.bindingdb.org), DrugBank (www.drugbank.ca) as well as seamless integration with other EBI resources.

Outreach and training: We are planning a wide range of community engagement and training for our new resources in 2010, including conference presentations, lab visits to geographical clusters of labs interested in the data, and also a series of training workshops.

Selected references

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

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

Agüero, F. *et al.* (2008). Innovation – Genomic-scale prioritization of drug targets: the TDR Targets database. *Nat. Rev. Drug. Discov.*, 7, 900-907

Berriman, M. (2009). The genome of the blood fluke *Schistosoma mansoni*. *Nature*, 460, 352-358

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 also 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 eScience project) and for the EMBRACE project that integrates access to bioinformatics tools and data content through standard-compliant web services.

To date, Grid development has focused 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 participation in 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 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 fourth 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 was 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 group, 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 themselves to be highly useful to prototype and develop service and metadata standards. We are looking, especially through the EMBRACE project, to migrate to true Grid services, but like many other groups we are waiting for the long-anticipated merging of web and grid service standards.

The EMBOSS project plans to expand in the coming few years to cover bioinformatics more generally, including genomics, protein structure, gene expression, proteomics, phylogenetics, genetics and biostatistics. This will require the participation of external groups to expand the project beyond its current EBI base, and we are actively seeking potential partners in each area. We will expect to build a service-based eScience architecture around the applications and data resources through the EMBRACE project, with support and guidance from the community of users in academia and industry.

The EMBRACE project is now in its final phase. The services have been developed and standardised. We are now developing metadata standards including an ontological description of the data types used and the methods provided. These will be annotated within each service's Web Service Definition Language (WSDL) file, and queries through the EMBRACE registry portal.

Selected references

Carver, T. *et al.* (2009). DNAPlotter: Circular and linear interactive genome visualization. *Bioinformatics*, 25, 119-120

Gibson, A. *et al.* (2009). The data playground: An intuitive workflow specification environment. *Fut. Gen. Comp. Systems*, 25, 453-459

Haider, S. *et al.* (2009). BioMart central portal – Unified access to biological data. *Nucleic Acids Res.*, 37, W23-27

Smedley, D. *et al.* (2009). BioMart – Biological queries made easy. *BMC Genomics*, 10, 22

Ugis Sarkans

PhD in Computer Science, University of Latvia, 1998.
Postdoctoral research at the University of Wales, Aberystwyth.
At EMBL–EBI since 2000.
Technical team leader since 2006.

The Microarray Software Development Team

Previous and current research

Our team has been developing software for ArrayExpress since 2001. As of October 2009, ArrayExpress holds data from more than 260,000 microarray hybridisations and is one of the major data resources of EMBL-EBI. ArrayExpress supports the following community standards: MIAME (Minimum Information About a Microarray Experiment), MAGE-ML (Microarray Gene Expression Markup Language), and MAGE-TAB (MAGE tabular).

In 2009 we have been working on replacing the MAGE-ML centred infrastructure with one based around MAGE-TAB. This effort will significantly simplify all internal data management tasks and will enable us to concentrate more on providing added value for our users.

The software development team has built the following components of the ArrayExpress infrastructure:

- Repository – the archival database for the data that support publications;
- Data Warehouse – a query oriented database of gene expression profiles (in 2009 replaced by the Gene Expression Atlas, built by Misha Kapushesky in the Microarray Informatics team);
- MIAMExpress – a data annotation and submission system;
- Expression Profiler – a web-based data analysis toolset;
- Components used internally by the ArrayExpress production team.

The ArrayExpress Repository interface has been restructured by introducing a new powerful search mechanism, based on Apache Lucene, the most popular open source, full text indexing and search technology. This development makes it easier to add new visual search capabilities to the interface. Keyword-based search behaviour is closer to that provided by PubMed which is widely regarded as the *de facto* standard in bioinformatics.

The web services for accessing ArrayExpress data programmatically were improved by adding more powerful search capabilities and making the interface more consistent. The user interface has also been improved by making the experiment view layout easier to use.

A continuing aspect of ArrayExpress maintenance has been improvements to the robustness and usability of ArrayExpress, and troubleshooting when necessary. The other area of significant effort has been the system for reannotating the array design descriptions submitted to ArrayExpress and used in the Gene Expression Atlas.

We also continued working in the field of medical informatics with MolPAGE Data Warehouse (MoDa), a proof of concept database for managing multi-omics data. Significant effort was devoted to reannotating bioentities measured in methylation, metabolomics, protein and tissue array experiments. A report of this work has been submitted for publication.

Future projects and goals

The main goal for 2010 is finishing the testing of the new generation ArrayExpress infrastructure, data migration and roll out. The ArrayExpress repository interface will continue to receive incremental updates; in particular, there are still some aspects of the interface that are served by older software (viewing array designs and protocols), and these will be replaced with new components.

There is a recent initiative in EMBL-EBI to clean up and aggregate aspects of biological sample information that are served by different EBI data resources. We will exploit our experience with dealing with various aspects of biological sample information management and reuse and adapt the relevant parts of the ArrayExpress software for these purposes.

Selected references

Krestyaninova, M. *et al.* (2009). A system for information management in BioMedical studies – SIMBioMS. *Bioinformatics*, 25, 2768-2769

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

Brazma, A. Krestyaninova, M. & Sarkans, U. (2006) Standards for systems biology. *Nat. Rev. Genet.*, 7, 593-605

Sarkans, U. *et al.* (2005) The ArrayExpress gene expression database: A software engineering and implementation perspective. *Bioinformatics*, 21, 1495-1501

Christoph Steinbeck

PhD 1995, Rheinische Friedrich-Wilhelm-Universität, Bonn. Postdoctoral research at Tufts University, Boston and the MPI for Chemical Ecology, Jena, 1997-2002.

Habilitation 2003, Organic Chemistry, Friedrich-Schiller-Universität, Jena, 2003.
Head of group, Cologne University Bioinformatics Center, 2002-2007.

Lecturer in Chemoinformatics, Tübingen University, 2007.
Team leader at EMBL-EBI since 2008.

Chemoinformatics and metabolism

Previous and current research

The Chemoinformatics and Metabolism team works on methods to elucidate, organise and publish the small molecule metabolism 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. Further, current research is aimed at developing methods to predict the metabolome of organisms based on their genetic blueprint and other boundary information.

We are also 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.

In addition the team hosts the following resources:

Chemical Entities of Biological Interest (ChEBI) database: a freely available dictionary of molecular entities focused on 'small' chemical compounds. It was initiated to provide standardised descriptions of molecular entities that enable other databases at EMBL-EBI and worldwide to annotate their entries in a consistent fashion. ChEBI focuses on high-quality manual annotation, non-redundancy and provision of a chemical ontology rather than full coverage of the vast range of chemical entities.

Intenz: The Integrated relational Enzyme database (IntEnz; www.ebi.ac.uk/intenz/) provides a complete, freely available database focused on enzyme nomenclature approved by the NC-IUBMB (www.iubmb.org), combined with additional information from the ENZYME database (www.expasy.org/enzyme/). Currently, IntEnz contains 4,150 approved entries as well as proposed new entries and revisions of previously published entries. IntEnz is jointly maintained by SIB (the Swiss Institute of Bioinformatics) and EMBL-EBI.

Rhea: a reaction database where all reaction participants (reactants and products) are linked to ChEBI (see above) which provides detailed information about structure, formula and charge. While the main focus of Rhea is enzyme-catalysed processes, it can include reactions not contemplated in IUBMB enzyme nomenclature, or even reactions not catalysed by enzymes.

Future projects and goals

In 2010 we will increase our research activities in modelling small molecule metabolism and predicting metabolomes. This work will be done in collaboration with external laboratories and with a focus on particular application domains and organisms.

We will also work closely with our collaborators at SIB to produce an enzyme portal involving the Rhea and IntEnz databases, providing a one-stop shop for enzyme related data. Another focus for the team will be to ensure a sustainable growth for the ChEBI database and automatic classification of existing entries within the ChEBI ontology. Not only do these tasks 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.

The integration of major open source software such as OrChem and JChemPaint into our services projects has been a major activity in 2009. We will continue to improve these projects, for example adding SMARTS querying and extending the structure editor to deal with all types of structures, such as polymers.

Selected references

de Matos, P. *et al.* (2010). Chemical entities of biological interest: an update. *Nucleic Acids Res.*, 38, D249-254

Spjuth, O. *et al.* (2009). Bioclipse 2: A scriptable integration platform for the life sciences. *BMC Bioinformatics*, 10, 397

Rijnbeek, M. & Steinbeck, C. (2009) OrChem – An open source chemistry search engine for *Oracle. J. Cheminf.*, 1, 17

Blinov, K.A. *et al.* (2008). Performance validation of neural network based 13C NMR prediction using a publicly available data source. *J. Chem. Inf. Model.*, 48, 550-555

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.

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 requires upgrading in order to realise the full potential of data-rich resources, and optimise the usage of our human and hardware resources.

Data distribution or data synchronisation has been a very active research area in the information technology sector for quite a few years. As a result, some useful tools, such as rsync and its variants, have been developed. The core of the technology, also known as delta compression or delta encoding, is to find the differences (deltas) between two sets of files located remotely from each other, and only those deltas are transferred to the target computer to allow it to rebuild a new version of the files based on the older version. If the deltas are significantly smaller than the full data files, a significant network saving can be achieved. The technology is working very well for some applications, such as internal data synchronisation and remote software distributions. However, it has not been successfully adopted by the database community for the distribution of large datasets, although it has been attempted by many people. When applied to large datasets, the latency of the runtime calculation to identify the deltas between source and target files becomes a new major bottleneck. Also, unclustered changes in files, which are common in bioinformatics databases, can result in full data transfer instead of deltas. A few other technologies, such as peer-to-peer solution and data replication, have also been tried by some bioinformaticians. Although some progress has been made, their application in distributing large-scale biological databases is still very limited.

This year we are focused on developing a new algorithm, sdeltac, to shorten the network latency for distributing large datasets across the network. The algorithm is a delta encoding-based solution but by taking advantage of some of the unique characteristics of biological databases, the algorithm is also a structure-based differential distribution system that overcomes the limitations of existing delta encoding solutions.

Future projects and goals

The continuing development and optimisation on sdeltac, on both algorithm and implementation, will be our focus for next year. New developments will include:

- algorithm adaptation and implementation work on compressed file formats;
- further development of Relational Database Management Systems, to allow selection of a subset of a database and version skipping;
- collaborations with database projects and data centres to test the utilisation of the program as a potential data distribution and data mirroring solution;
- commencing work on interfaces for the client to retrieve delta data from the RDBMS database.

EMBL Grenoble France

The EMBL outstation in Grenoble, France, 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. The ESRF crystallography beamlines are now highly automated and all are equipped with EMBL-designed high-precision diffractometers and frozen crystal sample changers. A new X-ray small-angle scattering instrument built by ESRF and EMBL is now operational with a custom designed small-volume automatic sample changer.

High-throughput methods have also been introduced in other steps of the structure determination process, a development closely connected with the outstation's involvement in several European structural proteomics integrated projects. 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 are now available to external users under the EU funded P-CUBE project (www.p-cube.eu). They also 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 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 and 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 2010 a new top-end electron microscope with cryo- tomography capability will become available.

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 groups working on TFIID, complex enhanceosomes and the dosage compensation complex. A molecular cell biology group is also working on the biology of micro-RNAs, in particular trying to elucidate the role of piRNAs in the germ line.

Another major focus is the study of RNA viruses, particularly influenza virus, with the aim of understanding how it replicates and also as a target for anti-viral drug design. 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.

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 Genome Biology Unit.
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.

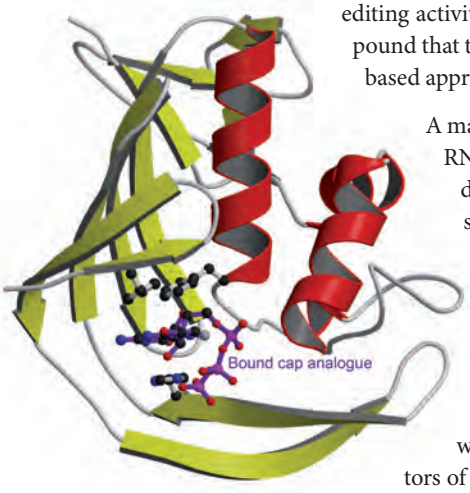
In eukaryotic cells, nascent Pol II RNA transcripts (e.g. mRNA or snRNA) are rapidly given an m7Gppp cap at the 5' end. The nuclear cap-binding complex (CBC) binds to this tag and mediates interaction with nuclear RNA processing machineries for splicing, poly-adenylation and transport. We determined the structure of human CBC, a 90KDa heterodimeric protein and its complex with a cap analogue, and are currently working on structures of several other proteins involved in cap-dependent processes. We have also worked on the structure of the protein PHAX which binds to CBC and is specifically involved in nuclear transport and export of small capped non-coding RNAs (e.g. snRNAs). 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. We have also determined the structure of the complete, heterodimeric, tri-functional vaccinia virus mRNA capping enzyme, which caps viral transcripts in the cytoplasm of infected cells.

Aminoacyl-tRNA synthetases play an essential role in protein synthesis by charging specifically their cognate tRNA(s) with the correct amino acid. We aim to obtain atomic resolution structural information to help us understand the catalytic mechanism of the enzymes and their substrate specificity for ATP, cognate amino acid and tRNA. 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 contain a large inserted editing domain able to recognise and hydrolyse mischarged amino acids, essential for maintaining translational fidelity.

Future Projects and Goals

We have several ongoing projects related to RNA metabolism, aiming to obtain and use structures of the complexes involved to understand function. These include continued studies on PHAX and ARS2, both of which bind CBC and are linked to the metabolism of small RNAs. Concerning NMD, we recently published the structure of the UPF1-UPF2 complex, and are now trying to determine the architecture of the entire trimeric UPF1-UPF2-UPF3 complex. Work is continuing on the leucyl-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 that are active against multi-drug resistant strains.

A major goal is the structure determination 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 co-founded a company to pursue this). We are also engaged in fluorescence studies of the transport and assembly of the influenza polymerase in living cells. Finally we work 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.



Ribbon diagram showing the cap-binding domain of influenza virus polymerase subunit PB2 (yellow and red) with bound cap analogue (purple).

Selected references

Clerici, M., Mourao, A., Gutsche, I., Gehring, N.H., Hentze, M.W. *et al.* (2009). Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. *EMBO J.*, 28, 2293-2306

Dias, A., Bouvier, D., Crepin, T., McCarthy, A.A., Hart, D.J., Baudin, F., Cusack, S. & Ruigrok, R.W. (2009). The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*, 458, 914-8

Guilligay, D., Tarendeau, F., Resa-Infante, P., Coloma, R., Crepin, T. *et al.* (2008). The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat. Struct. Mol. Biol.*, 15, 500-506

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

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.

Structural complexomics of eukaryotic gene expression

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 focused 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.

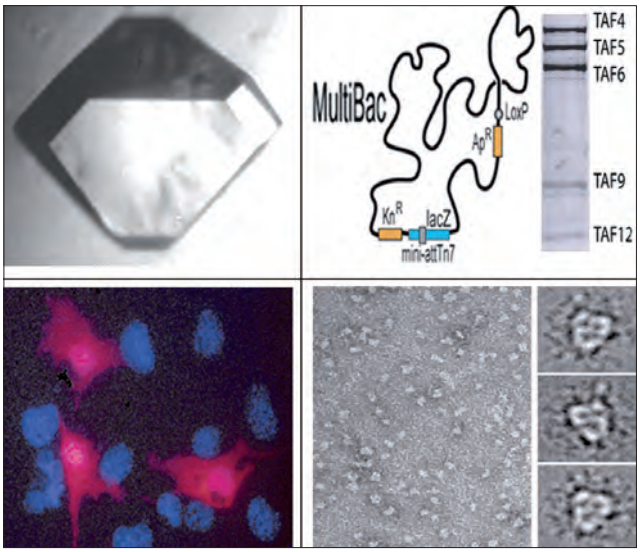
Our major methodologies include molecular biology and X-ray crystallography. However, we readily apply techniques from other fields that are required for our research, both by ourselves and through collaborative efforts. A prerequisite for successful structural study of many complexes, both by electron 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 developed our MultiBac system, a modular, baculovirus-based technology specifically designed for eukaryotic multiprotein expression. MultiBac is now being used successfully in many laboratories worldwide, not only by structural biologists but also for applications as diverse as the development of vaccines and gene therapy vectors. In our lab, we have recently harnessed homologous and site-specific recombination methods in tandem for all steps involved in multigene assembly, and we have successfully implemented all steps involved in a robotics setup by developing ACEMBL, a proprietary automated suite for multigene recombineering on our TECAN EvoII platform. By using our technology, we produced numerous large multiprotein assemblies in sufficient quantity and quality for structural studies, including large multicomponent membrane protein complexes and a 1 MDa core assembly of human TFIID general transcription factor.

Future projects and goals

At EMBL Grenoble, we 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, its various isoforms and other components of the preinitiation complex. In collaboration with the Schaffitzel group (page 96) and the Schultz group at IGBMC Strasbourg, we subject the complex specimens produced to electron microscopic analyses. We use homogenous complexes thus identified for X-ray crystallography. We strive to understand physiological function, and we explore and challenge our structural findings by *in vitro* and *in vivo* biochemical analysis.

By enlisting state-of-the-art mass spectrometric methods from systems biology, we are addressing a further bottleneck in complex crystallography, namely the challenge of defining crystallisable core assemblies of multiprotein complexes in a reasonable time frame (a collaboration with ETH Zürich and Lund University). Also, we are expanding 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., Nie, Y., Frey, D., Olieric, N., Schaffitzel, C., Collinson, I., Romier, C., Berger, P., Richmond, T.J., Steinmetz, M.O. & Berger, I. (2009). Automated unrestricted multigene recombineering for multiprotein complex production. *Nat. Methods*, 6, 447-50

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

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

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

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

We develop instruments and methods for optimising X-ray scattering experiments in collaboration with the Synchrotron Crystallography Team (page 93) and the ESRF Structural Biology group, as well as contributing to the development of the EMBL@PETRA-III beamlines (see page 104). Dissemination of our technology to other synchrotrons is also an important aspect of our work.

The macromolecular crystallography (MX) beamlines rely on our MD2/MD2M diffractometers, SC3 sample changers and C3D crystal centring software to process several hundred crystals per day. The EMBL@PETRA-III MX1 beamline will also be equipped with a customised version of the MD2 diffractometer. An important task for our team is the maintenance and continuous upgrade of these instruments.

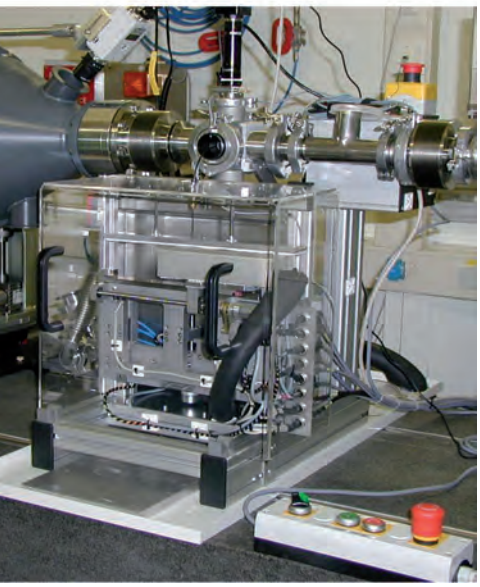
After a year of use at the ESRF MX-beamlines, the HC1b crystal dehydration device has proven to be efficient for improving the diffraction quality of protein crystals and contributed to the success of challenging projects such as the MSL proteins (Cusack group, page 88). MAX-lab and the DLS have recently acquired an HC1b, and a collaboration defines protocols for improving the dehydration of protein crystals. A new version of the dehydration system with a sample temperature control to 5 °C, HC2, is under development.

In the framework of our collaboration with EMBL Hamburg and the ESRF, an automated BioSAXS sample environment with liquid transfer robot has been developed and installed on the ESRF/EMBL ID14-3 BioSAXS beamline. Up to eight samples and associated buffers can be automatically exposed to X-rays, and the system is optimised for sample volumes down to 5 µls with temperature control of both the sample storage and exposure cell. A fully integrated software environment with the beamline has been in use since June 2009, a copy of which is currently being validated at the X33 beamline at EMBL Hamburg. The development of a high capacity version of the machine has already started.

MX demand has now entered a phase where the capacity of existing beamline robotics is insufficient. The design of sample holders strongly impacts on the number of samples that can be affordably shipped and processed, so we have started preliminary studies on a new sample holder that is more precise and compact.

Crystal harvesting is the last step that has remained resistant to automation in MX. In collaboration with the Grenoble HTX team (page 92), we are exploring a new technology to automatically harvest crystals grown in specifically designed crystallisation supports on which we have filed a patent.

The BioSAXS sample changer at ID14-3.



Future projects and goals

The ESRF and EMBL@PETRA-III BioSAXS beamlines will both be equipped with the final version of the automated BioSAXS sample environment. Meanwhile the HC2 dehydration device will become operational and dehydration protocols made available to users.

We will start developing a vertical kappa goniometer with a sub micrometer sphere of confusion to equip the MX2 beamline of EMBL@PETRA-III, and to upgrade the EMBL/ESRF ID23-2 micro-focus beamline. We also plan to contribute in the development of the future generation of automated beamlines, playing a driving role in the definition of a future sample holder standard for frozen crystals and associated robotic tools.

Together with the HTX team, our main objective for the future is to bridge crystallisation and data collection. Beyond automating the harvesting step, we propose a fully integrated screening platform where users can bring either protein in solution or crystallisation plates and obtain X-ray data with minimal handling. The EMBL/ESRF/Indian BM14 beamline could host such a system and become a fully integrated pilot screening platform.

Selected references

Sanchez-Weatherby, J., Bowler, M.W., Huet, J., Gobbo, A., Felisaz, F., Lavault, B., Moya, R., Kadlec, J., Ravelli, R.B. & Cipriani, F. (2009). Improving diffraction by humidity control: a novel device compatible with X-ray beamlines. *Acta Crystallogr. D Biol. Crystallogr.*, 65, 1237-46

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

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

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.

High-throughput protein technologies

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.

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.

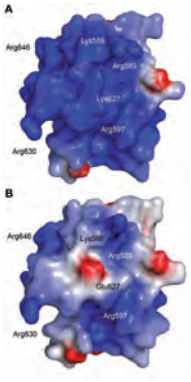
Selected references

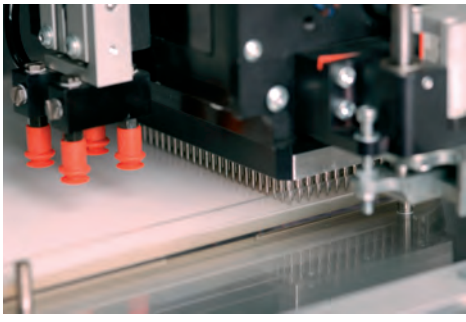
Moriniere, J., Rousseaux, S., Steuerwald, U., Soler-Lopez, M. *et al.* (2009). Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature*, 461, 664-668

Dias, A., Bouvier, D., Crepin, T., McCarthy, A.A., Hart, D.J., Baudin, F., Cusack, S. & Ruigrok, R.W. (2009). The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*, 458, 914-8

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. *et al.* (2007). Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat. Struct. Mol. Biol.*, 14, 229-233





Top: A previously unsuspected domain from influenza polymerase, identified by HT 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.

Right: 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.

PhD 1997, University of Valencia.
Postdoctoral research at EMBL.
Staff scientist at EMBL Grenoble since 2003.
Team leader since 2005.

José A. Márquez

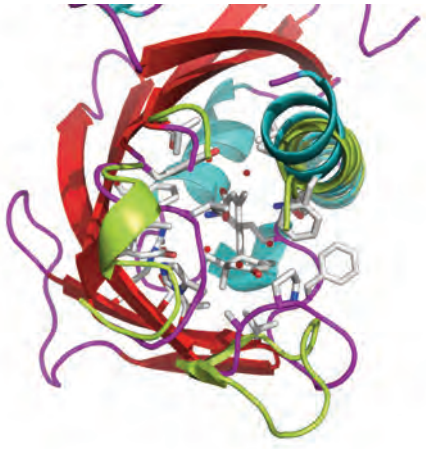
The high throughput crystallisation lab at the PSB

Previous and current research

Automated crystallisation: Obtaining crystals from biological molecules is 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, which can consume large amounts of sample and time. We have established a fully automated high-throughput crystallisation screening facility, the HTX lab (https://htxlab.embl.fr), which processes large numbers of samples using very low sample volumes. Since its opening, we have offered services to hundreds of scientists performing several millions of experiments. The HTX lab is next to the ESRF synchrotron and offers crystallisation screening services to scientists working in European research institutions through the EU-funded FP7 programme PCUBE.

One of the major problems we face is accurately capturing the enormous flux of information generated by experimental parameters and results. As a result, we have developed the Crystallisation Information Management System (CRIMS), which tracks experiments and makes results available to users via the web in real-time along with all the experimental parameters. It is also synchronised with the automated data collection systems at the ESRF and has been distributed to other European labs. The HTX lab not only increases efficiency in the process of structure solution, but also represents a technological advantage critical for the success of challenging projects like those studying protein complexes and large macromolecular assemblies.

Sensing and signalling at a structural level: Our research focus is in understanding the mechanisms of signaling at a structural level and specially sensing processes. In the past we have contributed to the understanding of the mechanisms of action of the HPrKinase/phosphatase, a sensor of the metabolic state in bacterial cells. More recently, we have obtained the structure of the Abscisic Acid (ABA) receptor, a hormone regulating the response to environmental stress in plants. This receptor belongs to the so called PYR/PYL/RCAR, family and is able to bind ABA and inhibit the activity of specific protein phosphatases of the type 2C (PP2Cs) leading to the activation of the signalling pathway controlling the stress response. Our work shows how the hormone is bound in a large cavity inside the PYR1 ABA receptor. The loops surrounding the entry to this cavity act as 'gates', closing over the hormone once it is inside. Amino acids in these same loops are also important for the interaction between PYR1 and the phosphatases, implying that binding of the hormone to the receptor and the closure of the gating loops generates an optimal surface for binding of the PP2Cs. This work represents the definitive confirmation of the PYR/PYL/RCAR protein family as ABA receptors, but also provides insights into the basic mechanism of hormone sensing.



Structure of the Abscisic acid hormone receptor showing the gating loops in the closed (magenta) and open (green) conformations.

Future projects and goals

Developing new technologies in crystallisation: With the new EC-funded PCUBE programme giving access to the platform to European scientists we will extend the base of HTX lab users. We will also develop an improved version of the CRIMS data management software specially designed to allow easy installation and customisation in other labs. In collaboration with the Syncrotron Instrumentation Group (page 90) and based on solutions we have already explored and patented, we plan to develop an integrated screening station where users can bring either the protein in solution or crystallisation trays, and get X-ray data with minimum input. This will require the development of new crystallisation supports and crystal harvesting robotics.

In the laboratory, we will continue the study of signaling systems with a special focus on two major fields: TEC-family protein tyrosine kinases and the ABA signaling pathway. The structure of the ABA hormone-receptor complex paves the way for the design of small molecules able to bind to the ABA receptors and activate the stress signalling pathway. These molecules should be easier to synthesise and more stable than ABA itself and could potentially be used to improve the tolerance of crops to drought and other type of environmental stress. We will also focus on the structural study of other components of the ABA signalling pathway.

Selected references

Santiago, J., Dupeux, F., Round, A., Antoni, R., Park, S.Y. *et al.* (2009). The abscisic acid receptor PYR1 in complex with abscisic acid. *Nature*, 462, 665-668

Santiago, J., Rodrigues, A., Saez, A., Rubio, S., Antoni, R. *et al.* (2009). Modulation of drought resistance by the abscisic acid receptor PYL5 through inhibition of clade A PP2Cs. *Plant J.*, 60, 575-588

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

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.

Synchrotron Crystallography Team

Previous and current research

The team works in close collaboration with the Structural Biology group at the ESRF in the design, construction and operation of macromolecular crystallography (MX) beamlines. We are currently responsible for two MX beamlines, ID14-4 and ID23-2, and the BioSAXS beamline ID14-3. Last year the tuneable beamline ID14-4 was the the first European MX beamline to celebrate a millennium of PDB depositions. It continues to be routinely used by the groups of Venki Ramakrishnan and Ada Yonath, who were awarded the 2009 Nobel prize in chemistry with Thomas Steitz for their 'studies on the structure and function of the ribosome'. Equally, the microfocus beamline ID23-2 produces many structures of biologically important molecules, while the highly-automated BioSAXS beamline ID14-3 has already produced several high impact publications in its inaugural year of operation. In January 2010 the team also took over the management of the CRG beamline BM14 at the ESRF, which for the next five years will be run as a partnership with the Indian government and the ESRF, enabling access for Indian and 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. Recent examples that were funded by SPINE2-Complexes include the prototype BioSAXS liquid handling robot now installed on ID14-3, as well as software for optimal crystal reorientation strategies and the use of X-ray tomography in MX.

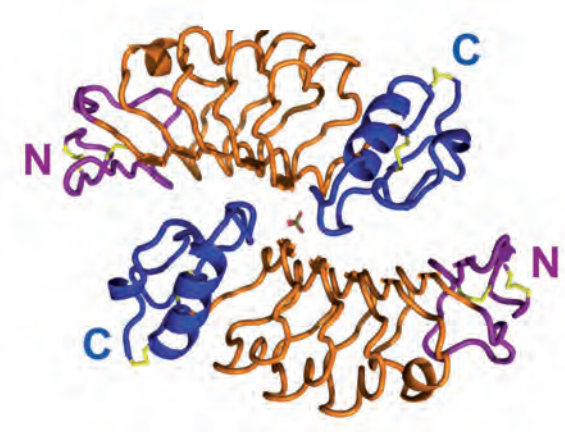
The team also studies proteins involved in neuronal development. We are particularly interested in the Slit-Robo signalling complex that is essential for the normal development of the central nervous system and has also been implicated in heart morphogenesis, angiogenesis and tumour metastasis. With part funding by SPINE2Complexes we have determined a number of structures (see figure) from this system that may be 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, already publishing 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 the creation of new ecologically-friendly and pest-resistant plants.

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 collection of optimal data from ever smaller crystals. On ID14-3 our team, in collaboration with the Instrumentation Team, the ESRF and EMBL Hamburg, will be actively involved in the provision of a highly automated BioSAXS beamline. On BM14 we will continue to build the new partnership with India and the ESRF for running the beamline in addition to managing a major upgrade of the optical components. We hope that all our combined efforts will push the boundaries of structural biology to better understand the function 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.

Structure of the Slit2 dimerisation domain. Slit2 D4 N- and C-terminal caps are in purple and blue, respectively, the LRRs 1-5 are in orange and the disulphide bridges are in yellow.



Selected references

McCarthy, A.A., Brockhauser, S., Nurizzo, D., Theveneau, P. *et al.* (2009). A decade of user operation on the macromolecular crystallography MAD beamline ID14-4 at the ESRF. *J. Synchrotron Radiat.*, 16, 803-812

Seiradake, E., von Philipsborn, A.C., Henry, M., Fritz, M., Lortat-Jacob, H., Jamin, M. *et al.* (2009). Structure and functional relevance of the Slit2 homodimerization domain. *EMBO Rep.*, 10, 736-741

Dias, A., Bouvier, D., Crepin, T., McCarthy, A.A., Hart, D.J., Baudin, F., Cusack, S. & Ruigrok, R.W. (2009). The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature*, 458, 914-8

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

PhD 1999, University of Basel.
Postdoctoral research at Harvard University, Boston.
Group leader at EMBL Grenoble since 2007.

Daniel Panne

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.

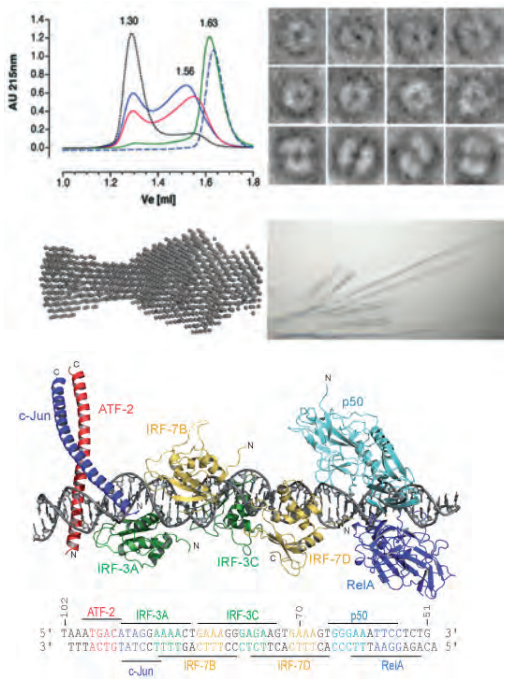


Figure 1 (top): We employ a number of different resolution techniques to visualise cellular structures. Figure 2 (bottom): Atomic model of the INF-β enhanceosome.

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-β 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

Ramesh Pillai

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

Regulation of gene expression by non-coding RNAs

Previous and current research

The aim of our research is to understand the molecular mechanisms by which non-coding RNAs regulate gene expression. We apply a combination of methods ranging from biochemistry, cell biology, bioinformatics and mouse genetics in these studies.

MicroRNAs (miRNAs) are an abundant class of ~21 nucleotide (nt) small non-protein-coding RNAs that function mostly as negative regulators of gene expression. Mis-expression of miRNAs has been implicated in human cancers, underscoring the relevance of these RNAs in human health. miRNAs recognise their mRNA targets by complementary base-pairing and prevent accumulation of their protein product. This is achieved by either mRNA degradation or repression of translation. We previously demonstrated that miRNAs inhibit an early step of translation initiation which then leads to accumulation of the mRNA targets in cytoplasmic granules called processing bodies (P-bodies). Our current research is aimed at understanding how miRNA homeostasis is maintained in the cell. MicroRNAs are processed from longer precursors via several intermediates and this processing is subject to several post-transcriptional regulatory mechanisms. We are investigating factors involved in modulating miRNA biogenesis and studying their regulatory role during early development.

Approximately 40% of the mouse genome is composed of repeats that pose a considerable threat to the genome as several maintain the ability to jump from one location to another. Germline genomes are particularly vulnerable as the protective DNA methylation that silences the transposons is removed for a short period during germ cell development to allow extensive epigenetic reprogramming. Piwi-interacting RNAs (piRNAs) are ~30 nt long germline small RNAs that form the RNA-mediated component of defence mechanisms against transposons and function as part of an RNP by associating with the piwi clade of Argonaute proteins. Our aim is to understand the mechanisms of piRNA biogenesis and function. Using biochemical methods we recently identified the Tudor domain-containing protein 1 (Tdrd1) as a key component in the piRNP complex. Along with other groups in the field we have shown that piwi proteins are post-translationally modified by symmetrical dimethyl arginines, and that the recognition of these modifications via the tudor domains mediates interaction of piwi proteins with tudor proteins. In the absence of Tdrd1, the piwi protein accumulates piRNAs with an altered profile, suggesting a role in piRNA biogenesis. We are now trying to gain mechanistic understanding of how tudor proteins contribute to the piRNA pathway.

Future projects and goals

Purification of piwi proteins has uncovered additional novel factors in the pathway and these will be further characterised. Collaborations with structural groups to obtain atomic resolution structures of pathway components will add another dimension to our understanding of small RNAs. In the future we hope to use live cell imaging techniques to study assembly of RNPs *in vivo* and define the contribution of the individual constituents of the complex to this process. Investigation of the regulatory role of long noncoding RNAs will be another major focus of investigation in the lab.

Selected references

Reuter, M., Chuma, S., Tanaka, T., Franz, T., Stark, A. & Pillai, R.S. (2009). Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat. Struct. Mol. Biol.*, 16, 639-46

Shoji, M., Tanaka, T., Hosokawa, M., Reuter, M., Stark, A. *et al.* (2009). The TDRD9-MIWI2 complex is essential for piRNA-mediated retrotransposon silencing in the mouse male germline. *Dev. Cell*, 17, 775-87

Pillai, R.S., Bhattacharyya, S.N. & Filipowicz, W. (2007). Repression of protein synthesis by miRNAs: how many mechanisms? *Trends Cell Biol.*, 17, 118-26

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, 1573-6

PhD 2001, University of Zürich, Switzerland.
Habilitation 2008, ETH Zürich, Switzerland.
Team leader at EMBL Grenoble since 2007.

Christiane Schaffitzel

Ribosomal complexes: targeting, translocation and quality control

Previous and current research

Research in our laboratory combines molecular biology, *in vitro* and *in vivo* biochemistry and single-particle cryo-electron microscopy (cryo-EM) to study the structure and function of ribosomal complexes. Synthesis and folding of proteins require concerted interactions of the translating ribosome with translation factors, regulatory factors, molecular chaperones and factors involved in the export of proteins. Structures of translating ribosomes in complex with these factors provide critical insight into the interaction networks, stoichiometry and molecular mechanism of these megadalton-size complexes. With cryo-EM, we can study the mutlicomponent translation machinery at close to physiological conditions. By using state-of-the-art electron microscopes and by image processing of large data sets, ribosomal structures at subnanomolar resolution have been obtained recently, demonstrating the power of this method.

A prerequisite for our functional and structural studies is the production of large amounts of homogenous, stable complexes in quantity and quality required for interaction assays, mass spectrometry and single-particle cryo-EM. In our laboratory, we established bacterial and eukaryotic cell-free translation systems for the *in vitro* generation of ribosomes displaying homogenous nascent polypeptide chains. We reconstitute the ribosomal complexes along the pathways of co-translational targeting and translocation and mRNA quality control. This approach was successfully applied in the case of the structures of a translating ribosome-signal recognition particle (SRP) complex (figure 1), of the complex of the ribosome with the translocation machinery SecYEG (figure 2), and of the ribosome in complex with the molecular chaperone Trigger Factor. The data from intermediate resolution structures derived from cryo-EM, in conjunction with high-resolution structures of the ribosome and of the isolated factors, were combined in a hybrid approach to generate quasi-atomic models of the ribosomal complexes involved. The structural data supported by biochemical data, provide important and detailed snapshots of the mechanisms underlying these cellular processes ensuring correct folding, targeting and translocation of nascent proteins.

Future projects and goals

We study ribosomal complexes involved in membrane protein integration, folding and assembly. We analyse the membrane protein complexes biochemically, by nanoelectrospray mass spectrometry (a collaboration with Carol Robinson, University of Cambridge, UK) and cryo-electron microscopy.

In collaboration with the groups of Stephen Cusack (page 88), Matthias Hentze (page 51) and Andreas Kulozik (Molecular Medicine Partnership Unit, EMBL/University of Heidelberg), we study human ribosomal complexes involved in nonsense-mediated mRNA decay. We produce these complexes by means of advanced recombinant eukaryotic technologies in collaboration with the Berger group (page 89).

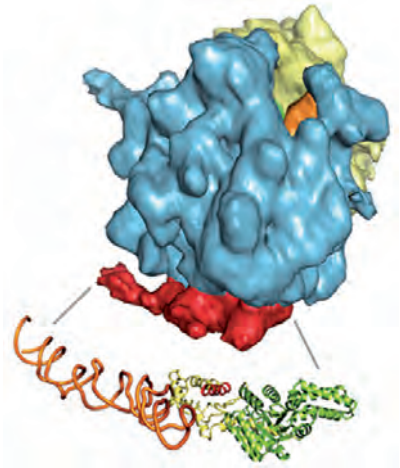


Figure 1: Cryo-EM structure of the E. coli *signal recognition particle bound to a translating ribosome. Ribosomal subunits are shown in yellow (30S) and blue (50S), tRNAs in green and orange, the SRP in red. Below the atomic model of* E. coli *SRP as fitted into the density is depicted.*

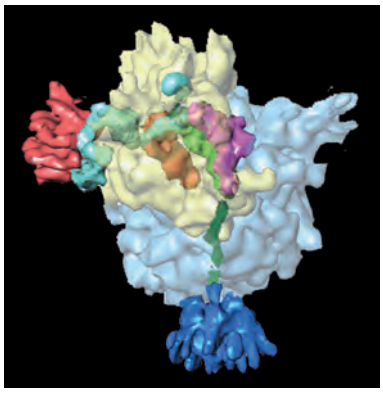


Figure 2: EM reconstruction of the ribosome nascent chain complex and the translocon. The 30S ribosomal subunit is shown in yellow, the 50S subunit in light blue, the A, P and E site tRNAs are in magenta, green and orange respectively. The mRNA is depicted in cyan and the nascent chain in the polypeptide 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

Kohler, R., Boehringer, D., Greber, B., Bingel-Erlenmeyer, R., Collinson, I., Schaffitzel, C. & Ban, N. (2009). YidC and Oxa1 form dimeric insertion pores on the translating ribosome. *Mol. Cell*, 34, 344-353

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, 7183, 108-11

Schaffitzel, C., Oswald, M., Berger, I., Ishikawa, T., Abrahams, J.P. *et al.* (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. *et al.* (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, 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 at the PETRA III ring (pictured). It will comprise three state-of-the-art beamlines for protein crystallography and small angle X-ray scattering, complemented by facilities for sample preparation and characterisation and data evaluation. EMBL Hamburg will also become one of the main partners in the future Centre for Structural Systems Biology (CSSB) on the DESY campus.

EMBL Hamburg has set up an ambitious research programme for structures of multifunctional proteins and protein complexes of biomedical relevance. Present research interests by group leaders include cell surface receptors, plant receptors, protein assemblies in muscle cells, protein kinases and phosphatases, and protein translocation into peroxisomes. In addition, several groups and teams have joined a common effort to determine 3D structures from potential drug targets of *Mycobacterium tuberculosis*, with the aim of contributing to the discovery of new drugs against this deadly disease. Beyond the tools in structural biology that are available on-site, EMBL Hamburg groups are engaged in many interdisciplinary collaborations with colleagues from other EMBL units, enabling access to a large variety of *in vitro* and *in vivo* functional techniques, including cellular imaging techniques.

EMBL Hamburg also 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 that allows automatic X-ray structure determination. It is integrated into an automated software pipeline, Auto-Rickshaw, which has also been compiled at EMBL Hamburg. Another package, ATSAS, allows the automatic interpretation of small angle X-ray scattering data for structural shape determination. Finally, there are two groups that focus on the development and construction of new equipment for experimental stations in structural biology, using synchrotron radiation. Present efforts focus on the installation of new robotics that allow automatic placement of biological samples into specialised synchrotron experiment facilities.

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.

Structure and function of protein complexes in biological systems

The architecture of the protein interactome in sarcomeric muscle cells. The study of the structure and function of muscle cells is of high scientific interest since many proteins found in these cells, when dysfunctional, are associated with cardiovascular diseases. The overall architecture of 'sarcomeric units' in muscle cells is established by several large protein filament systems such as actin, myosin, nebulin, titin, myomesin and obscurin. We investigate how these proteins are connected and interact with each other, frequently mediated via small scaffold proteins. Recently, we have determined the structure and function of some key complexes, including telethonin-mediated assembly of the N-terminus of titin (Zou *et al.*, 2006) and C-terminal self-assembly of myomesin (Pinotsis *et al.*, 2009, *EMBO J.*). Based on these findings, we have been able to unravel molecular key rules for complex formation of sarcomeric proteins with β-sheet domains (Pinotsis *et al.*, 2009, *TIBS*). Our future focus will be on novel protein interactions within the sarcomeric Z-disk and M-line region, and we will also increasingly investigate novel signalling functions of the protein partners involved.

Activity regulation in protein kinases. The human kinome – the part of the genome that expresses protein kinases – comprises about 600 genes. About 70 protein kinases that phosphorylate either serines or threonines share a common C-terminal autoregulatory domain that is thought to bind calcium/calmodulin (CaM). To investigate the mechanism of activity regulation in these kinases, we first determined the structure of the kinase domain from the giant filament protein titin, in the inhibited apo-conformation (Mayans *et al.*, 1998). Very recently, we were able to determine the structure of another kinase with apoptotic functions – Death Associated Protein Kinase – this time in the presence of CaM (de Diego *et al.*, 2010; figure 2). The structure provides first insight how CaM binding leads to kinase activation by withdrawing the autoregulatory domain from the kinase active site. Our future goal is to complement the ongoing structural studies by *in vitro* and *in vivo* functional studies, to decipher underlying, general molecular mechanisms that regulate the activity of members of the CaM-dependent protein kinase family. Our ultimate aim is to use these data to promote drug discovery against those kinases which frequently play critical roles in cancer formation and progression.

The architecture of the translocon of peroxisomes. Peroxisomes are cell organelles that allow sequestered metabolic processes that would interfere with other processes that generally 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) and alanine-glyoxylate aminotransferase (Fodor *et al.*, unpublished; see figure 1). Our present focus is on structural/functional studies of several other protein components of the peroxisomal translocation machinery, including the Pex19p receptor, which recognises proteins that are integrated into the peroxisomal membrane. Our ultimate goal is putting together data from known peroxisomal sub-complexes, to provide insight into the presently unknown overall architecture of the peroxisomal translocon by combined application of a variety of structural biology approaches.

Figure 1 (below): AGT-(Pex5p)2-AGT complex. Color codes: AGT, yellow, orange; Pex5p, cyan, magenta.

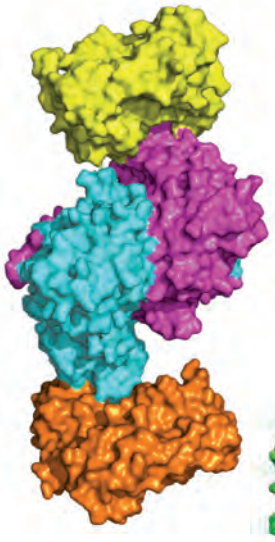
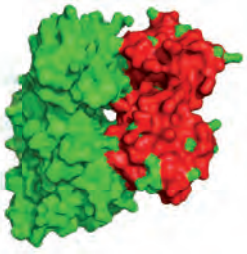


Figure 2: DAPK-CaM complex. Color codes: DAPK, green; CaM, red.



Structural systems biology in *M. tuberculosis*. 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 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). We are planning to make use of available structural data to investigate functional processes in living mycobacteria by systems biology-oriented approaches (including proteomics, metabolomics, modelling). The ultimate goal of our studies is to make the data available to promote the development of new drugs, vaccines and diagnostic markers against this pathogen.

Selected references

de Diego, I., Kuper, J., Bakalova, N., Kursula, P. & Wilmanns, M. (2010). Molecular basis of the death-associated protein kinase-calcium/calmodulin regulator complex. *Sci Signal.*, 3, ra6

Bertz, M., Wilmanns, M. & Rief, M. (2009). The titin-telethonin complex is a directed, superstable molecular bond in the muscle Z-disk. *Proc. Natl. Acad. Sci. USA*, 106, 13307-133310

Pinotsis, N., Abrusci, P., Djinovic-Carugo, K. & Wilmanns, M. (2009). Terminal assembly of sarcomeric filaments by intermolecular beta-sheet formation. *Trends Biochem. Sci.*, 34, 33-39

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, 253-64

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.

Instrumentation for synchrotron beamlines for structural biology at PETRA III

Previous and current research

EMBL is designing, building and will operate three beamlines for structural biology at the PETRA III synchrotron radiation source. The 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). Construction includes the provision of the X-ray optical elements, experimental endstations, vacuum system, control system, data acquisition system, technical infrastructure and parts of the civil engineering, for which our team provides expertise in X-ray optics, precision mechanical engineering, robotics, control software and electronics.

By 2009, most of the principal optical elements had been constructed and already partly manufactured, including the high-heatload monochromators and cooling systems, Kirkpatrick-Baez focusing optics with adaptive bimorph mirrors, deflection mirrors for angular separation of adjacent canted undulator beams and cooled UHV slit systems. Our work focuses on the selection, customisation and integration of mechanics, control electronics and control software. In addition, our activity ranges from the development of instruments from scratch to the performance of relevant validation experiments and commissioning of instrumentation for user experiments. Examples of this are the construction of a focusing 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) or the development of nanometer resolution slits.

Another important project is the development of a completely updated version of a robotic sample mounting system for protein crystals, MARVIN (see figure). The system is characterised by its high sample storage capacity, high sample mounting speed, improved geometry and maintainability. It has been commissioned at the BW7b beamline at DORIS and will be now cloned for the MX beamlines at PETRA.

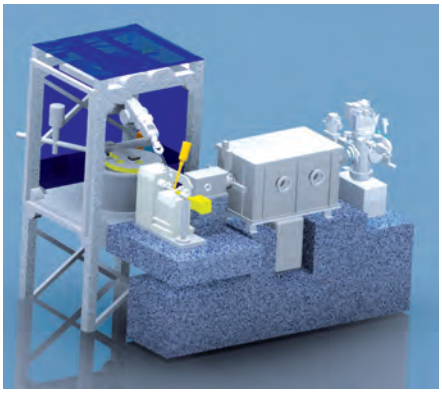
As a side project, a new high precision goniometer axis with sub-micrometer precision has been built. These elements are now integrated into a software-based modern control system which allows for a heterogeneous control environment and provides distributed access and monitoring. This is a prerequisite for the planned remote-controlled user experiments at the PETRA III beamlines.

Future projects and goals

In the next year our efforts will concentrate on:

- installation and commissioning of the new beamline elements so that first users can be received;
- 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 number of selected instrumental development projects (e.g. robotic sample changer for MX, 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, as well as exploring the possibilities and requirements for biological imaging research at PETRA III.



Selected references

Keyriläinen, J., Fernández, M., Karjalainen-Lindsberg, M.L., Virkkunen, P., Leidenius, M., von Smitten, K., Sipilä, P., Fiedler, S., Suhonen, H., Suortti, P. & Bravin, A. (2008). Toward high-contrast breast CT at low radiation dose. *Radiology*, 249, 321-327

Ristau, U., Di Castro, M., Pazos, A. & Fiedler, S. (2008). The concept of EMBL beamline control at PETRA III. *Proceedings of PCaPAC08*, MOZ02, 22-24

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

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

Christoph Hermes

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.

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.

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.

Selected references

Roessle, M.W., Klaering, R., Ristau, U. *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

Hermes, C. Gehrmann, T., Jahn, D. *et al.* (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. *et al.* (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

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 of outstation since 1997.

Development of an integrative modelling platform for structural biology

Science fiction often predicts the direction of cutting-edge research and suggests a future in which humans live and interact in physical realities modelled *in silico* (e.g. James Cameron's download of the human consciousnesses into foreign bodies in *Avatar*). A key element in such models would be a complete description of all molecular interactions within living organisms.

We are fascinated by complex computational methods in information processing that can address data interpretation problems as we encounter them in structural biology. Indeed, recognising patterns in experimental data that describe macromolecules is itself an application of artificial intelligence (see figure). Structure determination provides essential data for integrative modelling of the basis of life: DNA, RNA, proteins, macromolecular complexes and assemblies. Current approaches, largely based on macromolecular X-ray crystallography, are static in nature and concentrate on a reductionist view of a single structure from a single method or experiment. Future applications (e.g. a quantitative description of the living cell) will necessitate radically different approaches where a wider context of information, using data from complementary tools, is implemented in computational methods serving as an integrated platform for a model of life.

Previous and current research

Some of aspects of integrative modelling are addressed in one of the group's main foci, the ARP/wARP software project (Langer *et al.*, 2008) for protein/DNA/ligand crystal structure determination. The comprehensive nature of its methods makes ARP/wARP applicable to the examination of protein assemblies and complexes at both high and low resolution. The design of artificial intelligence to find the best way through the maze of structure determination is implemented in the group's AutoRickshaw software for validation of synchrotron beamline experiments. The group continues to supply advanced in-house computational facilities for hundreds of researchers worldwide to perform remote structure determination.

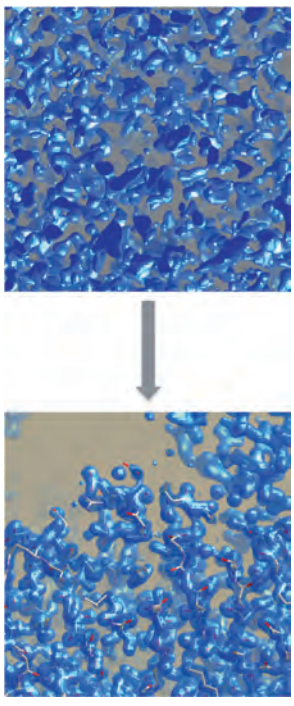
One of our emerging methods aims at combining X-ray diffraction with low-resolution methods such as electron microscopy. Recognising structural motifs and fold patterns at resolution as low as 20 Å (Heuser *et al.*, 2009) greatly aids investigation of large macromolecular complexes, and we make use of our expertise to interpret maps associated with important membrane proteins. The ultimate result is a complete and precise model of the structure and associated co-factors/ligands. Making use of crystallographic data obtained at ultra-high resolutions, we have applied quantum chemical calculations to the analysis of atomic movements and charge shifts; the results often reveal subtle but important aspects of the protein's function. These findings are essential in computational and pharmaceutical drug design.

The group also uses its expertise to develop methods that aid the analysis of crystal growth. We evaluated images and developed innovative methods to automatically detect crystals. These procedures (the group's XREC suite) are being incorporated into EMBL's new PETRA III beamlines.

On the application side, we are interested in human parasite genome analysis and enzymes involved in drug and vitamin syntheses (Ortiz de Orué Lucana *et al.*, 2009; Lapkouski *et al.*, 2009). These examples, in addition to illustrating complex biological pathways, also drive the development of our integrative model building platform, spanning the resolution range from 'globular' to 'atomic'.

Future projects and goals

Our activities will continue to focus on arising trends in the field, aiming to push currently perceived boundaries and helping to shape future structural biology research. The field will play an increasing role in investigations of biological interactions, and we will remain at the forefront of developments.



Evolution of macromolecular patterns in an electron density map during X-ray crystal structure determination.

Selected references

Langer, G., Cohen, S.X., Lamzin, V.S. & Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nature Protocols*, 3, 1171-1179

Heuser, P., Langer, G.G. & Lamzin, V.S. (2009). Interpretation of very low resolution X-ray electron density maps using core objects. *Acta Crystallogr. D Biol. Crystallogr.*, 65, 690-696

Ortiz de Orue Lucana D., Bogel G., Zou P. & Groves M.R. (2009). The oligomeric assembly of the novel haem-degrading protein HbpS is essential for interaction with its cognate two-component sensor kinase. *J. Mol. Biol.*, 386, 1108-1122

Lapkouski, M., Panjikar S., Janscak, P., Smatanova, I.K., Carey, J., Ettrich, R. & Csefalvay, E. (2009). Structure of the motor subunit of type I restriction-modification complex EcoR124I. *Nat. Struct. Biol.*, 16, 94-95

Rob Meijers

PhD 2001, EMBL Hamburg/ University of Amsterdam.

Postdoctoral research at the Dana Farber Cancer Institute, Boston.

Staff scientist at the Synchrotron Soleil, Saint Aubin, France, 2006-2009.

Group leader at EMBL Hamburg since 2009.

Selected references

Meijers, R., Lai, C.C., Yang, Y., Liu, J.H., Zhong, W., Wang, J.H. & Reinherz, E.L. (2005). Crystal structures of murine MHC Class I H-2 D(b) and K(b) molecules in complex with CTL epitopes from influenza A virus: implications for TCR repertoire selection and immunodominance. *J. Mol. Biol.*, 345, 1099-110

Wang, J.H., Meijers, R., Xiong, Y., Liu, J.H., Sakihama, T., Zhang, R., Joachimiak, A. & Reinherz, E.L. (2001). Crystal structure of the human CD4 N-terminal two-domain fragment complexed to a class II MHC molecule. *Proc. Natl. Acad. Sci. USA*, 98, 10799-804

Meijers, R., Puettmann-Holgado, R., Skiniotis, G., Liu, J.H., Walz, T., Wang, J.H. & Schmucker, D. (2007). Structural basis of Dscam isoform specificity. *Nature,* 449, 487-91

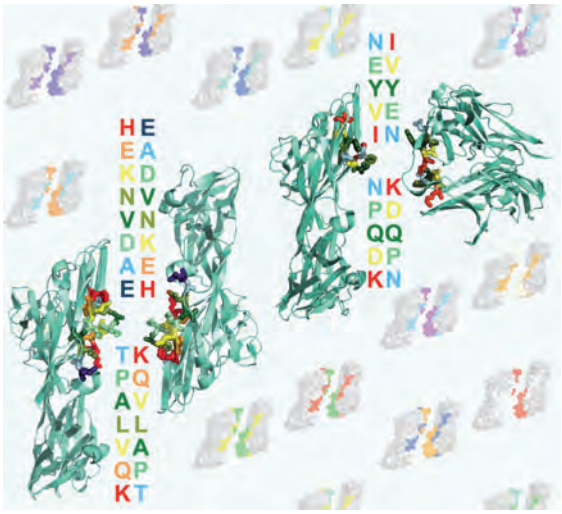
Structural biology of cell surface receptors

Previous and current research

Intercellular interactions occur through supramolecular clusters that form asymmetric kinapses when a cell is scanning the environment, and symmetric synapses when cell–cell recognition is established. This ancient recognition process has been observed in such diverse systems as predatory amoeba, lymphocytes scanning for antigens and nerve cells seeking connections to form a brain map. At the centre of these interactions are very specific molecular recognition events that trigger a reorganisation of the cluster on the cell surface. This in turn amplifies the recognition event, resulting in the activation of a signalling cascade within the cell that leads to physiological changes within the cell. We study the molecular basis of cell surface receptor recognition in the context of the dynamics of the supramolecular cluster as a whole.

Immune receptors: T cell receptors (TCR)s on lymphocytes are the central recognition unit of a supramolecular complex that scans for pathogenic peptides loaded on MHC molecules at the surface of host cells. We have studied the molecular basis for the dominant response of the adaptive immune system to certain influenza peptides, which provides insights that should help in the design of a peptide vaccine against influenza (Meijers *et al.*, 2005). We have also studied the interaction between the T cell receptor/MHC complex and CD4, a cell surface receptor that acts as a coreceptor in T cell recognition (Wang *et al.*, 2001). CD4 is a prime fusion target of the HIV virus, and the structure between CD4 and the MHC class II molecule I-Ak shows that the viral envelope protein gp120 and the MHC molecule bind the same region of CD4. The structure confirmed that immunodeficiency is caused in part by the disruption of the binding of MHC class II molecules to CD4, an essential coreceptor in the supramolecular TCR complex.

Neuronal receptors: The nervous system consists of neuronal circuits, and it is thought that the individual neurons find their targets and establish synaptic connections within the circuit following a specific developmental program. Some of the molecules that guide the neurons to their targets are cell surface receptors that provide a unique identity tag to each neuron. The Down syndrome cell adhesion molecule (Dscam) from *Drosophila* was identified as an axon guidance receptor that has the potential to provide a large number of unique identity tags. Dscam contains three variable extracellular immunoglobulin domains, which can provide 19,008 unique cell surface receptor identity tags through splicing. We have studied the structural basis for the remarkable specificity of the receptor, which seems to interact only with receptors that consist of identical isoforms (Meijers *et al.*, 2007). The structure shows how variable regions of the Dscam receptor interact through a palindromic hydrogen bonding network that is unique for each isoform.



The identification tag of the Dscam receptor is encoded in a palindromic hydrogen bonding network. Two receptors that contain the same code bind together, but a slight difference in amino acid sequence prevents binding. The Drosophila *Dscam gene is spliced to generate 19,018 different receptors that provide each neuron in the brain with a unique identity tag. This helps the neurons to orient themselves, and to decide where to form a connection with another neuron.*

Future projects and goals

Viruses and bacterial pathogens use cell surface receptors to invade host cells, but they also deregulate the order established in the supramolecular cluster to jam the recognition machinery of the host. The same (de)regulation mechanism is used by the immune system itself which employs internal regulators that act on the synapse, such as hormones and cytokines, to tune the response of the immune system. We are using molecular fragments of pathogenic and self-regulating factors to gauge the interactions with and within the supramolecular complex. A detailed understanding of these interactions will allow us to tinker with cell surface receptors in order to manipulate the behaviour of certain individual cells.

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.

X-ray crystallography, high-throughput crystallisation

Previous and current research

Structural biology: Our goal is to understand the molecular principles of infectious disease and important cellular processes at a structural level. In the area of infectious disease, our research focuses on target proteins from *Mycobacterium tuberculosis* (Mtb). We work with proteins 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 figure shows the structure of the rare iso-citrate dehydrogenase II (Rv0066c), whcih displays unexpected oligomeric characteristics which we characterised.

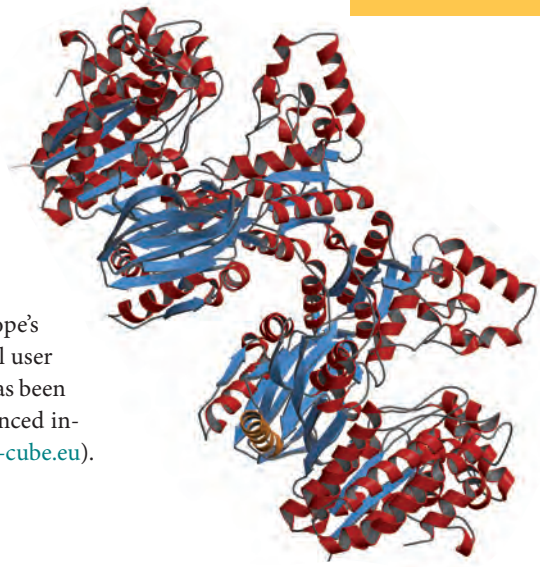
Posttranslational modification is an important cellular mechanism to control enzyme activity, subcellular localisation or interaction with specific binding partners. There is a variety of chemical entities used in the covalent modification of proteins which enables the cell to integrated intra- and extracellular signals and establish an intricate network of interactions. ADP-ribosyl transferases (ARTs) catalyse the covalent attachment of ADP-ribose to specific amino acid side chains while ADP-ribosyl hydrolases (ARHs) reverse the reaction. Akin to phosphorylation, ADP-ribosylation modifies its target activity. Ribosylation was originally observed by bacterial toxins manipulating functionally important proteins of their host cells. Since then, numerous regular functions of the cell, like DNA-repair, chromatin decondensation, transcription, telomere function, mitotic spindle formation and apoptosis have been discovered. Whereas the characterisation of ADP-ribosylation has made some progress in recent years, the removal of ADP-ribose is still very poorly understood. The human genome encodes for at least three different ADP-ribosyl hydrolases, ARH1, ARH2 and ARH3 which share little (<20%) sequence identity. We have structurally characterised human ARH1 and have obtained crystals of human ARH1.

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 from eukaryotic two-component systems to classical MAPK signalling pathways 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 an aim to functionally characterise the early steps of ethylene signaling through inter- and cross-domain activation of its components.

High-throughput crystallisation: EMBL Hamburg operates one of Europe's largest high-throughput crystallisation facilities, which is open to the general user community. Users come from 20 different nations. Since 2009 our platform has been part of the European FP7 initiative P-CUBE, which supports access to advanced infrastructures. More information on this initiative is available at (http://www.p-cube.eu).

Future projects and goals

We will continue structural analysis of biological macromolecules and their complexes. Ethylene reception in particular requires an interdisciplinary approach due to transient complex formation which additionally depends on signal induced auto-phosphorylation. To this end, we will use a combination protein characterisation tools, X-ray crystallography and SAXS.



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.

Selected references

Kernstock, S., Koch-Nolte, F., Müller-Dieckmann, J., Weiss, M.S. & Mueller-Dieckmann, C. (2009). Cloning, expression, purification and crystallization as well as X-ray fluorescence and preliminary X-ray diffraction analyses of human ADP-ribosylhydrolase 1. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.*, 65, 529-532

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

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

Thomas Schneider

PhD 1996, Technical University of Munich/EMBL.

Postdoctoral research at the MPI 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.

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

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).

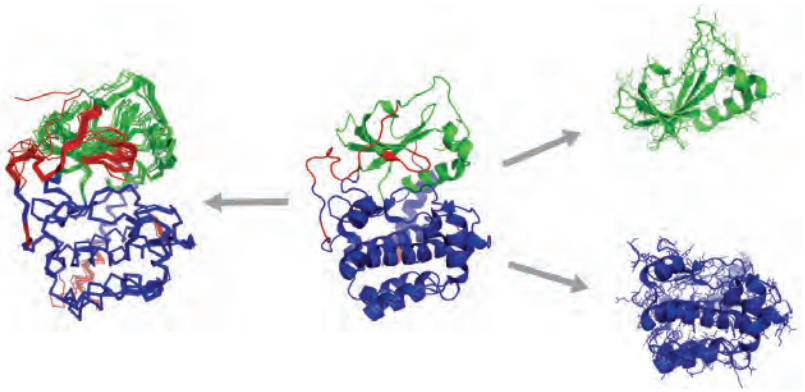
Partly due to the enormous progress in synchrotron radiation-based structural biology, structural data on biological macromolecules are produced at an ever-increasing rate, creating the need to develop tools for efficient mining of structural data. We are developing tools for which the central concept 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.

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-focusing and energy tunability, allowing imaging of the content of small crystals containing large objects such as multi-component complexes.

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



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

Dmitri Svergun

PhD 1982, Institute of Crystallography, Moscow.
Dr. of Science 1997, Institute of Crystallography, Moscow.
At EMBL since 1991. Group leader since 2003.

Small-angle X-ray scattering from macromolecular solutions

Previous and current research

Small-angle X-ray scattering (SAXS) reveals low resolution (1-2 nm) structures of biological macromolecules in close-to-native solutions for an extremely broad range of sizes from small peptides to huge macromolecular machines and at variable conditions. This versatility and universality – and the fact that it does not need crystals to characterise the structure – makes SAXS an ideal tool for a systemic structural biology approach. For many complicated biological systems which may be flexible or have a dynamic nature, SAXS is the only method capable of providing structural information. Recent experimental and methodical developments have significantly enhanced the resolution and reliability of SAXS-based structural models, and the last decade saw biological solution SAXS become a streamlined tool in structural molecular biology.

Our group leads the development of novel computational methods for constructing 3D structural models from the scattering data. Special attention is given to the joint use of SAXS with other structural, biophysical and biochemical techniques including crystallography, NMR, electron microscopy, neutron scattering and bioinformatics. We developed the world most used program package, ATSAS, employed in over 1,000 laboratories, and we continue providing the scientific community with novel approaches.

We run a synchrotron beamline X33, dedicated to biological solution SAXS, at DESY's storage ring, DORIS-III. The rapidly-growing demand for SAXS in the biological community has led to a sixfold increase in the user turnover at X33 during the last decade. X33 is the first synchrotron SAXS beamline with a robotic sample changer and a data analysis pipeline for building structural models online. The level of automation allows for remote access to the beamline, and the world's first remote SAXS experiment was performed at X33 in 2009 (see figure).

Most of the external users of X33 are seeking collaborative projects where the SAXS group members help not only with data collection but also with modelling. In numerous exciting applications, SAXS is employed 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, hierarchical sytems, processes of amyloid fibrillation and many other objects of high biological and medical importance.

Future projects and goals

The present and future work of the group includes:

- further development of novel methods and approaches for the reconstruction of tertiary and quaternary structure of macromolecules and complexes from X-ray and neutron scattering data;
- the use of bioinformatic to construct and validate SAXS-based models and the joint applications of SAXS with crystallography, NMR and other methods;
- participation in collaborative projects at X33 beamline employing SAXS to study the structure of a wide range of biological systems in solution;
- a complete automation of biological SAXS experiment and data analysis at X33, and, in collaboration with the PETRA III group, at the new high-brilliance BioSAXS beamline at the third-generation PETRA storage ring at DESY.

The world's first remote SAXS experiment in the lobby of Nanyang Technological University during a biological SAXS Course (Singapore, 26 May 2009). Left screen: remote access interface with cameras displaying the SAXS robot and the sample cell. Right screen: Skype showing the members of the EMBL BioSAXS group monitoring the experiment in Hamburg.

Selected references

Cheng, Z., Saito, K., Pisarev, A.V., Wada, M., Pisareva, V.P. *et al.* (2009). Structural insights into eRF3 and stop codon recognition by eRF1. *Genes Dev.*, 23, 1106-18

De Marco, V., Gillespie, P.J., Li, A., Karantzelis, N., Christodoulou, E., Klompmaker, R. *et al.* (2009). Quaternary structure of the human Cdt1-Geminin complex regulates DNA replication licensing. *Proc. Natl. Acad. Sci. USA*, 106, 19807-12

Fagan, R.P., Albesa-Jove, D., Qazi, O., Svergun, D.I., Brown, K.A. & Fairweather, N.F. (2009). Structural insights into the molecular organization of the S-layer from *C. difficile*. *Mol Microbiol.*, 71, 1308-22

Janssen, B.J., Gomes, L., Koning, R.I., Svergun, D.I., Koster, A.J., Fritzinger, D.C., Vogel, C.W. & Gros, P. (2009). Insights into complement convertase formation based on the structure of the factor B-cobra venom factor complex. *EMBO J.* 28, 2469-78

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 sensory perception. 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 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 use of 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 Postdoctoral (EIPOD) Fellowship Programme 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

Nadia Rosenthal

PhD 1981, Harvard Medical School.
Postdoctoral research at the NCI.
Assistant Professor, Boston University Medical Center.
Associate Professor, Mass. General Hospital, Harvard Medical School. Group leader and Head of EMBL Monterotondo since 2001.
Founding Director, Australian Regenerative Medicine Institute, Monash U., Melbourne, 2007.

Regenerative mechanisms in heart and skeletal muscle

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.

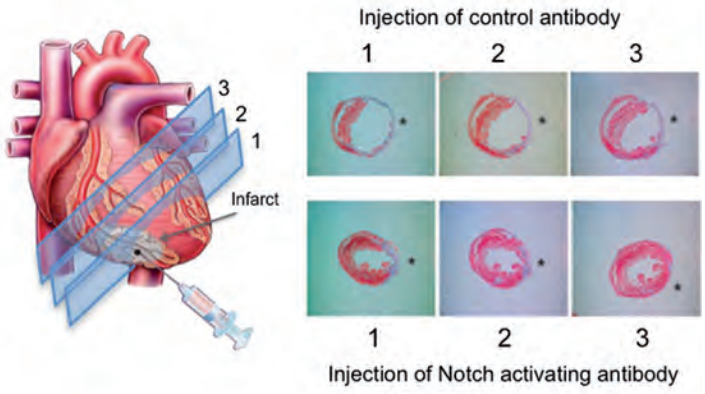
We found that insulin-like growth factors attenuate muscle atrophy and improve regeneration in ageing, muscular dystrophy and cardiomyopathies. Delivery of unprocessed IGF-1 isoforms to various mouse models of human muscle pathologies implicate this growth factor as a powerful enhancer of the regeneration response. Selective muscle fibre loss, fibrosis and functional decline in ageing or diseased muscle and heart tissues can be blocked by localised transgenic or viral delivery of IGF-1 isoforms, which also promotes recruitment of stem cells to sites of injury.

Expression of IGF-1 proteins *in vivo* has allowed us to assign specific functions of different peptide domains in muscle hypertrophy and regeneration. The varied responses evoked by different IGF-1 isoforms suggest specific mechanisms through which combinations of supplemental growth factors can improve regeneration, providing new targets for clinical intervention. Studies in skeletal and cardiac muscle have implicated FGF, NFkB and Notch -mediated signalling pathways in the intervention of damage and disease. New avenues of research are investigating regulatory networks in the epicardium, the outermost cell layer of the heart, which may contribute to the regenerative response in that organ.

Since IGF-1 modulates the expression of inflammatory cytokines, we are currently investigating the role of the innate immune system in the regeneration process. Several lines of evidence suggest that improvement both skeletal muscle and cardiac regeneration operates in part by modulation of the inflammatory response. As macrophages have been proposed to play important roles in tissue repair, we have recently collaborated with the Nerlov group (formerly EMBL) to develop a genetic model in which prevention of macrophage polarization by the regenerating environment blocks muscle regeneration. Similar roles played by related immune cells are currently being explored in the adult heart.

Future projects and goals

In our future research, we will exploit new conditional and inducible mouse genetic models and transgenic markers to characterise key mechanisms governing the regeneration of mammalian tissues. Our studies are designed to define the common nodal points of signaling in regenerative processes as they relate to embryonic development. We aim to define the molecular action whereby selected growth factors and their intracellular intermediates improve healing. At the cellular level, we are particularly interested in the role played by immune 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.



Activation of the Notch signaling pathway in the heart improves tissue repair after injury. Adult wild-type mice were subjected to myocardial infarction by coronary artery ligation, and a Notch1 activating antibody or a control antibody was injected near the infarct 5 minutes later. Four weeks after the injury, trichrome staining on cross-sections of injected hearts shows reduced collagen-rich scar tissue (blue) near the infarct (asterisk) in hearts where Notch signaling was activated .

Selected references

Kratsios, P., Catela, C., Salimova, E., Huth, M., Berno, V., Rosenthal, N. & Mourkioti, F. (2010). Redeployment of Notch1 embryonic signaling mediates cardiac morphogenesis and promotes myocardial repair in the adult. *Circ. Research*, in press

Kratsios, P., Huth, M., Al Banchaabouchi, M., Sgoifo, A., Manghi, M. *et al.* (2010). Antioxidant amelioration of dilated cardiomyopathy caused by conditional deletion of NEMO/IKK in cardiomyocytes. *Circ. Research*, 106, 133-44.

Ruffell, D., Mourkioti, F., Gambardella, A. *et al.* (2009). A CREB-C/EBPβ cascade induces M2 macrophage-specific gene expression and promotes muscle injury repair. *Proc Natl. Acad. Sci*, 106, 17475-80

Catela, C., Bilbao-Cortes D. *et al.* (2009). Multiple congenital malformations of Wolf-Hirschhorn syndrome are recapitulated in Fgfrl1 null mice. *Disease Models and Mechs*, 2, 283-94

PhD 1995, Yale University.
Postdoctoral research at Columbia University.
Group leader at EMBL Monterotondo since 2003.
Senior scientist since 2009.

Cornelius Gross

Developmental programming of anxiety

Previous and current research

Anxiety disorders are debilitating mental illnesses characterised by excessive worry and rumination 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 of anxiety disorders in 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 human.

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 behavior. 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 signaling 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 validation of the neurophysiological correlates of anxiety in 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. *et al.* (2008). Sporadic autonomic dysregulation and death associated with excessive serotonin autoinhibition. *Science*, 321, 130-133

Lo Iacono, L. & Gross, C. (2008). Alpha-Ca^{2+}/calmodulin-dependent protein kinase II contributes to the developmental programming of anxiety in serotonin receptor 1A knock-out mice. *J. Neurosci.*, 28, 6250-6257

Carola, V., Frazzetto, G., Pascucci, T., Audero, E., Puglisi-Allegra, S. *et al.* (2008). Identifying molecular substrates in a mouse model of the serotonin transporter x environment risk factor for anxiety and depression. *Biol. Psychiatry*, 63, 840-846

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

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.

Molecular physiology of somatosensation

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.

We use a combination of molecular, imaging and electrophysiological techniques to examine functional properties of sensory neurons at their peripheral and central terminals. 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.

At the molecular level, we are interested in mechanisms of touch sensitivity of sensory neurons. Normal mechanical sensitivity is dependent upon a complex of proteins that are localised at the peripheral endings of sensory neurons. Evidence supports a central role for stomatin-like proteins and a family of ion channels called ASICs in this complex. Using cellular, electrophysiological and molecular imaging techniques we are probing the nature of interactions between these proteins and characterising their function in the mechanotransduction complex in detail.

Another focus of the group is to understand the biophysical properties of ion channels involved in sensory transduction. Much of our work has concentrated on the ion channel TRPA1, a member of the Transient Receptor Potential (TRP) family of channels. In mammals, TRPA1 is expressed by nociceptors and plays a key role in detecting noxious chemicals. We demonstrated that intracellular Ca^{2+} directly activates TRPA1 via an EF-hand domain in the N-terminus of the protein and that Ca^{2+} is essential for normal activation of the channel by noxious chemicals. We are now interested in how TRP channels have evolved to become multimodal sensors across several phyla. Using a combination of computational and electrophysiological methods we are examining activation mechanisms in order to understand how these channels function as sensors for a diverse range of physical stimuli.

Future projects and goals

A major focus of the laboratory is to correlate cellular studies on somatosensation with observations made at the physiological level. To this end we are employing genetic approaches combined with electrophysiological and molecular imaging techniques. Future goals include:

- identification of novel genes involved in touch and pain;
- mutagenesis of transduction channels and associated proteins to determine their mechanism of action;
- tissue-specific and conditional mutagenesis of sensory-related genes in defined subpopulations of sensory neuron;
- development of new techniques to measure functional properties of sensory neurons at their terminals.

Selected references

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, 277-279

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, 206-209

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, 8107-8112

PhD 1999, Research Institute of Molecular Pathology, Vienna.

Postdoctoral research at The Rockefeller University, New York.

Group leader at EMBL Monterotondo since 2007.

Adjunct Member of Faculty, The Rockefeller University, since 2007.

Dónal O'Carroll

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 multiprotein RNA-induced silencing complex (RISC). At the core of RISC is an Argonaute (Ago) protein that binds miRNA and executes their function. Using hematopoiesis in mice as a model system to study the physiological function of Ago proteins and the mechanism of miRNA-mediated gene silencing *in vivo*, 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. Our current interests regarding RISC now focus on the post-translational regulation of RISC *in vivo*. Having identified erythroid development as being sensitive to miRNA dosage, we now strive to mechanistically understand how key miRNA loci regulate terminal erythropoiesis. In addition, we are exploring the contribution of individual miRNAs to the development and homeostasis of spermatogenesis.

Transposable elements are mobile genetic elements that constitute a large fraction of mammalian genomes. In mice, these highly mutagenic elements are epigenetically silenced throughout life to avoid the deleterious effects of transposition. Transposon silencing is established during germ cell development and maintained post-fertilisation in all somatic cells. This epigenetic process of transposons silencing is of fundamental importance for germ cell development and the genomic integrity of the gametes and thus the future soma. Highly conserved ribonuclear particles (RNPs) have evolved that target transposons for DNA methylation and subsequent epigenetic silencing. In the male germline members of the Piwi subclade of the Argonaute family of proteins, Mili and Miwi2 are essential for *de novo* DNA methylation of transposons and spermatogenesis. Both Mili and Miwi2 bind a class of small non-coding RNAs known as Piwi-interacting RNAs (piRNAs) that are believed to act as guides for targeting of the respective RNPs. We currently address several basic questions on the intrinsic mechanism and function of mammalian Piwi proteins function.

Future projects and goals

- Identify miRNAs and their respective targets that control erythropoiesis;
- determine the *in vivo* significance of post-translation modifications of Ago2;
- explore the function of microRNAs during spermatogenesis;
- understand the mechanism of Miwi2 and Mili in the process of epigenetic transposon silencing.

We currently address our goals using state-of-the-art mouse genetic strategies coupled with high throughput sequencing approaches.

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, 1553-1558

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, 644-648

Yi, R., O'Carroll, D., Pasolli, H.A., Zhang, Z., Dietrich, F.S. *et al.* (2006). Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nat. Genet.*, 38, 356-362

Index

Notes

Notes

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 Heads of Units

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

Cover design:
Vienna Leigh

Photos: EMBL Photolab (page 5, Samuel Bader; page 97, Rosemary Wilson).

Printed by ColorDruck, Leimen

European Molecular Biology Laboratory

Heidelberg • Hinxton • Grenoble • Hamburg • Monterotondo